

ACOUSTICAL NEWS-USA		3133
USA Meeting Calendar		3134
ACOUSTICAL NEWS-INTERNATIONAL		3137
International Meeting Calendar		3137
REVIEWS OF ACOUSTICAL PATENTS		3139
LETTERS TO THE EDITOR		
Enhanced sound transmission from water to air at low frequencies (L)	B. Edward McDonald, David C. Calvo	3159
Negative axial radiation forces on solid spheres and shells in a Bessel beam (L)	Philip L. Marston	3162
GENERAL LINEAR ACOUSTICS [20]		
Acoustic diffraction by deformed edges of finite length: Theory and experiment	Timothy K. Stanton, Dezhong Chu, Guy V. Norton	3167
Higher-order acoustic diffraction by edges of finite thickness	Dezhong Chu, Timothy K. Stanton, Allan D. Pierce	3177
On appropriate equivalent monopole sets for rigid body scattering problems	Yves J. R. Gounot, Ricardo E. Musafir	3195
The integrated extinction for broadband scattering of acoustic waves	Christian Sohl, Mats Gustafsson, Gerhard Kristensson	3206
NONLINEAR ACOUSTICS [25]		
Acoustoelasticity in soft solids: Assessment of the nonlinear shear modulus with the acoustic radiation force	J.-L. Gennisson, M. Rénier, S. Catheline, C. Barrière, J. Bercoff, M. Tanter, M. Fink	3211
A two-dimensional pseudospectral model for time reversal and nonlinear elastic wave spectroscopy	Thomas Goursolle, Samuel Callé, Serge Dos Santos, Olivier Bou Matar	3220
Accelerometer measurements of acoustic-to-seismic coupling above buried objects	Keith Attenborough, Qin Qin, Jonathan Jefferis, Gary Heald	3230
AEROACOUSTICS, ATMOSPHERIC SOUND [28]		
On the one-dimensional acoustic propagation in conical ducts with stationary mean flow	Ana Barjau	3242

CONTENTS—Continued from preceding page

UNDERWATER SOUND [30]

Low frequency sound scattering from spherical assemblages of bubbles using effective medium theory	Thomas R. Hahn	3252
Geoacoustic inversion using multipath pulse shape	Mark K. Prior, Chris H. Harrison, Peter L. Nielsen	3268
Near-specular acoustic scattering from a buried submarine mud volcano	Anthony L. Gerig, Charles W. Holland	3280
Modeling scattering from azimuthally symmetric bathymetric features using wavefield superposition	John A. Fawcett	3286
Passive reverberation nulling for target enhancement	H. C. Song, W. S. Hodgkiss, W. A. Kuperman, K. G. Sabra, T. Akal, M. Stevenson	3296
Determining dominant scatterers of sound in mixed zooplankton populations	Andone C. Lavery, Peter H. Wiebe, Timothy K. Stanton, Gareth L. Lawson, Mark C. Benfield, Nancy Copley	3304
Full wave-field reflection coefficient inversion	Jan Dettmer, Stan E. Dosso, Charles W. Holland	3327
Subbottom profiling using a ship towed line array and geoacoustic inversion	T. C. Yang, Kwang Yoo, L. T. Fialkowski	3338
Restless rays, steady wave fronts	Oleg A. Godin	3353
Assessing a dual-frequency identification sonars' fish-counting accuracy, precision, and turbid river range capability	Suzanne L. Maxwell, Nancy E. Gove	3364
Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999–2002	Kathleen M. Stafford, David K. Mellinger, Sue E. Moore, Christopher G. Fox	3378
Environmental inversion using high-resolution matched-field processing	Cristiano Soares, Sérgio M. Jesus, Emanuel Coelho	3391

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

Resonant generation of surface acoustic waves between moving and stationary piezoelectric crystals	Vladimir N. Khudik, Constantine E. Theodosiou	3405
Entrainment and stimulated emission of ultrasonic piezoelectric auto-oscillators	Richard L. Weaver, Oleg I. Lobkis, Alexey Yamilov	3409

TRANSDUCTION [38]

Effects of coupled vibrations on the acoustical performance of underwater cylindrical shell transducers	Boris Aronov, David A. Brown, Corey L. Bachand	3419
Development of a micromachined piezoelectric microphone for aeroacoustics applications	Stephen Horowitz, Toshikazu Nishida, Louis Cattafesta, Mark Sheplak	3428

STRUCTURAL ACOUSTICS AND VIBRATION [40]

Reconstruction of Rayleigh–Lamb dispersion spectrum based on noise obtained from an air-jet forcing	Eric Larose, Philippe Roux, Michel Campillo	3437
Response variance prediction for uncertain vibro-acoustic systems using a hybrid deterministic-statistical method	R. S. Langley, V. Cotoni	3445

NOISE: ITS EFFECTS AND CONTROL [50]

Generating sensor signals in isotropic noise fields	Emanuël A. P. Habets, Sharon Gannot	3464
---	-------------------------------------	------

CONTENTS—Continued from preceding page

Analytic mode matching for a circular dissipative silencer containing mean flow and a perforated pipe	Ray Kirby, Francisco D. Denia	3471
Evaluating noise in social context: The effect of procedural unfairness on noise annoyance judgments	Eveline Maris, Pieter J. Stallen, Riel Vermunt, Herman Steensma	3483
ACOUSTIC SIGNAL PROCESSING [60]		
Coherent acoustic communication in a tidal estuary with busy shipping traffic	Paul A. van Walree, Jeffrey A. Neasham, Marco C. Schrijver	3495
Optimal illumination and wave form design for imaging in random media	Liliana Borcea, George Papanicolaou, Chrysoula Tsogka	3507
PHYSIOLOGICAL ACOUSTICS [64]		
A computer model of medial efferent suppression in the mammalian auditory system	Robert T. Ferry, Ray Meddis	3519
Multifield coupled finite element analysis for sound transmission in otitis media with effusion	Rong Z. Gan, Xuelin Wang	3527
Distortion product otoacoustic emissions: Cochlear-source contributions and clinical test performance	Tiffany A. Johnson, Stephen T. Neely, Judy G. Kopun, Darcia M. Dierking, Hongyang Tan, Connie Converse, Elizabeth Kennedy, Michael P. Gorga	3539
Cochlear reflectivity in transmission-line models and otoacoustic emission characteristic time delays	Renata Sisto, Arturo Moleti, Christopher A. Shera	3554
Comparing stimulus-frequency otoacoustic emissions measured by compression, suppression, and spectral smoothing	Radha Kalluri, Christopher A. Shera	3562
Wavelet and matching pursuit estimates of the transient-evoked otoacoustic emission latency	Giuseppe Notaro, Adnan Mohsin Al-Maamury, Arturo Moleti, Renata Sisto	3576
PSYCHOLOGICAL ACOUSTICS [66]		
The effect of hair on auditory localization cues	Bradley E. Treeby, Jie Pan, Roshun M. Paurobally	3586
Spectral integration and wideband analysis in gap detection and overshoot paradigms	Joseph W. Hall, III, Emily Buss, John H. Grose	3598
The Franssen effect illusion in budgerigars (<i>Melopsittacus undulatus</i>) and zebra finches (<i>Taeniopygia guttata</i>)	Micheal L. Dent, Elizabeth M. McClaine, Thomas E. Welch	3609
Detection and discrimination of simple and complex sounds by hearing-impaired Belgian Waterslager canaries	Amanda M. Lauer, Robert J. Dooling, Marjorie R. Leek, Kirsten Poling	3615
Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled	David R. R. Smith, Thomas C. Walters, Roy D. Patterson	3628
Perceptual thresholds for detecting modifications applied to the acoustical properties of a violin	Claudia Fritz, Ian Cross, Brian C. J. Moore, Jim Woodhouse	3640
Within- and across-channel gap detection in cochlear implant listeners	John H. Grose, Emily Buss	3651
SPEECH PRODUCTION [70]		
Unsteady behavior of flow in a scaled-up vocal folds model	Michael Krane, Michael Barry, Timothy Wei	3659
Effects of perturbation magnitude and voice F_0 level on the pitch-shift reflex	Hanjun Liu, Charles R. Larson	3671

CONTENTS—Continued from preceding page

SPEECH PERCEPTION [71]

Acoustic cues to lexical segmentation: A study of resynthesized speech	Stephanie M. Spitzer, Julie M. Liss, Sven L. Mattys	3678
On the perception of similarity among talkers	Robert E. Remez, Jennifer M. Fellowes, Dalia S. Nagel	3688
The effect of word learning on the perception of non-native consonant sequences	Lisa Davidson, Jason Shaw, Tuuli Adams	3697

BIOACOUSTICS [80]

Intra- and intergroup vocal behavior in resident killer whales, <i>Orcinus orca</i>	Brigitte M. Weiß, Helena Symonds, Paul Spong, Friedrich Ladich	3710
Flying big brown bats emit a beam with two lobes in the vertical plane	Kaushik Ghose, Cynthia F. Moss, Timothy K. Horiuchi	3717
Short- and long-term changes in right whale calling behavior: The potential effects of noise on acoustic communication	Susan E. Parks, C. W. Clark, P. L. Tyack	3725
Application of metrics constructed from vibrotactile thresholds to the assessment of tactile sensory changes in the hands	A. J. Brammer, P. Sutinen, U. A. Diva, I. Pyykkö, E. Toppila, J. Starck	3732

JASA EXPRESS LETTERS

Analysis of handling noises on wound strings	J. Pakarinen, H. Penttinen, B. Bank	EL197
Effect of auditory-nerve response variability on estimates of tuning curves	Ananthakrishna Chintanpalli, Michael G. Heinz	EL203
Histology-based simulations for the ultrasonic detection of microscopic cancer <i>in vivo</i>	Timothy E. Doyle, Keith H. Warnick, Brent L. Carruth	EL210
Comments on the limp frame equivalent fluid model for porous media	Raymond Panneton	EL217
Spatial structure of low-frequency wind noise	D. Keith Wilson, Roy J. Greenfield, Michael J. White	EL223
Perceiving the speech of multiple concurrent talkers in a combined divided and selective attention task	Valeriy Shafiro, Brian Gygi	EL229

INDEX TO VOLUME 122

How to Use This Index	3747
Classification of Subjects	3747
Subject Index to Volume 122	3751
Author Index to Volume 122	3801

Analysis of handling noises on wound strings

J. Pakarinen, H. Penttinen, and B. Bank

Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, P.O. Box 3000,
FI-02015 TKK, Finland

jyri.pakarinen@tkk.fi, henri.penttinen@tkk.fi, bbank@acoustics.hut.fi

Abstract: This study analyzes the handling noises that occur when a finger is slid along a wound string. The resulting noise has a harmonic structure due to the periodic texture of the wound string. The frequency of the harmonics and the root-mean-square amplitude of the noise were found to be linearly proportional to the sliding speed. In addition, the sliding excites the longitudinal modes of the string, thus resulting in a set of static harmonics in the noise spectrum. The sliding excites different longitudinal modes depending on the sliding location.

© 2007 Acoustical Society of America

PACS numbers: 43.75.Gh, 43.40.Cw [TDR]

Date Received: April 13, 2007 **Date Accepted:** August 20, 2007

1. Introduction

Although the basic vibrational behavior of musical strings is extensively studied in the literature (see, for example, the book by Fletcher and Rossing¹), not much is known about the unintentional handling sounds a musician makes when playing a stringed instrument. Interestingly, musicians, especially in classical music, often tend to avoid making these sounds, but artificially removing them completely will make the music sound unrealistic. This is one reason why synthetic music can sound less lively, or more machine-like, than a real recording.

One approach to correct this machine-like quality in synthesizers is to record a sample library of different handling sounds and trigger a sample whenever a specific sound is needed, as done by Laurson *et al.*² On the other hand, a parametric model would provide a more flexible and memory-efficient way of implementing the handling sounds. More information concerning parametric or model-based sound synthesis can be found, for example, in the report by Välimäki *et al.*³ Obviously, before a parametric model for the handling noises can be constructed, an in-depth analysis of the noise type must be presented.

This study analyzes the handling noise generated by sliding the fingertip or nail along a wound string, thus producing a squeaky sound. This type of handling noise, usually called “fret noise” in the guitar terminology, can be heard often wherever a wound string is played with fingers. For the remainder of the article, this squeaky sound is simply referred to as handling noise.

This article is organized as follows: The measurement setup is described in detail in Sec. 2, and the general structure of the handling noise is explained in Sec. 3. A thorough analysis of the time-varying noise components is presented in Sec. 4, while the static noise components are analyzed in Sec. 5. Finally, conclusions are drawn in Sec. 6.

2. Measurement setup

The wound string handling noises on a steel-string acoustic guitar (Landola D-805E) were recorded in the small anechoic chamber at the Helsinki University of Technology. The guitar was placed in the normal playing position (in the player’s lap), and the sound was picked up by a microphone (AKG C 480 B, cardioid capsule) 50 cm away from the soundhole along the line normal to the soundboard. The signal was recorded digitally (44.1 kHz, 16 bits) with a sound card (Edirol UA-101) to the hard drive of a Macintosh laptop.

The handling noises were created by sliding the fingertip or nail along the wound 6th, 5th, and 4th strings. During the measurements, the slide was performed on one string at a time,

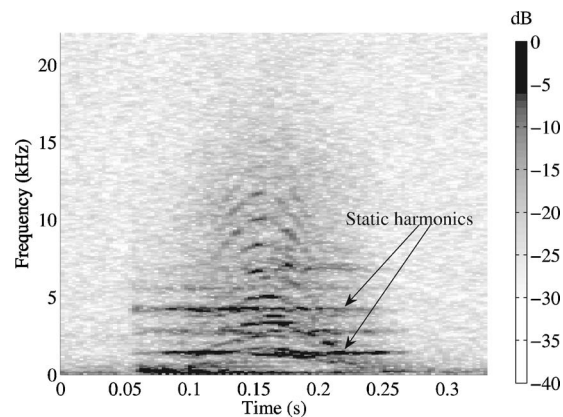


Fig. 1. A spectrogram of the handling noise created by sliding a finger on a wound guitar string. A 12 ms Hamming analysis window was used with 75% overlap. The noise clearly has a harmonic structure, where the frequency and amplitude of the harmonics increase with the sliding velocity. In addition to the moving harmonics, static harmonics (denoted by arrows) can be found in the spectrum.

and all other strings were damped with tape to prevent them from ringing. Although the analysis for only the 6th string is presented in the following, these results were found to apply to other wound strings as well.

3. General structure of the handling noise

The resulting spectrogram in Fig. 1 shows that the noise has a smooth lowpass character with a time-varying harmonic structure. The frequency of these harmonics depends on the slide velocity; a faster slide will lift the harmonics in frequency. This effect was found already earlier in a recent study.⁴ Also, the amplitude of the handling noise increases with slide velocity. It can be seen in Fig. 1, that the slide first increases and then decreases its velocity. This is natural, since the finger first accelerates and then decelerates when changing position on the string.

The presence of the time-varying harmonics can be explained by the surface texture of the wound string: each time the finger passes over a single winding turn, it will produce a noise pulse. Since the winding pattern is periodic, the resulting sound will be a train of noise pulses, and thus have a harmonic structure. Naturally, a faster slide will shorten the time interval between the noise pulses, thus raising the harmonics in frequency.

In addition to the moving harmonics, less intensive static harmonics can be found in the spectrum. These are illustrated with arrows in Fig. 1. The static harmonics are due to the longitudinal vibration of the string: the finger excites the longitudinal modes while scratching the surface of the string.

With this in mind, the handling noises can be thought of as consisting of an exciter and a resonator part. The excitation, created by the moving string-finger contact, is discussed in Sec. 4, while the resonator, consisting of the string vibrations, is discussed in Sec. 5.

4. Analysis of the time-varying finger-string excitation

4.1 Harmonic components

As explained above, an object (finger, nail, or plectrum) moving on a wound string creates a periodic pulse train, i.e., a velocity-dependent harmonic force excitation to the string. This force can be approximated as a periodic pulse train:

$$F(t) = \left[\sum_k \delta(t - t_k) \right] * f(t), \quad (1)$$

where t denotes time, δ is Dirac's delta function, t_k is the time instant of the k th pulse, and $f(t)$ is the impulse response of a single pulse that is generated when a finger slips from one winding.

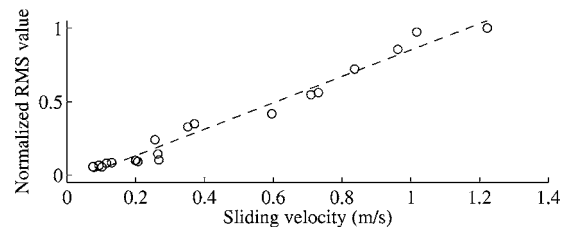


Fig. 2. Normalized rms noise value as a function of the sliding speed. The circles denote 20 different slide events performed on the fingernail. The dashed line represents a straight-line fit to the data, where the standard deviation of the residues is 0.053.

The operator $*$ denotes convolution. Thus, the excitation force presented in Eq. (1) can be interpreted as a periodic Dirac train filtered by the transfer function of a single pulse. When the sliding velocity is constant, t_k has the form

$$t_k = \frac{k}{d_w v_s}, \quad (2)$$

where d_w is the wound density (wounds per meter) and v_s is the sliding speed (meters per second).

When the sliding speed varies in time, the harmonic frequencies change. The relationship between the harmonic frequencies and the sliding speed can be given as $f_{Hn} = n v_s d_w$, where $n = 1, 2, 3, \dots$ is the mode number. This relation was found to coincide well with the results obtained by tracking the lowest harmonic frequency and the sliding speed on the recordings.

4.2 Handling noise amplitude

The amplitude of the handling noise as a function of the sliding velocity is illustrated in Fig. 2. The figure was obtained by evaluating 20 slide events, where the user slide his fingernail the distance of 23 cm on the surface of the string while attempting to maintain a constant sliding speed. The sliding velocity was approximated by dividing the distance (23 cm) by the duration of each slide event. The vertical axis in Fig. 2 denotes the root-mean-square (rms) value of the slide events, normalized to between zero and unity. It can be seen that the rms value of the handling noise is approximately linearly dependent on the sliding speed.

It must be noted that, due to the measurement setup, the finger speed could only be kept approximately constant during the slides. However, small variations in the sliding speed average out when the rms value is taken.

5. Analysis of the string resonances

Naturally, the force exerted by a sliding object to the string excites the transversal and longitudinal vibrational modes. The object can also, depending on its type, damp the string vibrations, especially in the transversal direction.

In other words, for rigid and relatively sharp objects, such as a plectrum, mode damping is minimal, and the transversal vibration plays a major part in the handling sound. The “pick scrape” effect in contemporary electric guitar playing is a good example of this; the player scrapes the plectrum against the string with a long movement, in order to produce a grinding sound with a changing pitch. The pitch change is caused by the change in the string’s length from the transversal vibration point of view: the moving plectrum acts as a rigid termination and divides the string into two segments. Since the magnetic pickup is located near one end of the string, it typically registers the vibration of only one of these segments.

On the other hand, a fingertip effectively attenuates the transversal vibrations, so the string vibrates mostly in the longitudinal direction. Since this type of handling noise generation is much more common, the longitudinal mode excitation is considered more thoroughly in the following.

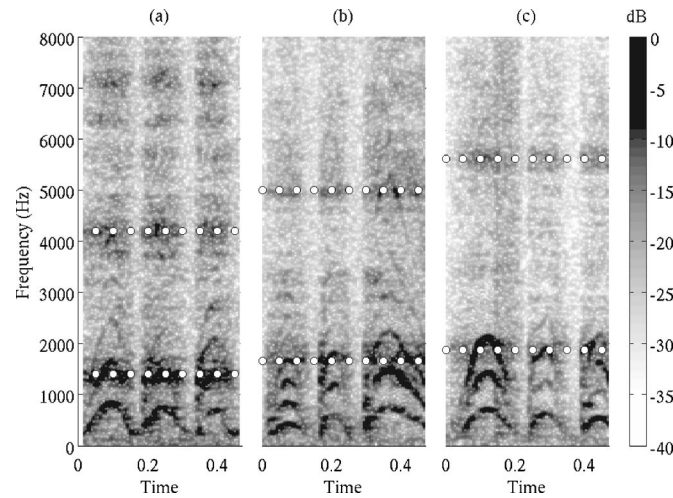


Fig. 3. Spectrograms of the scratch noise on the damped 6th string of an acoustic guitar. A 23 ms Hamming window with 75% overlap was used in analysis. The string was scratched with a fingertip, while (a) it had an open length. In (b) and (c), a capo was applied (b) at the third and (c) at the fifth fret. In (a), the dotted white line illustrates the frequency of the 1st and 3rd static harmonics (1400 and 4200 Hz, respectively). In (b) and (c) the white line denotes the frequency where the static harmonics should be located if they were a function of the length of the string. As can be seen, the static resonances coincide well with the dotted lines.

The partial differential equation for the longitudinal string vibration can be formulated as follows:⁵

$$\mu \frac{\partial^2 \xi}{\partial t^2} = ES \frac{\partial^2 \xi}{\partial x^2} - 2R(f)\mu \frac{\partial \xi}{\partial t} + d(x,t), \tag{3}$$

where $\xi(x, t)$ is the longitudinal displacement of the string, E is Young’s modulus, S is the cross-section area of the string, and μ is the linear mass density. The propagation speed is $c_L = \sqrt{ES/\mu} = \sqrt{E\rho}$, where ρ is the density of the material. Thus, the propagation speed is constant for a given material and does not depend on string tension, unlike in the case of the transverse vibration. The function $R(f)$ is the frequency-dependent frictional resistance. The excitation force density is denoted by $d(x, t)$. When the string is excited at one point x_{exc} , the spatial distribution of the force can be approximated by a Dirac function: $d(x, t) = \delta(x - x_{exc})F(t)$. The force acting on the bridge, F_b , can be computed as the tension variation at the bridge termination, that is $F_b = ES(\partial \xi / \partial x)|_{x=0}$.

For a given excitation force $F(t)$ at the position x_{exc} , the bridge force can be approximately computed as follows:⁶

$$F_b(t) = \frac{ES}{\mu L^2} \sum_{k=1}^{\infty} \left\{ \frac{k}{f_k} e^{-tR(f_k)} \sin(2\pi f_k t) \right\} * \left\{ \sin\left(\frac{k\pi x_{exc}}{L}\right) F(t) \right\}. \tag{4}$$

The longitudinal modal frequencies $f_k = kc_L / (2L)$ depend on the propagation speed c_L and string length L . In Eq. (4) it can be seen that the force signal excites a set of parallel resonances and that the excitation amplitudes depend on x_{exc} . As a special case, those modes that have a node at the excitation point will not be present. In addition to eliminating some harmonics, the excitation position x_{exc} has a strong influence on the general shape of the spectrum. For $x_{exc} \approx 0$ or $x_{exc} \approx L$ the first few longitudinal modes are only weakly excited.

The assumption that the static components are originating from the longitudinal vibration is confirmed by measurements. It can be seen in Fig. 1 that the static components have a

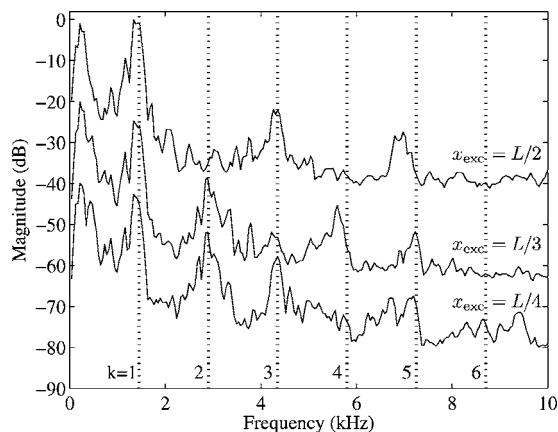


Fig. 4. The averaged power spectrum of the handling noise when the 6th string is scratched at 1/2 (top pane), at 1/3rd (middle pane), and at 1/4th (bottom pane) of the string's length. A magnitude offset of 20 dB was applied to the spectra for clarity. An averaging window of 23 ms was used. The k values and the vertical dotted lines denote the longitudinal mode numbers and locations, respectively.

clear harmonic structure and are around those frequencies where the longitudinal modes are expected when calculated from the physical parameters of the string. Moreover, the frequencies do not change as a function of tension. Instead, they are inversely proportional to string length, being in good agreement with theory. This is shown in Fig. 3. When the string is excited at different positions, the shape of the spectrum changes considerably. Figure 4 shows special cases of exciting the string at half, one third, and one fourth of its length, and leading to the expected result that every second, third, or fourth harmonic is missing from the power spectrum. The time-varying contact noise caused by the string excitation is averaged mostly at low frequencies. This explains the spectral peaks below the first longitudinal modes in Fig. 4.

6. Conclusions and discussion

The handling noises created by sliding a finger on a wound string were analyzed. The resulting noise can be interpreted to be a result of an exciter-resonator system. The moving finger-string contact forms the exciter part: a lowpass-type noise with a clear harmonic structure is created when the finger rubs against the string windings. The harmonic frequencies and the rms amplitude of the handling noise were found to be linearly dependent on the slide velocity.

The resonator part consists of the vibrational behavior of the string. The sliding contact excites mainly the longitudinal string modes, since transversal vibration is effectively damped by the soft finger-string contact itself. Naturally, different longitudinal modes are excited depending on the finger location. When a harder object, such as a plectrum or bottleneck, is used in the slide, transversal modes will be more significant. Sound examples of different handling noises are available in the companion web page of this article: <http://www.acoustics.hut.fi/publications/papers/jasa-el-handling-noise>.

A simple signal-based model for synthesizing contact sounds on wound strings has been introduced in an earlier study.⁴ However, based on the knowledge obtained in this article, a more sophisticated physics-based model for synthesizing wound string contact noises will be presented in an upcoming article on slide guitar synthesis.⁷ There, the contact noise between a wound string and the slide tube is synthesized with a noise pulse train-based parametric model. Transversal string vibration is implemented with digital waveguides, and the longitudinal modes are created with static resonators.

Acknowledgments

The work of J. Pakarinen is funded by the GETA graduate school, whereas the work of B. Bank is funded by the EC 6th Framework Program Marie Curie Intra-European Fellowships Contract

No. 041924, “Nonlinear Effects in String Instruments: Perception and Modeling.” The authors wish to thank Professor Vesa Välimäki and Dr. Tom Bäckström for helpful discussions.

References and links:

- ¹N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer-Verlag, New York, 1988).
- ²M. Laurson, C. Erkut, V. Välimäki, and M. Kuuskankare, “Methods for modeling realistic playing in acoustic guitar synthesis,” *Comput. Music J.* **25**, 38–49 (2001), URL <http://lib.tkk.fi/Diss/2002/isbn9512261901/>
- ³V. Välimäki, J. Pakarinen, C. Erkut, and M. Karjalainen, “Discrete-time modeling of musical instruments,” *Rep. Prog. Phys.* **69**, 1–78 (2006).
- ⁴H. Penttinen, J. Pakarinen, V. Välimäki, M. Laurson, H. Li, and M. Leman, “Model-based sound synthesis of the guqin,” *J. Acoust. Soc. Am.* **120**, 4052–4063 (2006).
- ⁵B. Bank, *Physics-Based Sound Synthesis of String Instruments Including Geometric Nonlinearities* (Budapest, Hungary, 2006), Chap. 5.3.2, p. 77; available online at <http://www.mit.bme.hu/~bank/phd>, (Budapest University of Technology and Economics).
- ⁶P. M. Morse, *Vibration and Sound*, reprint of 2nd ed. (1st ed. 1936) (American Institute of Physics, New York, 1983), Chap. III.10, p. 106.
- ⁷J. Pakarinen, T. Puputti, and V. Välimäki, “Virtual slide guitar,” *Computer Music Journal* (2007).

Effect of auditory-nerve response variability on estimates of tuning curves

Ananthakrishna Chintanpalli

*Weldon School of Biomedical Engineering, Purdue University, 206 S. Martin Jischke Drive, West Lafayette, Indiana 47907
cananthk@purdue.edu*

Michael G. Heinz

*Department of Speech, Language and Hearing Sciences and Weldon School of Biomedical Engineering, Purdue University, 500 Oval Drive, West Lafayette, Indiana 47907
mheinz@purdue.edu*

Abstract: Near-Poisson variability in auditory-nerve (AN) responses limits the accuracy of automated tuning-curve algorithms. Here, a typical adaptive tuning-curve algorithm was used with a physiologically realistic AN model with and without the inclusion of neural randomness. Response randomness produced variability in Q_{10} estimates that was nearly as large as in AN data. Results suggest that it is sufficient for AN models to specify frequency selectivity based on mean Q_{10} values at each characteristic frequency (CF). Errors in estimates of CF, which decreased from ± 0.2 octaves at low frequencies to ± 0.05 octaves at high frequencies, are significant for studies of spatiotemporal coding.

© 2007 Acoustical Society of America

PACS numbers: 43.64.Bt, 43.64.Pg [BLM]

Date Received: August 9, 2007 **Date Accepted:** September 14, 2007

1. Introduction

Automated tuning-curve algorithms have been used for many years to characterize the tuning properties of auditory-nerve (AN) fibers (Lieberman, 1978; Miller *et al.*, 1997; Heinz and Young, 2004). The sound pressure level necessary to increase the AN response to a 50-ms tone by a small criteria (typically 1 spike) is determined as a function of tone frequency in order to map the threshold tuning curve. The near-Poisson variability of AN responses limits the accuracy with which individual thresholds can be estimated quickly. For example, empirical data suggest that a standard deviation of 1.25 spikes is expected within a 50-ms window for a fiber with an overall discharge rate of 50 spikes/s (Young and Barta, 1986; Winter and Palmer, 1991). Thus, a typical high spontaneous-rate (SR) fiber has a standard deviation that is larger than the typical criterion used to estimate threshold. Adaptive tracking procedures are used as an efficient method to estimate thresholds given the random nature of AN responses. However, the relatively short holding time for AN-fiber recordings requires a compromise between an adaptive algorithm with great accuracy and one that is quick enough to allow other stimuli of interest to be presented.

Given the necessarily quick nature of the automated algorithms, there is some amount of variability in parameters that are estimated from tuning curves. Characteristic frequency [(CF) the frequency to which a fiber responds to the lowest sound level], fiber threshold (the threshold sound level at CF), and Q_{10} (a measure of sharpness of tuning equal to the ratio of CF to the bandwidth 10 dB above threshold) are typical parameters derived from tuning curves. Population plots of thresholds and Q_{10} values derived from tuning curves typically show a range of values at each CF (e.g., Lieberman, 1978; Miller *et al.*, 1997). Although much of the variability in threshold arises from the dependence of threshold on SR, thresholds for individual fibers have been shown to differ from thresholds estimated from rate-level functions by ± 10 dB (Lieberman, 1978). The observed variability in Q_{10} values has provided a challenge to AN modeling studies because AN fibers with similar CF and different Q_{10} values would be expected to

show somewhat different responses to complex stimuli, such as vowels (Bruce *et al.*, 2003). However, it is possible that much of the Q_{10} variation observed in the data may simply be due to variability in Q_{10} estimates derived from automated tuning-curve algorithms. Finally, although it is not apparent in typical AN-fiber population plots, there is also likely to be variability in estimates of CF. Although such variability may be insignificant for single-fiber response properties, there has been a long-standing interest in spatiotemporal coding mechanisms that rely on the relative temporal patterns across AN fibers with slightly different CFs (Shamma, 1985; Deng and Geisler, 1987; Heinz *et al.*, 2001; Carney *et al.*, 2002; Cedolin and Delgutte, 2007; Heinz, 2007). Such across-CF mechanisms have a strong parametric dependence on the CF difference between AN fibers and this parameter will have a variability that is roughly twice as large as the variability in individual CF estimates. Thus, it is also important to examine the precision with which CF is estimated from typical automated tuning-curve algorithms.

2. Methods

2.1 Auditory-nerve model

The AN model described by Zilany and Bruce (2006) was used in this study. This phenomenological AN model represents an extension of several previous versions of the AN model that have been extensively tested against physiological responses to both simple and complex stimuli, including tones, two-tone complexes, broadband noise, and speech (Zhang *et al.*, 2001; Bruce *et al.*, 2003). Important for the purposes of the present study, the model has been fit well to tuning curves and to the statistical properties of AN responses. Tuning-curve parameters derived from the model were shown to match very well with those derived from neurophysiological data collected from cats (Miller *et al.*, 1997). Model thresholds match well across CF to the lowest thresholds from the cat data, whereas Q_{10} values derived from model tuning curves fit well with the 50th percentile of the Q_{10} data from cats as a function of CF. The statistics of AN responses are represented in the model by using the time-varying discharge rate waveform as the driving function to a nonhomogeneous Poisson process, which has been modified to include both absolute and relative refractory effects. Although this Poisson-based model does not capture all of the detailed stochastic properties of AN fiber activity (e.g., Heil *et al.*, 2007), the main statistical properties that are most relevant to the present study are well represented by this model (e.g., Young and Barta, 1986).

In the present study, tuning curves were measured based on outputs at two different model stages to evaluate the effect of randomness on tuning-curve estimates. To include the neural variability associated with AN responses, tuning curves were first measured based on the spike outputs of the model. To remove the effects of randomness (and refractoriness), model tuning curves were also measured based on the time-varying discharge rate waveforms prior to the generation of spikes. Note that this version of the model contains all other response properties, including the adaptation associated with synaptic transmission. Because of the adaptive nature of the tuning-curve algorithm, separate tuning curves had to be derived for each of these cases. All model simulations are for high-SR fibers and are based on the SR value (50 spikes/s) for which this AN model was designed and tested (Zilany and Bruce, 2006).

2.2 Automated tuning-curve algorithm

The adaptive algorithm used here has been used in numerous experimental studies to measure frequency threshold tuning curves of AN fibers (e.g., Liberman, 1978; Miller *et al.*, 1997; Heinz and Young, 2004). This algorithm, which is described in detail by Liberman (1978), represents an efficient method to track the threshold sound level necessary to elicit a small increase in AN-fiber response as a function of frequency. At each frequency, the number of spikes occurring in the final 50 ms of a 60-ms tone (with 5-ms rise/fall ramps) is compared to the number of spikes occurring in the final 50 ms of the 60-ms window following the offset of the tone. If the difference is larger than a specified criterion (0 spikes here), then the sound level is decreased by one step (2 dB), otherwise the sound level is raised by two steps (4 dB). Tone bursts are presented every 120 ms and the sound level is tracked adaptively until a threshold is determined for

the current frequency. Threshold is determined when a specified pattern of tracked sound levels is observed for the current frequency; specifically the current sound level must be (1) equal to the sound level three and six steps previous in the track and (2) a decrease from the previous level. This pattern indicates that the current sound level is a reasonable estimate of the threshold level needed to increase the rate by 0–20 spikes/s based on a criterion of 0 spikes (Lieberman, 1978). Between 20 and 35 frequencies per octave are used, depending on the expected CF. The three parameters CF, threshold, and Q_{10} are estimated from each tuning curve, following five-point triangular smoothing (consistent with Miller *et al.*, 1997, see their Fig. 8).

2.3 Neurophysiological recordings

Repeated measures of tuning curves were collected for three AN fibers during two experiments that were part of other studies in the lab. Single-unit AN responses were measured from chin-chillas, weighing 400–600 g, using standard techniques (Heinz and Young, 2004) that were approved by the Purdue Animal Use and Care Committee. Animals were anesthetized initially with xylazine (0.5–1 mg/kg im) and ketamine (40–50 mg/kg im). Supplemental doses of sodium pentobarbital (~10 mg/h iv) were given to maintain an areflexic state of anesthesia throughout the experiment. A tracheotomy was performed to allow a low-resistance airway. The bulla was vented with PE tubing to equalize the middle-ear pressure. The animal's rectal temperature was maintained near 37.5 °C with a feedback heating pad. A craniotomy was made in the posterior fossa to expose the cerebellum that overlies the auditory-nerve/cochlear-nucleus complex. AN-fiber recordings were made by inserting a 15–30 M Ω glass micropipette filled with 3M NaCl into the AN under visual control after aspiration of the overlying cerebellum. Recordings were made in an electrically shielded, double-walled sound-proofed room, and computer-controlled stimuli were presented via a calibrated closed-field acoustic system. Isolated fibers were characterized by the automated tuning-curve algorithm described earlier. Spontaneous rate was estimated from the off-time of a CF-tone rate-level function. Between 5 and 11 repeated tuning curves were recorded from each AN fiber depending on the holding time.

3. Results

Figure 1 illustrates the effect of AN response variability on individual tuning curves estimated from a single model AN fiber. Figure 1(A) shows the resulting tuning curve as measured by the adaptive algorithm when the effects of randomness were excluded from the AN model. For comparison, Fig. 1(B) shows a tuning curve measured when the effects of randomness were included in the AN model. Note that the individual data points, which represent the thresholds obtained at each frequency from the adaptive algorithm, are extremely orderly in the non-random case and are quite variable when randomness was included in the model. The smoothed curve (as typically used with experimental data), eliminates much of the variability in the random case; however, some differences are still apparent. Figure 1(C) shows five repeated tuning curves using the same AN model CF with randomness included and illustrates the extent of variability across repeated measures. The derived parameter estimates from the five repeated measures of the tuning curve are listed in the figure caption. The CFs ranged from 1.97 to 2.08 kHz (a 0.08-octave range), thresholds ranged from 0.7 to 2.7 dB SPL, and Q_{10} values ranged from 3.0 to 4.6. Thus, even with the typical smoothing applied to the tuning curve data, variability remains in each of the parameters. Note that the range of estimated thresholds was 6–8 dB higher in the random case than the nonrandom case, whereas the range of Q_{10} and CF estimates with randomness surrounded the values obtained without randomness.

The effects of randomness on population estimates of threshold and Q_{10} values are illustrated in Figs. 2(A) and 2(B). Tuning curves were measured from model AN fibers with nominal CFs spaced logarithmically from 0.3 to 10 kHz. Red lines in Figs. 2(A) and 2(B) shows results when randomness was excluded from the AN model and blue symbols show derived parameters when randomness was included. In the nonrandom case [red lines, Fig. 2(A)], model thresholds were consistent with those reported by Zilany and Bruce (2006), ranging

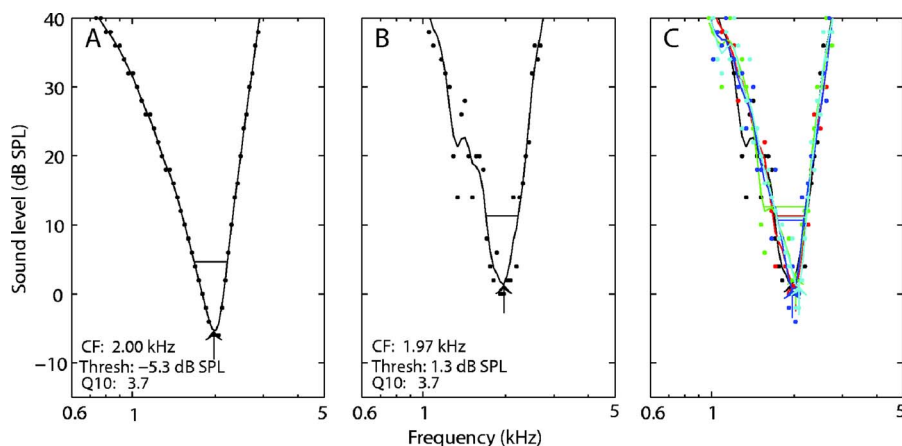


Fig. 1. (Color online) Tuning curves measured by the adaptive algorithm with and without randomness for an AN-model fiber with CF=2.0 kHz and SR=50 spikes/s. (A) AN model without randomness. Individual data points represent the thresholds estimated by the adaptive algorithm at each frequency. Solid lines represent a smoothed curve (see the text). Identical tuning curves were obtained with repeated measures. (B) AN model with randomness. (C) Five repeated tuning curves with randomness. Derived parameters for the individual repetitions are: CF (kHz) = 1.97, 2.02, 2.02, 1.97, 2.08; geometric mean=2.01; and range=0.08 octaves. Threshold (dB SPL) = 1.3, 1.3, 2.7, 0.7, 1.1; mean=1.4; range=2.0 dB. Q_{10} =3.7, 3.7, 3.0, 4.6, 4.2; geo. mean=3.8; range=0.62 octaves. The information may not be properly conveyed in black and white.

down to -5 dB SPL between 1 and 2 kHz and showing an increase at low frequencies and near 4 kHz based on the cat middle-ear filter. The inclusion of randomness [symbols, Fig. 2(A)] produced thresholds that were consistently higher than the nonrandom case and that spanned a range of about 10 dB in each CF region. This variation was not seen in the non-random case. The derived Q_{10} values in the nonrandom case [red lines, Fig. 2(B)] showed an increasing trend with CF that matched the mean of the physiological data (Miller *et al.*, 1997), but showed little variability in each CF region (except for the 4–5 kHz region corresponding to the middle-ear notch). When randomness was included [symbols, Fig. 2(B)], an increasing trend across CF was observed with large variability within each CF region. Predicted Q_{10} values spanned much of the 5th–95th percentile range (dashed lines) of Q_{10} values measured from cats (Miller *et al.*, 1997). Note that the Q_{10} variability spans above and below the nonrandom Q_{10} values, unlike the threshold estimates that were consistently above the nonrandom values.

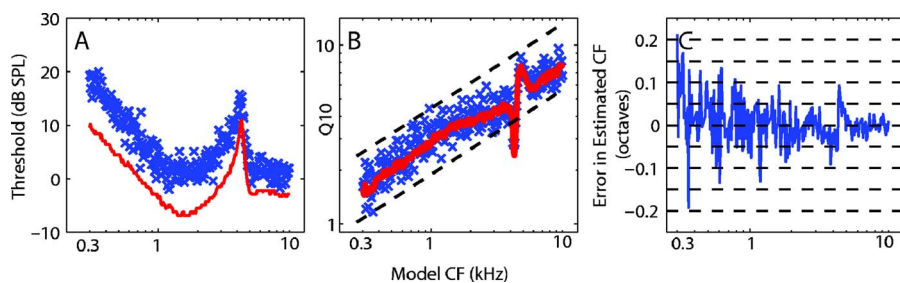


Fig. 2. (Color online) (A) Populations of thresholds, (B) Q_{10} values, and (C) CF errors estimated from tuning curves measured for 250 model AN fibers. (A and B) Each symbol (\times) represents the estimated value with randomness and red lines represent nonrandom estimates. (B) Black dashed lines represent the range between the 5th and 95th percentile of experimental Q_{10} values within each CF region (Miller *et al.*, 1997). (C) Error in estimated CFs from tuning curves measured when randomness was included. CF error in octaves is computed as \log_2 (estimated CF/model CF). The information in (A) and (B) may not be properly conveyed in black and white.

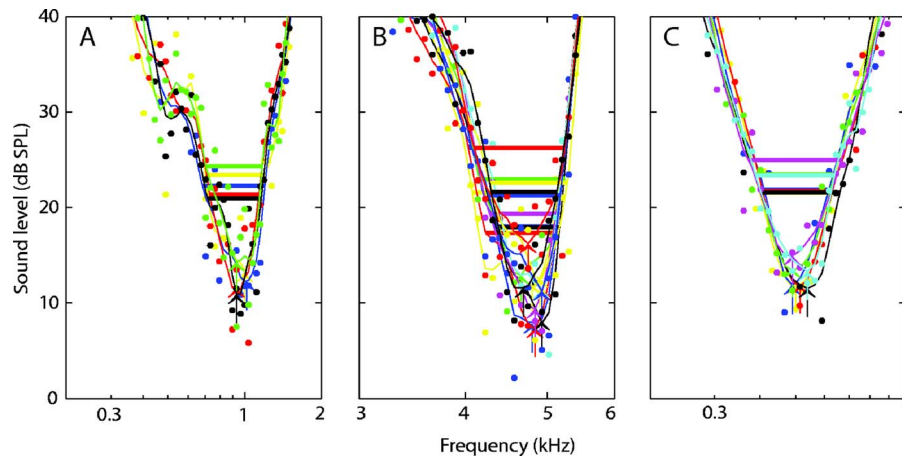


Fig. 3. (Color online) Repeated tuning curves measured from three AN fibers from chinchillas, which vary in SR and CF. Derived parameters for the individual repetitions are: (A: SR=58.8 spikes/s) CF (kHz)=1.00, 1.00, 0.96, 0.96, 0.93; geometric mean=0.97; range=0.11 octaves. Threshold (dB SPL)=13.4, 12.3, 11.4, 10.9, 14.3; mean=12.5; range=3.4. Q_{10} =2.3, 1.9, 2.4, 2.2, 2.0; geo. mean=2.2; range=0.34 octs. (B: SR=54.6) CF =4.75, 4.75, 4.84, 4.93, 4.66, 4.84, 4.93, 4.93, 4.93, 4.75, 4.66; geo. mean=4.78; range=0.08 octs. Threshold =11.5, 8.0, 7.3, 7.9, 12.9, 9.3, 11.5, 12.6, 11.2, 16.2, 11.6; mean=9.5; range=5.6. Q_{10} =4.4, 5.9, 5.5, 6.0, 4.5, 5.7, 6.1, 4.9, 5.4, 4.1, 5.6; geo. mean=5.2; range=0.45 octs. (C: SR=3.7) CF=0.50, 0.49, 0.51, 0.54, 0.49, 0.49, 0.54; geo. mean=0.51; range=0.14 octs. Threshold=11.5, 11.9, 11.7, 11.5, 13.4, 15.0, 13.3; mean=12.0; range=1.9. Q_{10} =2.2, 2.4, 2.3, 2.1, 2.0, 1.7, 2.2; geo. mean=2.2; range=0.26 octs. The information may not be properly conveyed in black and white.

The effect of neural randomness on the variability in tuning-curve estimates of CF is shown in Fig. 2(C). The error in the CF estimates for the 250 tuning curves that were included in Figs. 2(A) and 2(B) is plotted as a function of the model CF. Although the errors are centered around zero, i.e., tuning-curve estimates are essentially unbiased, there is a range of errors at all model CFs. The largest errors are observed at the lowest CFs (0.3–0.4 kHz), where errors fall within a ± 0.2 -octave range. For mid CF values (0.7–2.0 kHz), errors generally fall within a ± 0.1 -octave range. At the highest CFs, errors are restricted to within a ± 0.05 -octave range, with the exception of the region surrounding 4 kHz. The cat middle-ear transfer function has a sharp notch in gain (of almost 20 dB) centered between 4 and 5 kHz (Bruce *et al.*, 2003), which creates a sharp increase in CF thresholds near 4 kHz [Fig. 2(A)]. This sharp gain reduction leads to slightly biased CF estimates and larger errors at model CFs surrounding the notch. Note that the bias is in opposite directions above and below the center frequency of the notch, with CFs below the notch having CF estimates that are biased downwards and CFs above the notch having an upward bias. Overall, the observation of larger octave errors at lower CFs is consistent with shallower tuning curves (lower Q_{10} values) having a less-well defined CF, for which alternative estimates of CF based on spike timing have been proposed (Joris *et al.*, 2006).

The general model predictions were tested by measuring repeated tuning curves for several AN fibers recorded from chinchillas. Figure 3 shows the variability across repeated tuning curves for three AN fibers ranging in CF and SR. The variability in estimated tuning-curve parameters for each fiber is listed in the figure caption, along with the SR values. The qualitative variability in tuning-curve shape was larger for the high-SR fibers [Figs. 3(A) and 3(B)] as compared to the fiber with lower SR [Fig. 3(C)]. This resulted in less variability in threshold and Q_{10} estimates for the lower-SR fiber. However, the range of CF estimates was largest for the lower-SR fiber (0.14 octaves) because of its low CF, consistent with the model predictions shown in Fig. 2(C). The ranges of CF estimates for the three AN fibers (0.08–0.14 octaves) are quantitatively consistent with the modeling predictions shown in Fig. 2(C).

4. Discussion

The modeling and experimental results presented here are based on one specific automated tuning-curve algorithm; however, the effects shown are likely to play a significant role in the variability observed in tuning-curve estimates derived from other techniques as well, including nonadaptive methods (e.g., Kiang *et al.*, 1965; Evans, 1972). The modeling results shown in Fig. 2(B) suggest that the variability observed in Q_{10} values derived from experimental AN data is largely due to the effects of response variability on the precision of Q_{10} estimates derived from tuning curves. The Q_{10} values derived from the AN model without randomness fell in the middle of the range of Q_{10} values derived from the model with randomness. This results suggests that it is sufficient for AN models to fit frequency selectivity based on the 50th percentile of Q_{10} values at each CF (Bruce *et al.*, 2003; Zilany and Bruce, 2006). It does not appear to be necessary to assume that different AN fibers with similar CFs have inherently different frequency selectivity. The finding that the range of threshold estimates was above the nonrandom threshold, while the variability in Q_{10} estimates ranged above and below the nonrandom case supports the typical interpretation of such data (e.g., Miller *et al.*, 1997; Bruce *et al.*, 2003).

The range of errors in CF estimates was shown to decrease from ± 0.2 octaves at low frequencies to ± 0.05 octaves at high frequencies. These errors appear to be quite small, consistent with the qualitative impression that the automated algorithm provides a reasonably consistent estimate of frequency selectivity across repeated tuning-curve measurements [Figs. 1(C) and 3(A)–3(C)]. This degree of variability in CF estimates may be adequate for most population studies, including those that quantify rate-place coding (e.g., Sachs and Young, 1979) and those that quantify temporal-place coding in terms of metrics that depend on the average magnitude of phase locking within a range of CFs [e.g., the average localized synchronized rate (ALSR) (Young and Sachs, 1979)]. However, the variability in CF estimates is significant for studies of spatiotemporal coding that depend on the relative phase response across nearby CFs (e.g., Shamma, 1985; Palmer, 1990). This is because narrow cochlear filters have sharp phase transitions near CF that translate a small error in CF into a large phase shift. For example, two AN fibers with CFs surrounding (± 0.05 octaves) a vowel formant at 1 kHz are predicted to have a 1/4-cycle phase difference between their phase-locked responses (Heinz, 2007). The model predictions in Fig. 3 suggest that CF estimates near 1 kHz could be in error by up to ± 0.1 octaves, implying that the estimate of CF difference could be in error by up to ± 0.2 octaves. Given the sharp phase transition, a 0.4-octave range in CF difference corresponds to a $\pm 1/2$ -cycle range in phase shift. The large effect of small CF errors on spatiotemporal response patterns is likely to be why it has been hard to study spatiotemporal cues quantitatively with experimental data (e.g., Palmer, 1990), and why many spatiotemporal studies have used models where the CF difference is controllable (Deng and Geisler, 1987; Heinz *et al.*, 2001; Carney *et al.*, 2002). The limitations imposed by the variability in CF estimates for spatiotemporal coding studies can be overcome by techniques that predict the response of a population of AN fibers with similar CFs responding to a single stimulus from the response of a single AN fiber responding to frequency-shifted stimuli (Cedolin and Delgutte, 2007; Heinz, 2007).

Acknowledgments

The authors thank Ian Bruce for initial discussions and suggestions related to evaluating the variability in experimental Q_{10} values. Thanks are also given to Ian Bruce, Sushrut Kale, and Jayaganesh Swaminathan for providing valuable comments on a previous version of the manuscript. Special thanks are given to Sushrut Kale for his help with collection of the neurophysiological AN data. This research was supported by NIH-NIDCD.

References and links

- Bruce, I. C., Sachs, M. B., and Young, E. D. (2003). "An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses," *J. Acoust. Soc. Am.* **113**, 369–388.
- Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S. (2002). "Auditory phase opponency: A temporal model for masked detection at low frequencies," *Acust. Acta Acust.* **88**, 334–347.
- Cedolin, L., and Delgutte, B. (2007). "Spatio-temporal representation of the pitch of complex tones in the auditory nerve," 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, Berlin), pp. 61–70.
- Deng, L., and Geisler, C. D. (1987). “A composite auditory model for processing speech sounds,” *J. Acoust. Soc. Am.* **82**, 2001–2012.
- Evans, E. F. (1972). “The frequency response and other properties of single fibres in the guinea-pig cochlear nerve,” *J. Physiol.* **226**, 263–287.
- Heil, P., Neubauer, H., Irvine, D. R., and Brown, M. (2007). “Spontaneous activity of auditory-nerve fibers: Insights into stochastic processes at ribbon synapses,” *J. Neurosci.* **27**, 8457–8474.
- Heinz, M. G. (2007). “Spatiotemporal encoding of vowels in noise studied with the responses of individual auditory nerve fibers,” 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), pp. 107–115.
- Heinz, M. G., Colburn, H. S., and Carney, L. H. (2001). “Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection,” *J. Acoust. Soc. Am.* **110**, 2065–2084.
- Heinz, M. G., and Young, E. D. (2004). “Response growth with sound level in auditory-nerve fibers after noise-induced hearing loss,” *J. Neurophysiol.* **91**, 784–795.
- 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.
- Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (1965). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT Press, Cambridge, MA).
- Liberman, M. C. (1978). “Auditory-nerve response from cats raised in a low-noise chamber,” *J. Acoust. Soc. Am.* **63**, 442–455.
- Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (1997). “Effects of acoustic trauma on the representation of the vowel /e/ in cat auditory nerve fibers,” *J. Acoust. Soc. Am.* **101**, 3602–3616.
- Palmer, A. R. (1990). “The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers,” *J. Acoust. Soc. Am.* **88**, 1412–1426.
- Sachs, M. B., and Young, E. D. (1979). “Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate,” *J. Acoust. Soc. Am.* **66**, 470–479.
- Shamma, S. A. (1985). “Speech processing in the auditory system. I: The representation of speech sounds in the responses of the auditory nerve,” *J. Acoust. Soc. Am.* **78**, 1612–1621.
- Winter, I. M., and Palmer, A. R. (1991). “Intensity coding in low-frequency auditory-nerve fibers of the guinea pig,” *J. Acoust. Soc. Am.* **90**, 1958–1967.
- Young, E. D., and Barta, P. E. (1986). “Rate responses of auditory nerve fibers to tones in noise near masked threshold,” *J. Acoust. Soc. Am.* **79**, 426–442.
- Young, E. D., and Sachs, M. B. (1979). “Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers,” *J. Acoust. Soc. Am.* **66**, 1381–1403.
- Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (2001). “A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression,” *J. Acoust. Soc. Am.* **109**, 648–670.
- Zilany, M. S. A., and Bruce, I. C. (2006). “Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery,” *J. Acoust. Soc. Am.* **120**, 1446–1466.

Histology-based simulations for the ultrasonic detection of microscopic cancer *in vivo*

Timothy E. Doyle, Keith H. Warnick, and Brent L. Carruth

Department of Physics, Utah State University, Logan, Utah 84322-4415

tim.doyle@usu.edu, khwarnick@cc.usu.edu, carruth@cc.usu.edu

Abstract: Ultrasonic spectroscopy may offer an alternative to imaging methods for the *in vivo* detection of microscopic cancer. To investigate this potential, a numerical model that incorporates multiple scattering, wave-mode conversion, and hierarchical microstructures was developed to simulate ultrasonic interactions in biological tissue at the microscopic level. Simulated high-frequency (20–75 MHz) spectra of up to 2137 cells displayed significant correlations to nucleus diameter and malignant cell infiltration, and indicated as few as 300 malignant cells may be detectable in normal tissue. The results suggest that ultrasonic spectroscopy combined with simulation-based interpretive models may provide real-time histopathology during surgeries, biopsies, and endoscopies.

© 2007 Acoustical Society of America

PACS numbers: 43.80.Cs, 43.80.Jz, 43.80.Qf [CC]

Date Received: August 31, 2007 **Date Accepted:** September 24, 2007

The detection of microscopic cancer *in vivo* holds substantial promise for improving the early diagnosis, treatment, and monitoring of the disease. Current research into *in vivo* detection methods includes optical, terahertz wave (T-wave), and nuclear imaging methods. Many of these methods rely on chemical mechanisms for identifying cancer cells such as molecular absorption, molecular scattering, or molecular imaging agents. *In vivo* detection methods for resolving cancer-induced alterations to cell and tissue structures are also needed, however, to obtain vital information on staging (invasion) and grading (differentiation status). To date, these methods have focused on imaging approaches such as optical coherence tomography and confocal laser scanning microscopy.¹

Ultrasonic spectroscopy has the potential to reveal detailed histological information on tissues, and may offer a viable alternative to imaging methods for the *in vivo* characterization of malignant cell and tissue structures. Extensive experimental data show tissue modifications due to breast, prostate, and other cancers affect ultrasonic backscatter measurements, and both empirical and analytical models have been developed to estimate tissue parameters such as cell sizes and distributions.^{2–7} Empirical models statistically define spectral parameters from tissue measurements, and studies have shown excellent agreement for homogeneous tissues but discrepancies for heterogeneous tissues.⁷ Analytical models average the scattering from uniformly distributed cells or cell nuclei, and employ simplifications such as single scattering, weak scattering (Born approximation), and fluidlike tissue properties (no shear waves).⁵

Biological tissues are heterogeneous at multiple scales. Microscopic cancer increases tissue heterogeneity by introducing cell mutations that have different cellular structures and physical properties, that are mixed with normal cells to various degrees, and that modify the tissue organization. The enormous variety of morphologies that cancer exhibits renders the development of empirical or analytical models to characterize microscopic cancer extremely challenging. A first-principles, direct-simulation approach to quantitatively determine the effects of tissue structure on ultrasonic spectra may therefore offer a more efficient, versatile, and accurate methodology for *in vivo* cancer detection.

To explore this possibility, a multiscale numerical model was developed to simulate ultrasonic wave propagation in heterogeneous tissues and to model the effects of cancer-induced tissue modifications on ultrasonic measurements. A simulation approach is attractive for mod-

Table 1. Tissue and cell properties used for simulations.

	Longitudinal wave speed (m/s)	Shear wave speed (m/s)	Density (g/cm ³)
Extracellular matrix	1570	106	1.06
Cytoplasm	1483	0–50	0.998
Nucleus	1509	0–50	1.43

eling the infiltration of normal tissue by cancerous cells, nuclear pleomorphism, and other microscopic structural changes associated with cancer since it has the ability to include heterogeneous distributions of scatterers (cells) with specific histological features and internal cell structures. A simulation approach can also include hierarchical tissue structures, multiple scattering, and wave-mode conversion, which have been largely ignored by previous methods. Although high-frequency shear waves have been considered insignificant in tissues due to high attenuation, mode conversion of longitudinal waves to shear waves at the cell membrane or nuclear envelope may be an important loss mechanism for ultrasonic energy. Additionally, the role of multiple scattering in tissues needs to be rigorously modeled and tested before being assumed negligible, particularly for cancerous tissues where the cell packings, cell densities, and nuclear diameters are typically higher than for normal tissues.

The numerical model for this study was based on an iterative multipole method that simulates the multiple scattering of elastic waves in arbitrary, three-dimensional agglomerations of spherical particles.⁸ The method was modified to incorporate scattering from inclusions (nuclei) within the particles (cells). A concentric core-shell structure embedded in a medium represented the cell nucleus, cell cytoplasm, and extracellular matrix, respectively. Each of the components could be assigned fluid, solid, or viscoelastic properties. Vector multipole expansions simulated both longitudinal and shear waves in the model. The model calculated the multiple scattering interactions both within and between the cells using boundary conditions, addition theorems, and iteration. The concentric core-shell approach was verified by testing the model under two conditions that reduced the solutions to single-sphere scattering: (1) equal matrix and shell properties and (2) equal shell and core properties. Computation times were approximately 24 h for simulations of roughly 2000 cells on a single-CPU desktop computer having 1.0 GB RAM and a 1.66-GHz processor. To make the computations more manageable an effective interaction limit was used that restricted the cell-to-cell multiple scatterings to a range of six cell diameters.

Table 1 presents the elastic wave properties used for the extracellular matrix, cytoplasm, and nucleus. The properties of freshly excised myocardium were chosen for the extracellular matrix since they are intermediate between those of stiff biomaterials such as collagen and those of softer tissues in the body.⁹ The properties of the cell components are not well known and vary widely in the literature. The cytoplasm was therefore given waterlike properties with a room-temperature wave speed to correspond to published measurements of cell cultures, tissue cultures, and excised tissues. The properties of the nucleus were those used by Baddour *et al.* to model ultrasonic scattering from single cells and were obtained from typical eukaryotic nuclear properties and measurements from condensed cell nuclei.¹⁰ Both the cytoplasm and nucleus were given variable shear moduli. Although varying the shear wave speed of the extracellular matrix affected the simulation results, varying the shear wave speed of the cytoplasm and nucleus had no significant effect, including cases where the nucleus wave speed was higher than that of the cytoplasm. Viscoelastic (attenuation) properties are known to preferentially absorb shear and high-frequency waves in tissues, but were not included in this study since the objective was to demonstrate the feasibility of the approach and not to obtain detailed predictions. Attenuation will be included in future work, as well as comparison of model predictions with measurements from well-characterized tissues and cell cultures to refine the tissue and cell properties.

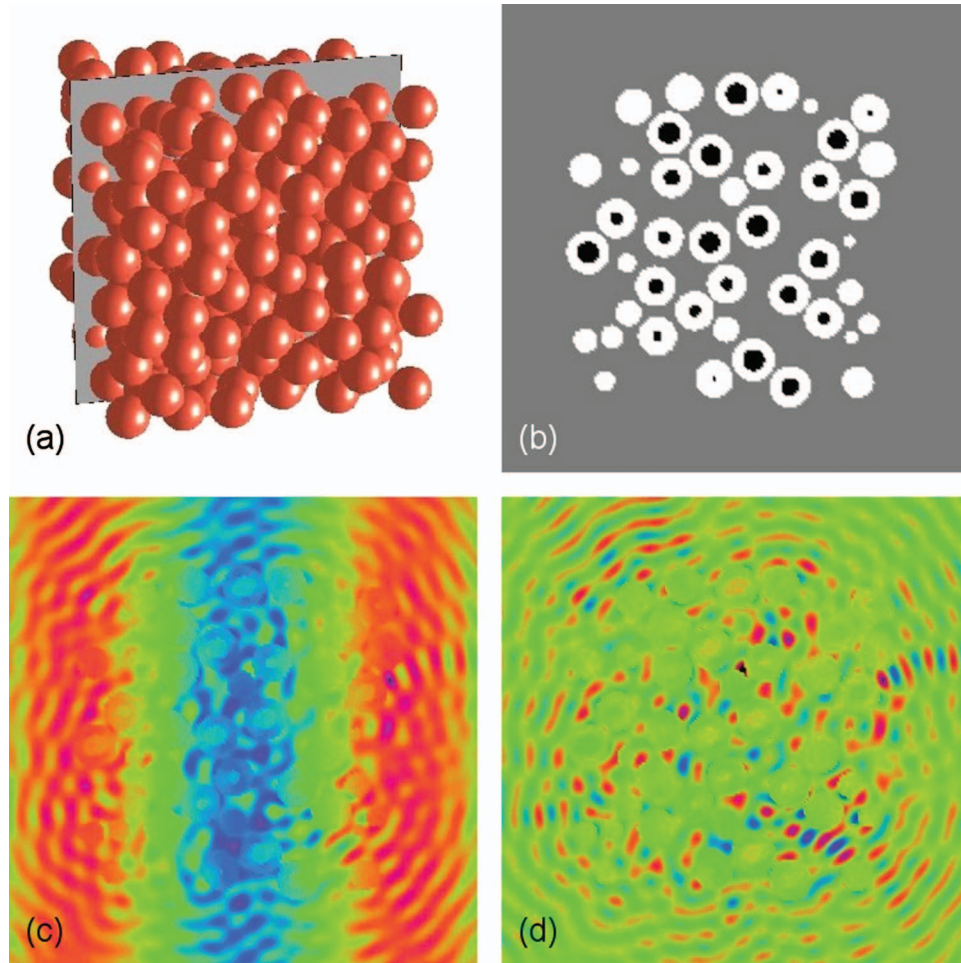


Fig. 1. (a) Three-dimensional depiction of a random cluster of 327 cells with a 50% volume density in an extracellular matrix. (b) Two-dimensional slice through the cell cluster. The cells and nuclei are 20 and 10 μm in diameter, respectively. (c) Scattered+incident wave field displacements from an incident 10-MHz longitudinal wave. (d) Scattered wave field displacements only. Red and blue denote negative and positive field amplitudes, respectively. Incident wave propagation is from left to right. This figure is intended for color viewing. The information may not be properly conveyed in a black and white printout.

Simulations of single cells determined that changes in the nuclear diameter had the greatest effect on the backscatter frequency spectra as compared to changes in cell size, density, or shear modulus. The spectra displayed several sharp peaks at high frequencies (20–100 MHz) that shifted with nuclear diameter. Recent experimental data from single cells and isolated nuclei also show strong correlations between backscatter amplitudes and nuclear diameter at high frequencies (20–60 MHz).¹¹ Both the single-cell simulations and experimental results strongly suggest the internal structure of cells should appreciably affect the propagation of ultrasound in tissues. However, the experimental spectra do not show the strong, sharp features as predicted by the models, probably due to the wide distributions of nuclear diameters and shapes for the cell types investigated.^{10,11} Details of the single-cell results will be presented in a forthcoming article.

Simulated wave field images from clusters of several hundred cells (Fig. 1) indicated that tissues do not behave acoustically as either homogeneous fluids or as a medium with simple

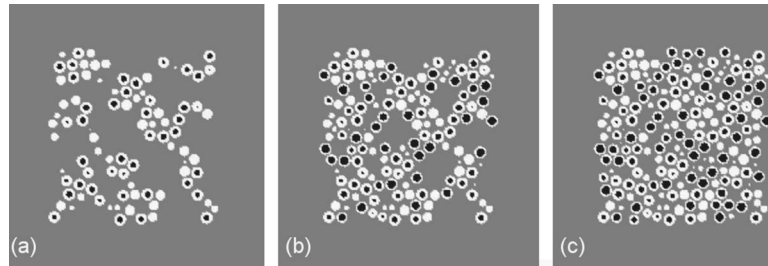


Fig. 2. Two-dimensional slices through random tissues consisting of 1049–2075 cells and randomly distributed cavities. Decreasing cavity size represents infiltration of malignant cells with larger nuclei (nuclear pleomorphism) into the cavities, with 0% (a), 57% (b), and 100% (c) infiltration denoting a corresponding filling of the cavity volume.

scatterers at the microscopic scale. Rather, extensive scattering can occur at both the cell membrane and nuclear envelope. The short-wavelength fields in Fig. 1(d) are shear waves in the extracellular matrix that arise from mode conversion, whereas the long-wavelength fields in Fig. 1(c) are the incident longitudinal waves. These results demonstrate the necessity of a modeling approach that differentiates the nucleus from the cytoplasm in the cell structure, and the importance of modeling tissues as semi-solids since fluids cannot support shear waves.

High-frequency (20–75 MHz) ultrasonic backscatter spectra were also acquired from tissue simulations containing up to 2137 cells. Both random and ordered (hexagonal-close-packed) cell packings were simulated, as well as random tissue structures where the nucleus diameter was varied for all of the cells. Extrapolating three-dimensional model tissue structures from two-dimensional histology micrographs is a difficult task and was not attempted in this preliminary study. Rather, as a first approximation the random tissue structures were constructed from a random particle packing generated by a Monte Carlo algorithm.¹²

Complex hierarchical tissues were also simulated consisting of large, randomly distributed cavities within a random cell packing. These cavities are analogous to structures such as lobules in breast tissue, and histological modifications to these tissue structures that corresponded to neoplastic changes were modeled. Figure 2 shows the simulated infiltration of malignant cells into tissue cavities, with both the normal and malignant cells having diameters of 20 μm . The malignant cells, however, were modeled with larger nuclei (14 μm) than those for the normal cells (10 μm) to represent nuclear pleomorphism. Average spectra and standard deviations were computed from the simulation of five tissue structures where the cell and cavity locations were varied, providing five test cases for each nucleus size or percent infiltration. Figure 3 displays representative spectra resulting from averaging simulation results from the five cases.

The simulation results demonstrated that both tissue structure and internal cell structure significantly affected the spectra. Comparisons between random and ordered cell packings displayed substantial differences in the spectra due to the structural ordering, and correctly predicted an acoustic band gap for the ordered cell packing. The random tissue spectra where the nuclear diameters were uniformly varied [Fig. 4(a)] exhibited substantial changes with nucleus size across a wide frequency range (20–56 MHz). Similarly, the backscatter spectra resulting from malignant cell invasion [Fig. 4(b)] displayed sizable amplitude changes at 47 MHz that were indicative of the degree of infiltration. Using the largest standard deviation for the 47-MHz peak as a conservative estimate, the results indicate that the detection limit would be approximately 300 malignant cells in a normal cell population. The simulations also suggest that spectral features can be used to differentiate malignant neoplasms from benign neoplasms such as hyperplasia.

Figures 4(a) and 4(b) display spectral features finer than those typically observed for high-frequency ultrasonic spectra of cells and tissues.^{6,10} This is most likely attributable to variations in the size and shape of cells and nuclei in actual specimens that smooth out the spectral structure. These variations can be incorporated into the simulations to generate more realistic spec-

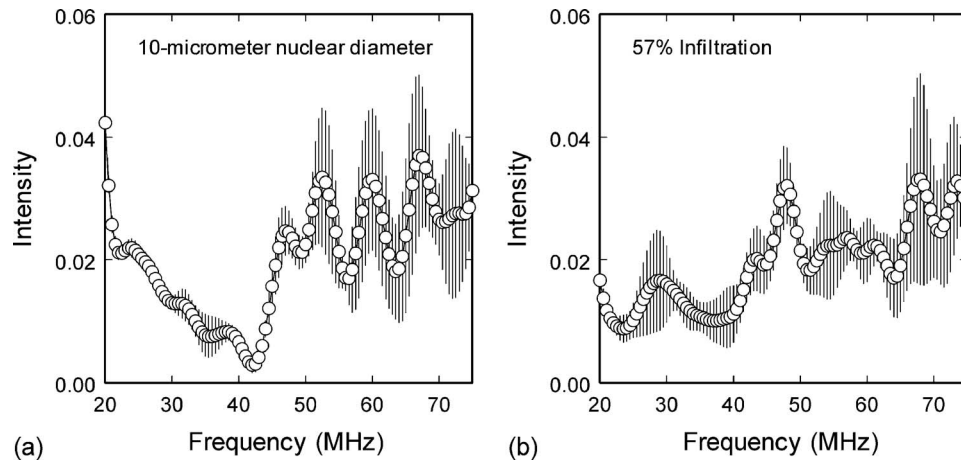


Fig. 3. Simulated spectra for (a) random tissues containing 2075 cells with a nucleus diameter of 10 μm and (b) 57% infiltration of malignant cells into tissue cavities [see Fig. 2(b)]. Plots show average intensities (circles) and standard deviations (bars) for the simulation of five tissue structures where the cell and cavity locations were varied.

tra. Additionally, Figs. 4(a) and 4(b) predict changes in broad spectral regions that should be observable in experimental spectra. An important conclusion from these simulations is that spectral resolution of histological modifications may not be limited by ultrasonic image resolution since many of the modeled changes are smaller than the ultrasonic wavelengths. For example, the changes in nucleus size shown in Fig. 4(a) were smaller by a factor of 10–40 than the wavelengths at which the changes were manifested in the spectra.

Comparison of the spectra in Fig. 4 emphasizes the importance of tissue heterogeneity in ultrasonic cancer detection. Both common and dissimilar spectral features arise from and are sensitive to different types of malignant processes in tissues. For example, both nuclear pleomorphism in homogeneous tissues [Fig. 4(a)] and infiltration of malignant cells into heterogeneous tissues [Fig. 4(b)] produce significant intensity increases at 47 MHz. The 47-MHz trends are nearly linear, with coefficients of linear correlation of $r=0.92$ [Fig. 5(a)] and $r=0.87$ [Fig.

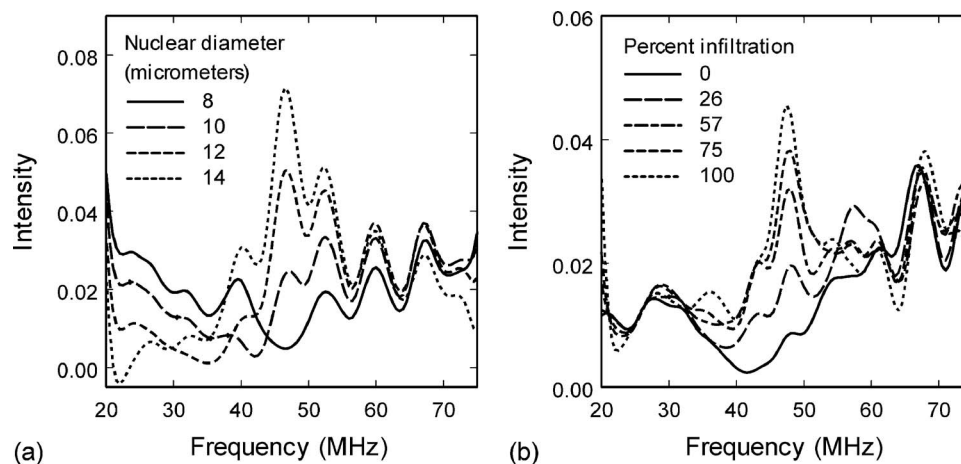


Fig. 4. (a) Simulated spectra for random tissues containing 2075 cells with the nucleus diameter varied uniformly for all of the cells. (b) Simulated spectra for random tissues with random cavities (e.g., Fig. 2) infiltrated by malignant cells with larger nuclei than the surrounding tissue. Spectra were averaged from simulations of five tissue structures where the cell and cavity locations were varied.

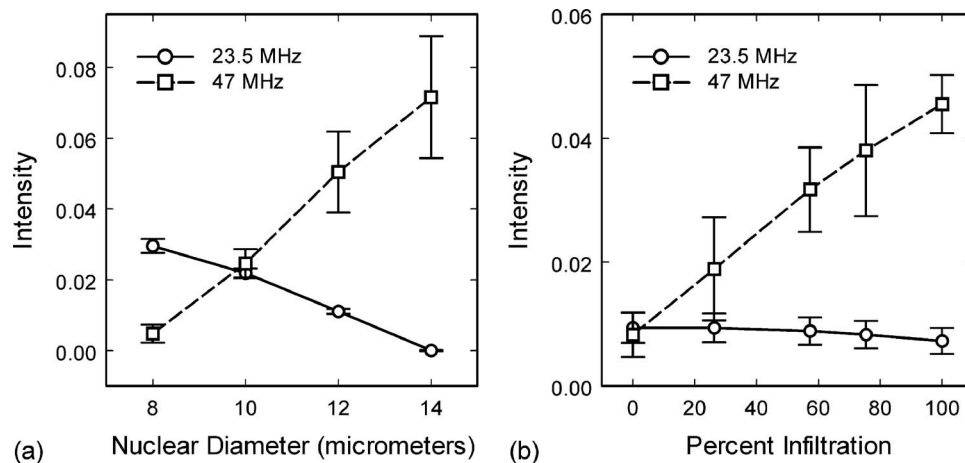


Fig. 5. Intensity trends at 23.5 and 47 MHz for (a) variation in nucleus size and (b) percent infiltration into tissue cavities, corresponding to the spectra shown in Figs. 4(a) and 4(b), respectively. The error bars are the standard deviations in the spectra arising from the five structural variations.

5(b)], respectively. In contrast, the spectral intensity at 23.5 MHz displays a negative correlation to nuclear diameter [Fig. 5(a), $r=-0.99$] and is insensitive to percent infiltration [Fig. 5(b), $r=-0.30$]. These differences in spectral responses may permit both the *in vivo* grading of homogeneous tissues in tumors and the *in vivo* staging of microscopic or micro-invasive cancer in heterogeneous tissues.

Researchers have acquired ultrasonic spectra in the 10–25-MHz range from 1-cm mouse mammary tumors to distinguish between carcinomas and sarcomas.⁶ The tumors exhibited different tissue morphologies (homogeneous for carcinoma versus heterogeneous for sarcoma), and the greatest spectral variations occurred from 16 to 25 MHz. Significant spectral differences in this band are also displayed in the simulation results between homogeneous [Fig. 4(a)] and heterogeneous [Fig. 4(b)] tissues. However, the heterogeneous tissue structures in the mouse sarcomas were not equivalent to the tissue models for invasive microscopic cancer simulated in this study (Fig. 2).

High-frequency ultrasound has not been as extensively developed for cancer detection as it has for intravascular ultrasound (IVUS) due to its shallow penetration into tissue (1–10 mm) and the lack of adequate data analysis methods to resolve sub-wavelength tissue structure.¹ Although low penetration is not suitable for clinical screening of deep tissues, it is ideally suited for differentiating normal from malignant tissues during needle biopsies, endoscopies, and surgical margin assessments. Additionally, the results from this study suggest ultrasonic spectroscopy can surpass the resolution limits of ultrasonic imaging for delineating cancer-associated histological changes. A spectral rather than an imaging approach to histopathology may also provide greater efficiency and sensitivity for real-time detection due to the relative simplicity of spectral measurements and availability of straightforward interpretation methods (for example, principal components analysis). In comparison to *in vivo* optical methods, ultrasonic spectroscopy assesses the elastic wave properties of cell and tissue components as well as their structure, thereby providing additional information for histological characterization.

Ultrasonic spectra and wave field images have been numerically simulated for representative tissue pathologies associated with cancer. The objective of this study was to demonstrate the feasibility of directly linking ultrasonic measurements to the histology of cancer by constructing a model that incorporates effects from cell structure, tissue heterogeneity, and multiple scattering. These first results suggest that microscopic changes in malignant tissues may significantly affect specific ultrasonic features at high frequencies. The results also demonstrate the potential uses of computer simulation for the development of *in vivo* cancer detection meth-

ods and data interpretation tools. Simulations of more complexly structured and histologically accurate tissue structures are planned to refine and validate these results.

Acknowledgments

The authors thank David Britt, Leigh Neumayer, Reyhan Baktur, Raymond DeVito, and Nicholas Flann for their encouragement, assistance, and suggestions. This work was supported by the Utah State University Community/University Research Initiative.

References and links

- ¹F. Koenig, J. Knittel, and H. Stepp, "Diagnosing cancer *in vivo*," *Science* **292**, 1401–1403 (2001).
- ²E. J. Feleppa, A. Kalisz, J. B. Sokil-Melgar, F. L. Lizzi, T. Liu, A. L. Rosado, M. C. Shao, W. R. Fair, Y. Wang, M. S. Cookson, V. E. Reuter, and W. D. W. Heston, "Typing of prostate tissue by ultrasonic spectrum analysis," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **43**, 609–619 (1996).
- ³B. Alacam, B. Yazici, N. Bilgutay, F. Forsberg, and C. Piccoli, "Breast tissue characterization using FARMA modeling of ultrasonic RF echo," *Ultrasound Med. Biol.* **30**, 1397–1407 (2004).
- ⁴Y. Bige, Z. Hanfeng, and W. Rong, "Analysis of microstructural alterations of normal and pathological breast tissue *in vivo* using the AR cepstrum," *Ultrasonics* **44**, 211–215 (2006).
- ⁵M. L. Oelze, J. F. Zachary, and W. D. O'Brien, Jr., "Characterization of tissue microstructure using ultrasonic backscatter: theory and technique for optimization using a Gaussian form factor," *J. Acoust. Soc. Am.* **112**, 1202–1211 (2002).
- ⁶M. L. Oelze and J. F. Zachary, "Examination of cancer in mouse models using high-frequency quantitative ultrasound," *Ultrasound Med. Biol.* **32**, 1639–1648 (2006).
- ⁷F. L. Lizzi, M. Astor, E. J. Feleppa, M. Shao, and A. Kalisz, "Statistical framework for ultrasonic spectral parameter imaging," *Ultrasound Med. Biol.* **23**, 1371–1382 (1997).
- ⁸T. E. Doyle, "Iterative simulation of elastic wave scattering in arbitrary dispersions of spherical particles," *J. Acoust. Soc. Am.* **119**, 2599–2610 (2006).
- ⁹M. Yang, S. L. Baldwin, K. R. Marutyan, K. D. Wallace, M. R. Holland, and J. G. Miller, "Elastic stiffness coefficients (c_{11} , c_{33} , and c_{13}) for freshly excised and formalin-fixed myocardium from ultrasonic velocity measurements," *J. Acoust. Soc. Am.* **119**, 1880–1887 (2006).
- ¹⁰R. 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).
- ¹¹L. R. Taggart, R. E. Baddour, A. Giles, G. J. Czarnota, and M. C. Kolios, "Ultrasonic characterization of whole cells and isolated nuclei," *Ultrasound Med. Biol.* **33**, 389–401 (2007).
- ¹²J. L. Finney, "Random packings and the structure of simple liquids: I. The geometry of random close packing," *Proc. R. Soc. London, Ser. A* **319**, 479–493 (1970).

Comments on the limp frame equivalent fluid model for porous media

Raymond Panneton

*GAUS, Department of Mechanical Engineering, Université de Sherbrooke, Sherbrooke, Qc, Canada, J1K 2R1
raymond.panneton@usherbrooke.ca*

Abstract: In this letter, the low and high frequency limits of the effective density characterizing a limp frame porous medium are investigated. These theoretical limits are compared to the ones found for a classical rigid frame porous medium, and to experimental measurements. While the high frequency asymptotic behaviors of both limp and rigid effective densities are usually only slightly different, their low frequency behaviors are significantly different. Compared to experimental measurements performed on a limp frame fibrous layer, only the limp frame effective density yields good correlations over the whole frequency range.

© 2007 Acoustical Society of America

PACS numbers: 43.20.Jr, 43.20.Hq, 43.55.Ev, 43.55.Rg [MS]

Date Received: July 19, 2007 **Date Accepted:** September 17, 2007

1. Introduction

Open-cell porous materials used to attenuate sound may be categorized in three frame types: elastic, rigid, and limp. While polymeric foams are common examples of elastic frame porous materials, metal foams and soft fibrous layers are examples of rigid and limp frame materials, respectively. For an elastic porous medium, the Biot theory^{1,2} is commonly used to describe the propagation of waves in the medium. This theory can be used to model limp and rigid porous media; however, extreme values in the rigidity may create instabilities in a numerical implementation. To prevent instabilities and to reduce the number of degrees of freedom in a numerical poroelastic modeling,³ an equivalent fluid model is often used. In this case, only the acoustic compression wave propagating in the porous medium is considered.⁴ This propagation is governed by the Helmholtz equation in which the equivalent fluid is characterized by an effective density and an effective bulk modulus accounting for the viscous and thermal dissipations of the acoustic wave. For rigid frame porous media, different models were proposed to evaluate these effective properties.⁴⁻¹⁰ On the other hand, only a few models were proposed for limp porous media, and few comments on their frequency limits were given.¹¹⁻¹⁴ However, the work by Lai *et al.*¹³ has underlined the importance of the limp model in view to prevent erroneous predictions of global acoustic indicators, such as the sound transmission loss, when using a rigid frame model.

The intent of this letter is to develop and experimentally validate the low and high frequency limits of the effective density of the limp model derived as a limit case of the Biot poroelastic mixed model.³ Also, it is intended to show how these limits compare to the classical rigid frame models to explain the aforementioned erroneous predictions.

2. Theory

2.1 Poroelastic model

For an open-cell porous medium made up from an elastic solid phase (the frame) and a fluid phase, the Biot theory states that three waves propagate simultaneously in the medium: one elastic compression wave, one elastic shear wave, and one acoustic compression wave.^{1,2} Defining the fluid pore pressure by p and the solid phase displacement vector by \underline{u} , the dynamic behavior of the three waves is governed by the following two coupled equations:³

$$\operatorname{div} \hat{\sigma} + \omega^2 \tilde{\rho}_s \underline{u} = -\tilde{\gamma} \operatorname{grad} p \quad \text{and} \quad \frac{\phi}{\omega^2 \tilde{\rho}_f} \Delta p + \frac{\phi}{\tilde{K}_f} p = \tilde{\gamma} \operatorname{div} \underline{u} \quad (1)$$

with the volume coupling coefficient given by

$$\tilde{\gamma} = \frac{\phi \rho_0}{\tilde{\rho}_f} + \frac{K_b}{K_s} - 1 \quad (2)$$

and the effective solid phase density

$$\tilde{\rho}_s = \rho_1 + \phi \rho_0 \left(1 - \frac{\rho_0}{\tilde{\rho}_f} \right). \quad (3)$$

In these equations, Δ is the Laplace operator ($\Delta = \operatorname{div} \operatorname{grad}$), the tilde symbol represents a variable that is complex and frequency dependent, ϕ is the open porosity of the porous medium, ρ_0 is the density of the fluid saturating the pores, ρ_1 is the *in vacuo* bulk density of the medium, $\hat{\sigma}$ is the *in vacuo* stress tensor, K_s is the bulk modulus of the elastic material from which the frame is made, and K_b is the *in vacuo* bulk modulus of the frame. The effective density $\tilde{\rho}_f$ and bulk modulus \tilde{K}_f of the fluid phase may be estimated using different general semiphenomenological models taking into account the viscous and thermal losses of the acoustic compression wave.⁴⁻¹⁰

2.2 Rigid frame porous model

When the frame of the porous medium is assumed motionless, the frame does not undergo any displacement (i.e., $\underline{u}=0$) and deformation. This situation occurs under acoustic excitations when the frame is constrained and rigid, heavy, or when the solid-fluid coupling is negligible for an elastic frame.¹⁵ Consequently, only the acoustic compression wave propagates in the porous medium, and only the second equation of Eq. (1) remains and simplifies to the following Helmholtz equation:

$$\frac{1}{\omega^2 \tilde{\rho}_{\text{eq}}} \Delta p + \frac{1}{\tilde{K}_{\text{eq}}} p = 0, \quad (4)$$

where $\tilde{\rho}_{\text{eq}} = \tilde{\rho}_f / \phi$ and $\tilde{K}_{\text{eq}} = \tilde{K}_f / \phi$ are the effective density and bulk modulus of the so-called *rigid frame equivalent fluid* medium. These quantities can be used to deduce the more common characteristic impedance and propagation constant of the equivalent fluid medium by $Z_{\text{ceq}} = \sqrt{\rho_{\text{eq}} K_{\text{eq}}}$ and $\gamma = j\omega \sqrt{\rho_{\text{eq}} / K_{\text{eq}}}$, respectively. Note that the “rigid frame model” appellation is common in the literature to define Eq. (4), and its effective properties. However, it may seem abusive since in dynamics “rigid” only implicates no deformation. As a result, the rigid frame model eliminates *de facto* the rigid body motion of the solid phase. This points out the potential abusive use of the rigid frame model in the case where the frame is rigid but unconstrained (e.g., highly resistive light foam with a sliding edge condition). This problem was stressed by Lai *et al.*¹³

2.3 Limp frame porous model

When the frame of the porous medium is assumed limp (i.e., flexible), the frame does not resist to external excitations and its stress field vanishes (here, assuming $K_b/K_s \cong 0$). This situation occurs for solid particles in suspension in a fluid medium or for porous media with very low shear modulus—in the limit case, the shear modulus is zero. Again, as for the rigid frame model, only the acoustic compression wave exists. Consequently, since the stress field vanishes, taking the divergence of the first equation in Eq. (1), then substituting the result into the second equation of Eq. (1) yields the following *limp frame equivalent fluid* equation

$$\frac{1}{\omega^2 \tilde{\rho}'_{\text{eq}}} \Delta p + \frac{1}{\tilde{K}_{\text{eq}}} p = 0. \quad (5)$$

While the effective bulk modulus stays unchanged (compared to rigid frame media), the effective density of the equivalent fluid is modified as follows:

$$\frac{1}{\tilde{\rho}'_{\text{eq}}} = \frac{1}{\tilde{\rho}_{\text{eq}}} + \frac{\tilde{\gamma}^2}{\tilde{\rho}_s}. \quad (6)$$

Since K_s is generally much greater than K_b , the limp effective density may simplify to

$$\tilde{\rho}'_{\text{eq}} \approx \frac{\tilde{\rho}_{\text{eq}} M - \rho_0^2}{M + \tilde{\rho}_{\text{eq}} - 2\rho_0}, \quad (7)$$

where $M = \rho_1 + \phi \rho_0$ is the total apparent mass of the equivalent fluid limp medium.

Equation (5) is similar to Eq. (4); however, this time the effective density takes into account the inertia added by the limp solid phase. Taking the high bulk density limit of Eqs. (6) or (7), one can show that $\tilde{\rho}'_{\text{eq}} \approx \tilde{\rho}_{\text{eq}}$ (as ρ_1 or $M \rightarrow \infty$). This limit proves that high density limp porous media can be modeled as rigid frame porous media, as mentioned in the previous section. On the other hand, taking the low bulk density limit of Eq. (6) and assuming porosity tends to unity (logical for low density materials), one can show that $\tilde{\rho}'_{\text{eq}} \approx \rho_0$ (as $\rho_1 \rightarrow 0$ and $\phi \rightarrow 1$). Consequently, as ρ_1 reduces, the viscous losses in the limp frame material reduce, and the thermal loss (introduced by K_{eq}) stays unchanged.

Finally, it is worth mentioning that the characteristic impedance and propagation constant of the limp frame medium can be computed as for the rigid frame medium; however, this time the limp effective density is used: $Z'_{\text{ceq}} = \sqrt{\rho'_{\text{eq}} K_{\text{eq}}}$ and $\gamma' = j\omega \sqrt{\rho'_{\text{eq}} / K_{\text{eq}}}$. In this case, both properties are modified by the limpness of the frame. In the particular case of the aforementioned low bulk density limit, since the viscous losses vanish and the thermal losses are generally small in porous media, both Z'_{ceq} and γ' approach the characteristic impedance and propagation constant of air. Inversely, as the bulk density increases, the viscous losses take more importance and Z'_{ceq} and γ' may diverge strongly from the air values, and reach the rigid frame behavior when $\rho_1 \rightarrow \infty$.

3. Results and discussions

Both rigid and limp effective densities are now compared to experimental results, and their high and low frequency limits analyzed. For the simulation results, the semiphenomenological model worked out by Johnson *et al.*⁶ is used to compute $\tilde{\rho}_{\text{eq}}$. This model writes

$$\tilde{\rho}_{\text{eq}} = \frac{\rho_0 \alpha_\infty}{\phi} \left(1 - j \frac{\sigma \phi}{\omega \rho_0 \alpha_\infty} \sqrt{1 + j \frac{4 \alpha_\infty^2 \eta \rho_0}{\sigma^2 \phi^2 \Lambda^2} \omega} \right), \quad (8)$$

where the material properties are given in Table 1 and correspond to a soft fibrous material. For the experimental results, the method worked out by Utsuno *et al.*¹⁶ was used in conjunction with a 100-mm-diam Brüel and Kjaer 4206 impedance tube to deduce the measured effective density.

Figure 1 presents the comparison between the predictions and the measurement of the effective density of the soft fibrous material. As one can note, both rigid and limp models yield good predictions at higher frequencies [see Fig. 1(a)]; however, when zooming at high frequencies [see Fig. 1(b)], a slight discrepancy is observed for the real part of the rigid model. This may be explained by looking at the high frequency limits of the models:

Table 1. Material properties of the studied soft fibrous material.

	Symbol	Value	Units
Open porosity	ϕ	0.98	
Static airflow resistivity	σ	25 000	N s/m ⁴
Tortuosity	α_∞	1.02	
Viscous characteristic length	Λ	90	μm
Bulk density	ρ_1	30	kg/m ³
Density of saturating air	ρ_0	1.208	kg/m ³
Viscosity of saturating air	η	1.81×10^{-5}	N s/m ²
Elastic bulk modulus ratio	K_b/K_s	≈ 0	

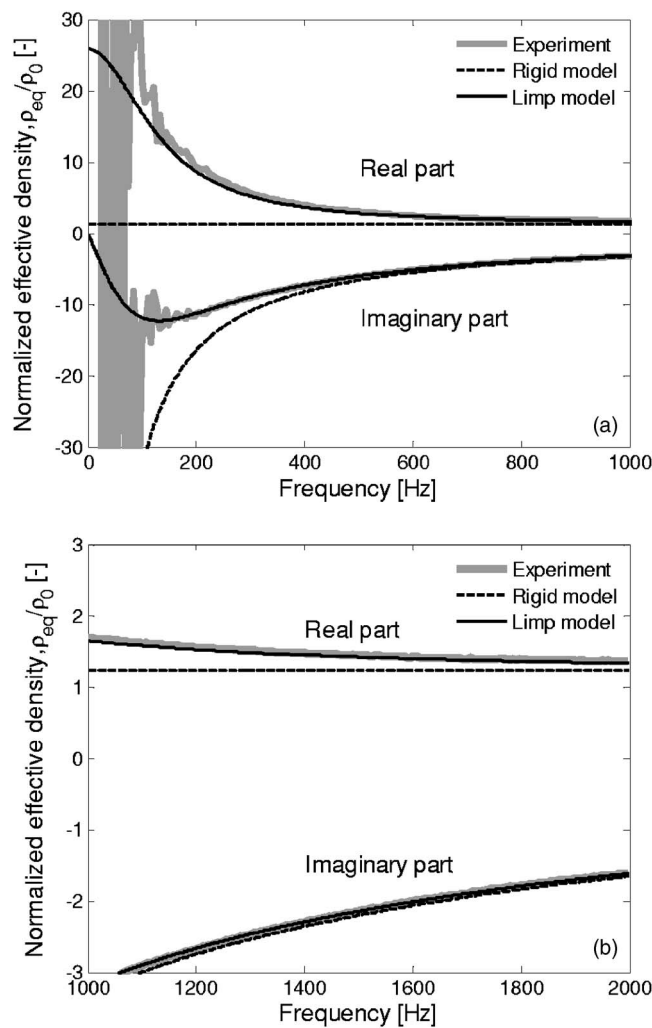


Fig. 1. Normalized effective density of the soft fibrous material. Comparison between rigid frame predictions, limp frame predictions, and experimental results. (a) Zoom between 0 and 1000 Hz. (b) Zoom between 1000 and 2000 Hz.

$$\lim_{\omega \rightarrow \infty} \tilde{\rho}_{\text{eq}} = \frac{\rho_0 \alpha_\infty}{\phi} \quad \text{and} \quad \lim_{\omega \rightarrow \infty} \tilde{\rho}'_{\text{eq}} = \frac{\rho_0 \alpha_\infty}{\phi} \left(1 + \rho_0 \frac{\phi / \alpha_\infty + \alpha_\infty / \phi - 2}{M - \phi / \alpha_\infty} \right)^{-1}. \quad (9)$$

From these limits, it is clear that the asymptotic high frequency behaviors of both models are only slightly different. For a porosity and a tortuosity nearly equal to unity (typical for light fibrous materials), both limits are equal.

On the other hand, at low frequencies, the prediction obtained with the rigid model significantly diverges from the measured real and imaginary parts of the effective density. In this case, only the limp model correctly predicts the measured low frequency behavior of the limp material. Again, this may be explained by analyzing the following low frequency limits of the models:

$$\lim_{\omega \rightarrow 0} \tilde{\rho}_{\text{eq}} = \frac{\rho_0 \alpha_\infty}{\phi} \left(1 + \frac{2\alpha_\infty \eta}{\phi \sigma \Lambda^2} \right) - j \frac{\sigma}{\omega} \quad \text{and} \quad \lim_{\omega \rightarrow 0} \tilde{\rho}'_{\text{eq}} = M. \quad (10)$$

At the low frequency limit, one can observe that the rigid effective density has an asymptotic behavior dominated by a strong imaginary part. On the contrary, the limp effective density yields a purely real static value equal to the global apparent density $M = \rho_1 + \phi \rho_0$ of the equivalent fluid medium. This static value does not depend on the semiphenomenological model used to compute $\tilde{\rho}_{\text{eq}}$ —in that sense, the limits on $\tilde{\rho}'_{\text{eq}}$ may be seen as exact.

The fact the low frequency limit of the limp model is real and equal to the global apparent density of the medium has a major impact on the prediction of global acoustic indicators (absorption and transmission coefficients) of limp porous materials. In fact, the limp model implicitly accounts for the rigid body motion defined at 0 Hz for which all the solid phase particles move in phase (i.e., no deformation). This rigid mode strongly controls the dynamic behavior of the limp porous medium at low frequencies. Also, it explains why the rigid frame model may yield erroneous results at low frequencies.¹³

Also, beyond the use of the limp model for limp frame porous materials, one can also use the limp model to capture the rigid body motion of an unconstrained rigid frame porous material. In fact, if the sound absorption coefficient or the sound transmission loss of a relatively stiff porous sample (or poroelastic with weak solid-fluid coupling as some polymeric foams¹⁵) is measured in an impedance tube, and assuming the sample can move freely along the axis of the tube (i.e., sliding boundary conditions), the measured coefficient will correlate with the limp frame model and not with the rigid frame model at low frequencies.

Finally, if one prefers to directly input the effective properties $\tilde{\rho}_{\text{eq}}$ and \tilde{K}_{eq} obtained from measurements in Eq. (1) (instead of measuring all the physical properties of the medium and using a semiphenomenological model), one needs to ensure that the measured effective density corresponds to $\tilde{\rho}_{\text{eq}}$. From the previous analysis, if one note that the tested porous sample has a rather limp frame (or unconstrained rigid frame), or that the measured effective density has a limp-like behavior as the one shown in Fig. 1(a), the measured effective density needs to be corrected. Derived from Eq. (7), the correction to apply is

$$\tilde{\rho}_{\text{eq}} = \frac{(M - 2\rho_0)\tilde{\rho}'_{\text{eq}} + \rho_0^2}{M - \tilde{\rho}'_{\text{eq}}}, \quad (11)$$

where $\tilde{\rho}'_{\text{eq}}$ is assumed to be the measured effective density.

4. Conclusions

From the previous results and discussions, it was shown that some precautions have to be considered when using an equivalent fluid model for porous media. In many situations, it is important to take into account the static rigid body mode of the porous medium to obtain the proper low frequency behavior in terms of sound absorption and, mainly, in terms of sound transmission. It was shown that the rigid body mode can be captured by the limp frame equivalent fluid model, and not by the rigid frame equivalent fluid model. Consequently, prior to simulating a

system containing a porous medium, the practitioner should ensure to select the proper porous model. If some uncertainties remain on the selection of the model, it is suggested to use the full poroelastic model to prevent any erroneous results. However in this case, the elastic properties of the medium are required, and larger numerical systems with potential instabilities may result.^{3,13}

Acknowledgment

The author thanks N.S.E.R.C. Canada and F.Q.N.R.T. Quebec for their financial support.

References and links

- ¹M. A. Biot, "The theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," *J. Acoust. Soc. Am.* **28**, 168–178 (1956).
- ²M. A. Biot, "The theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher-frequency range," *J. Acoust. Soc. Am.* **28**, 179–191 (1956).
- ³N. Atalla, R. Panneton, and P. Debergue, "A mixed displacement pressure formulation for poroelastic materials," *J. Acoust. Soc. Am.* **104**, 1444–1452 (1998).
- ⁴J.-F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Elsevier, New York, 1993).
- ⁵K. Attenborough, "Acoustical characteristics of rigid fibrous absorbents and granular materials," *J. Acoust. Soc. Am.* **73**, 785–799 (1983).
- ⁶D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," *J. Fluid Mech.* **176**, 379–402 (1987).
- ⁷Y. Champoux and J.-F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media," *J. Appl. Phys.* **70**, 1975–1979 (1991).
- ⁸D. K. Wilson, "Relaxation-matched modeling of propagation through porous media, including fractal pore structure," *J. Acoust. Soc. Am.* **94**, 1136–1145 (1993).
- ⁹S. R. Pride, F. D. Morgan, and A. F. Gangi, "Drag forces of porous-medium acoustics," *Phys. Rev. B* **47**, 4964–4975 (1993).
- ¹⁰D. Lafarge, P. Lemarinier, J.-F. Allard, and V. Tarnow, "Dynamic compressibility of air in porous structures at audible frequencies," *J. Acoust. Soc. Am.* **102**, 1995–2006 (1997).
- ¹¹K. U. Ingard, "Locally and nonlocally reacting flexible porous layers: A comparison of acoustical properties," *J. Eng. Ind.* **103**, 302–313 (1981).
- ¹²P. Göransson, "Acoustic finite element formulation of a flexible porous material—a correction for inertial effects," *J. Sound Vib.* **185**, 559–580 (1995).
- ¹³H.-Y. Lai, S. Katragda, and J. S. Bolton, "Layered fibrous treatments for sound absorption and sound transmission," SAE paper No. 972064, 1553–1560 (1997).
- ¹⁴O. Dazel, B. Brouard, C. Depollier, and Stéphane Griffiths, "An alternative Biot's displacement formulation for porous materials," *J. Acoust. Soc. Am.* **121**, 3509–3516 (2007).
- ¹⁵D. Pilon, F. Sgard, and R. Panneton, "Frame acoustical excitability: A decoupling criterion for poroelastic materials," *Proc. of Euronoise*, paper ID: **9**, 1–6 (2003).
- ¹⁶H. Utsuno, T. Tanaka, and T. Fujikawa, "Transfer function method for measuring characteristic impedance and propagation constant of porous materials," *J. Acoust. Soc. Am.* **86**, 637–643 (1989).

Spatial structure of low-frequency wind noise

D. Keith Wilson and Roy J. Greenfield

*U.S. Army Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory,
72 Lyme Rd., Hanover, New Hampshire 03755
d.keith.wilson@erdc.usace.army.mil
roy@essc.psu.edu*

Michael J. White

*U.S. Army Engineer Research and Development Center, Construction Engineering Research Laboratory,
2902 Newmark Dr., Champaign, Illinois 61822
michael.j.white@erdc.usace.army.mil*

Abstract: The distinguishing spatial properties of low-frequency microphone wind noise (turbulent pressure disturbances) are examined with a planar, 49-element array. Individual, propagating transient pressure disturbances are imaged by wavelet processing to the array data. Within a given frequency range, the wind disturbances are much smaller and less spatially coherent than sound waves. Conventional array processing techniques are particularly sensitive to wind noise when sensor separations are small compared to the acoustic wavelengths of interest.

© 2007 Acoustical Society of America

PACS numbers: 43.28.Ra, 43.60.Cg, 43.28.Vd [NX]

Date Received: July 13, 2007 **Date Accepted:** August 5, 2007

1. Introduction

Turbulent pressure disturbances, or *wind noise*, are most often the primary source of unwanted low-frequency noise on microphones. Morgan and Raspet¹ have shown that, for a single microphone, wind noise is caused primarily by turbulence already present in the ambient atmospheric flow, as opposed to turbulence generated by flow interaction with the microphone assembly itself.

A single pressure sensor is generally unable to distinguish the turbulent pressure from the acoustic pressure signal of interest. However, arrays of sensors can potentially distinguish turbulence and sound waves based on their differing spatial and propagation characteristics. Turbulence translates near the mean wind speed, which is two orders of magnitude slower than sound. (Hence Bass *et al.*² were able to estimate wind speed and direction by tracking turbulent pressure disturbances moving across a microphone array.) Therefore, in a particular frequency range, the wavelength of sound is much longer than the size of the turbulent eddies. Turbulence is also comparatively incoherent. Shields³ showed that when wind noise is narrowband filtered, the correlation length is about 1/3 the wavelength (eddy size) corresponding to that band.

This letter describes the comparative spatial structure of wind noise and sound waves by examining data from a large, planar microphone array. A smaller sensor spacing (0.15 m) is used than would normally be chosen for a low-frequency array design, so that pressure disturbances associated with individual turbulent eddies can be discerned. In order to preserve the integrity of the pressure disturbances in space and time, wavelet processing is applied to the signals. This type of analysis is particularly helpful for determining how wind noise interferes with discrimination of acoustic transients such as blasting noise.

2. Description of experiment and apparatus

The measurements reported here were all made with a 7×7 planar, horizontal microphone array. The 49 individual microphones were 13-mm-diam free-field condenser transducers and were covered with conventional 10-cm-diam spherical foam windscreens. Spacing between neighboring microphones was 0.152 m, so that the total aperture of the array was 0.914 m by 0.914 m. The microphone height was 0.46 m above the ground, and 0.24 m above a smooth,



Fig. 1. (Color online). Left: W. Fairlee, VT site, showing the microphone array and solar panels. Right: Tolono, IL site, showing the microphone array and ultrasonic anemometers.

plastic plate. The microphone signals were acquired with three, 24-channel, 24-bit seismographs. The sampling rate of the seismographs was 1000 Hz and the signals were passed through a low-pass, anti-aliasing filter with cutoff at 250 Hz. Three ultrasonic anemometers, with 30 Hz sampling rate, arranged in an equilateral triangle with 1 m side length, were positioned approximately 3 m from the array and with the same nominal height as the microphones.

Data were collected in two quiet, rural locations, as shown in Fig. 1. One location, a farm in West Fairlee, VT (43.9371°N, 72.2279°W) was selected for the presence of large objects to disturb the atmospheric flow. The other location, the Monticello Road Field Station near Tolono, IL (40.0180°N, 88.3282°W), was selected for its flat terrain. The VT experiment was conducted from 25 Apr to 1 May 2006. The immediate vicinity of the microphone array was fairly flat, mowed grassland. A row of mature trees stood about 30 m from the array. Buildings, large hills, and more trees were all within several hundred meters. The IL experiment was performed from 10 to 13 July 2006. The area around the array was mowed grass, although 1.5-m-high corn was present on all sides. The closest corn row was about 50 m to the north of the array.

The VT experiment encompassed a variety of weather conditions, including very weak (0.2 m s^{-1}) to moderately strong (6 m s^{-1}) wind. The background noise at this location consisted of infrequent vehicle traffic on a nearby small roadway, occasional distant aircraft, and some natural sounds such as birds. For this experiment, we only monitored the existing environment; no additional sound sources were introduced. For the IL location, occasional road, air, and rail traffic were audible. Wind speeds were less than 4 m s^{-1} during the entire data collection. In addition to monitoring the existing environment, in some trials we intentionally introduced music from a car stereo, random noise from loudspeakers, controlled vehicle pass-bys, and propane cannon blasts. For both experiments, the incidental vehicle traffic was carefully documented and only intervals without such traffic are analyzed here.

The analyses here are based on five particular 180 s data segments, which were selected for their distinctive characteristics. A summary of these five cases is provided in Table 1.

3. Pressure field images

This section provides some example images and animations of the pressure signals from the 7×7 array. In order to preserve the structure of the signals in space and time, they were processed with a continuous wavelet transform (CWT). Since a wind gust may consist of a single burst with positive and negative pressure phases, a multi-cycle Fourier transform cannot localize it well. As is often the case, the results of the CWT were found to depend significantly on the choice of analyzing wavelet. The localization properties of the wavelet in both time and frequency are important. Since the wind disturbances propagate across the array, poor temporal localization is also apparent as poor spatial localization. For the widely used derivative-of-Gaussian (DOG) family of wavelets, frequency localization improves, but temporal localization

Table 1. Date segments selected for processing. The segments extend 180 s past the start time. Wind measurements were made at 0.46 m height.

Case	Location	Start date, time (UTC)	Description	Wind speed, mean \pm std (m/s)	Wind direction (deg, 0=from N)
A	W. Fairlee	28 Apr. 2006, 16:35:53	Gusty wind, quiet	3.51 \pm 1.19	51
B	W. Fairlee	1 May 2006, 12:26:10	Very low wind, quiet	0.46 \pm 0.20	148
C	Tolono	11 Jul 2006, 22:52:00	Mod. wind, quiet	3.37 \pm 0.79	204
D	Tolono	11 Jul 2006, 23:43:14	Mod. wind, loud music	2.65 \pm 0.70	200
E	Tolono	12 Jul 2006, 15:54:32	Low wind, cannon	1.63 \pm 0.45	233

degrades, as the derivative order m increases. In particular, we found that for $m \geq 2$, the images give a false sense of large-sized eddies at frequencies above roughly 30 Hz. (This is likely due to the spectrum of the wind noise, which contains much more large-scale energy.) Hence, for the images presented here, we have used a DOG1 ($m=1$) wavelet. As discussed in Ref. 4, the DOG m wavelet with time scale s responds most strongly to a harmonic signal with frequency $f = \sqrt{(m+1/2)}/2/2\pi s$. For convenience, in this letter we consistently characterize the scale of the wavelets by referring to this frequency, rather than to the scale itself.

Figures 2(a)-2(c) show CWTs at 5, 15, and 50 Hz. These spatial images are snapshots for a particular moment in time near the middle of case A, which was notable for its gusty winds. The snapshots appear to be dominated by a sharp wind gust oriented diagonally across the array. The DOG1 wavelets, at all three frequencies, respond to the edge of this feature. Mm 1 and Mm 2 provide 10 s animations of the CWTs at 5 and 50 Hz. The animations show that the disturbances at 5 Hz move erratically but with a speed near the mean wind speed. At 50 Hz, the disturbances tend to be smaller and even more erratic. Turbulence dominates the pressure field for all frequencies shown here.

Figures 2(d)-2(f) are the same as 2(a)-2(c), except they are for Case D. Mm 3 and Mm 4 provide corresponding animations. In comparison to Case A, the field is smoother and more coherent at 50 Hz. Since the wavelength of sound at this frequency is much larger than the size of the array, these smooth pressure disturbances are almost certainly music from the car stereo positioned north of the array during this trial. The animation confirms the rapid propagation speed of these features.

Figure 2(g) is the same as (a), except that delays have been introduced to the signals. The delay times were chosen to track the estimated velocity of the wind gust, which by trial and error was determined to be 3.5 m s^{-1} from the direction of 10° (measured clockwise from N, as in the meteorological convention). It is seen that the wind gust is fairly persistent as it propagates over the array, as consistent with the results of Bass *et al.*² However, it does not extend substantially in the directions perpendicular to its propagation, as would be the case for a nominal plane wave. Figure 2(h) is the same as (g), except that delay times corresponding to the acoustic propagation speed, 330 m s^{-1} (still from 10°), are chosen. The image is essentially identical to Fig. 2(a). This demonstrates that the wind noise is essentially independent of the “look” direction for a conventional, plane-wave beamformer operating at acoustic wavelengths.

Lastly, Fig. 2(i) shows the CWT at 100 Hz, squared and then averaged over a 15 min interval including Case A. This processing, which indicates the relative strength of the pressure fluctuations across the array, helps to reveal the flow modification by the array and windscreens. Since the wind direction for this interval is from NNE, it appears that the row of microphones on the far upstream side of the array experience systematically less turbulence than the following two rows, which exhibit wind noise levels several dB higher. The microphones on the downstream side appear to be somewhat shielded from the turbulence. The array thus appears to locally modify the turbulent flow. Analyses for other frequencies show similar trends, although the spatial variation is less at lower frequencies.

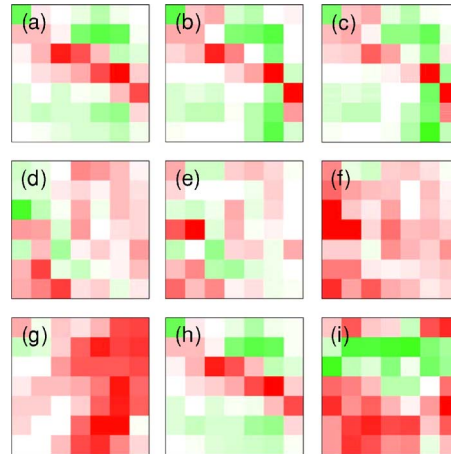


Fig. 2. (Color online). Images of DOG1 continuous wavelet transforms (CWTs) of signals from the 7×7 array. In all images, except as noted, fully saturated green corresponds to twice the maximum of the standard deviation of the time series, and fully saturated red corresponds to the negative of that value. Top of the page is north. (a) Snapshot from CWT for Case A, 5 Hz (b) Same as (a), except 15 Hz. (c) Same as (b), except 50 Hz. (d) Snapshot from CWT for Case D, 5 Hz (e) Same as (d), except 15 Hz. (f) Same as (d), except 50 Hz. (g) Same as (a), except that the signals have been delayed to adjust for the velocity of a wind gust moving across the array. (h) Same as (a), except that the signals have been delayed to adjust for the velocity of a hypothetical sound wave moving across the array. (i) Average power (dB) in the 100 Hz wavelets over 15 min interval including Case A. The total dynamic range is 5 dB from red (lower) to green (higher).

[Mm 1. 10 s animation of the CWT from Case A, 5 Hz. This is a file of type “mpg” (1.8 Mb).]

[Mm 2. 10 s animation of the CWT from Case A, 50 Hz. This is a file of type “mpg” (2.7 Mb).]

[Mm 3. 10 s animation of the CWT from Case D, 5 Hz. This is a file of type “mpg” (1.5 Mb).]

[Mm 4. 10 s animation of the CWT from Case D, 50 Hz. This is a file of type “mpg” (3.6 Mb).]

4. Wavelet spectra and spatial coherence

When the turbulent eddies responsible for wind noise are within the inertial subrange, as is normally the case, the wind noise spectrum should have a power-law dependence $f^{-7/3}$ or $f^{-5/3}$, depending on whether the dominant pressure disturbances are intrinsic to the turbulence or due to flow stagnation.⁵ This behavior is expected for frequencies low enough that the responsible eddies are much larger than the windscreen. At higher frequencies, when the responsible eddies are smaller than the windscreen, a faster decay is expected since the microphone senses an average of the pressure disturbances over the surface of the windscreen.^{6,7} Heuristically this averaging introduces an additional f^{-2} diminishment,⁷ so the wind noise should decay as $f^{-13/3}$ or $f^{-11/3}$ in this region.

In the context of wavelet analysis, we can examine these relationships by calculating the *global wavelet spectrum*, which is the average of the wavelet spectrum over time.⁴ It is analogous to the Fourier power spectrum. To obtain good spectral resolution, we calculate the global wavelet spectrum with the DOG4 wavelet, which has better spectral resolution than DOG1. Empirically, we found that DOG1 and DOG2 sometimes smooth away the spectral relationships of interest, whereas DOG4 and higher provide fairly consistent results. The spectra were calculated from the CWTs of the measured sound pressures in Pa. The wavelet amplitudes were then divided by the standard reference pressure ($20 \mu\text{Pa}$) and converted to dB. Hence they represent sound pressure level filtered to the passband of the wavelets.

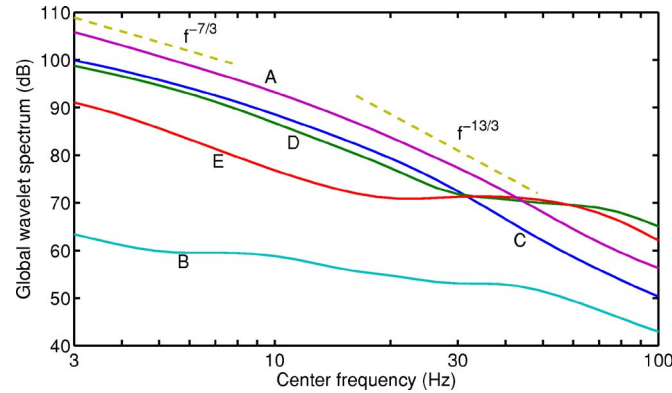


Fig. 3. (Color online). Global wavelet spectra (dB re $20 \mu\text{Pa}$) corresponding to the five cases in Table I. The analyzing wavelet was DOG4.

Wavelet spectra for the center array microphone for each of these five cases are shown in Fig. 3. Except for the extremely low wind case (E), the spectra consistently display a slope of about $f^{-7/3}$ in at low frequency. For frequencies above several Hz, there is a transition to a steeper slope. According to van den Berg⁷ this transition should occur near the frequency $f_c = S_c V/D$, where $S_c \approx 1/3$ is a critical Strouhal number, D the windscreen diameter, and V the mean wind speed. The data appear to be consistent with this prediction. At frequencies above 20–30 Hz, the spectra for cases D and E cease to decay as in cases A and C. This is due to the acoustic energy present in these cases (loud music and a propane cannon, respectively). A slight inflection in the spectrum is also evident for cases A and C at 70–80 Hz. This is likely indicative of acoustic background noise, but at a much lower level than in cases D and E.

Since the turbulent eddies propagate much more slowly than the sound waves, the noise they produce is essentially independent of the direction to which a conventional delay/sum beamformer is steered [e.g., Fig. 2(i)]. Hence the noise suppression generally attainable with the microphone array (at a particular moment in time) can be characterized as the coherent average of the individual signals without introducing delays for different steering directions. At one extreme, if the wind noise happens to be the same at all microphones, the coherent average would have the same average power as any of the individual microphones, and there would be no noise suppression. If the signals are completely uncorrelated, it can be shown that the expected power in the coherent average is $1/N$ times the expected power at any single microphone. Hence noise suppression should vary from 0 dB to $10 \log_{10} N$ dB, depending on the spatial correlation of the wind noise.

Figure 4 shows the noise suppression for several array configurations, as estimated from the global wavelet spectrum of the coherent array average divided by the global wavelet spectrum of the center microphone. The DOG1 wavelet was used. The configurations considered are the entire 49-element array, a square four-element array with side length 0.304 m (taken from the four microphones diagonal from the center microphone), and a square four-element array with side length 0.914 m (taken from the four corners of the array). Based on the previous discussion, the maximum noise suppression $10 \log_{10} N$ dB is 16.9 dB for the 49-element array and 6.0 dB for the four-element arrays. The best performance of the 49-element array occurs around 12 Hz, where 13 dB suppression is obtained. For lower frequencies, spatial correlation of the turbulence diminishes the suppression. At higher frequencies, particularly for case D (with music), the spatial correlation increases due to the presence of sound waves. For the four-element arrays, suppression is close to 6 dB in the mid-frequency range. At lower frequencies, the larger array is less affected by spatial correlation of the turbulence.

5. Conclusions

The data reported here demonstrate that, within a particular frequency range, turbulent pressure disturbances (wind noise) and sound waves have very different spatial structure. The turbulent

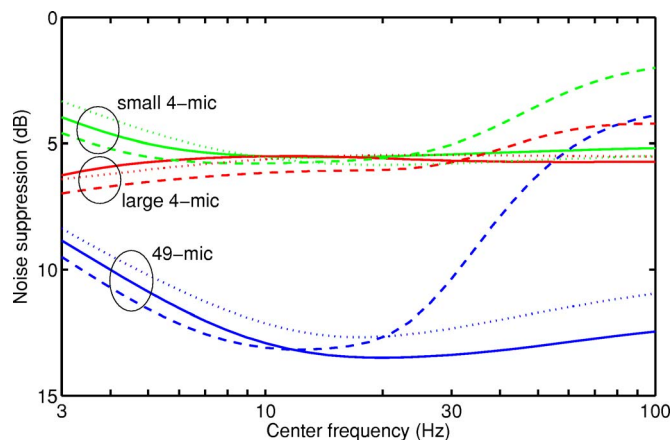


Fig. 4. (Color online). Noise suppression for various array configurations as described in the text. Shown are results for Cases A (dotted), C (dashed), and D (solid). The former two cases are predominantly wind noise, whereas the latter case includes music.

disturbances are smaller, by about two orders of magnitude, than the sound waves. For microphone separations smaller than the size of the turbulent eddies, the wind noise is substantially correlated. Such spatial correlation diminishes the performance of typical beamforming algorithms and is particularly an issue for compact acoustic arrays (those much smaller than the acoustic wavelength). As the microphone separation increases beyond the eddy size, the wind noise becomes nearly uncorrelated. Hence the design of a microphone array for outdoor use should generally take into account the spatial structure of the wind noise as well as the acoustic wavelength range of interest. Perhaps new processing algorithms, specifically designed to separate propagating, nonplane-wave disturbances (wind noise) from propagating plane-wave disturbances (sound waves), could be developed and would improve the performance of compact arrays.

Acknowledgments

Funding was provided by the U.S. Army Basic Research In-House Laboratory Independent Research (ILIR) Program. The authors recognize the outstanding efforts of David Fisk and Douglas Punt (ERDC-CRREL) and Jefferey Mifflin (ERDC-CERL) in constructing the array and performing the field measurements. We also thank Richard Raspet (Univ. of Miss.) for many helpful discussions.

References and links

- ¹S. Morgan and R. Raspet, "Investigation of the mechanisms of low-frequency wind noise generation outdoors," *J. Acoust. Soc. Am.* **92**, 1180–1183 (1992).
- ²H. E. Bass, R. Raspet, and J. O. Messer, "Experimental determination of wind speed and direction using a three microphone array," *J. Acoust. Soc. Am.* **97**, 695–696 (1995).
- ³F. D. Shields, "Low-frequency wind noise correlation in microphone arrays," *J. Acoust. Soc. Am.* **117**, 3489–3496 (2005).
- ⁴C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bull. Am. Meteorol. Soc.* **79**, 61–78 (1998).
- ⁵R. Raspet, J. Webster, and K. Dillion, "Framework for wind noise studies," *J. Acoust. Soc. Am.* **119**, 834–843 (2006).
- ⁶Z. C. Zheng and B. K. Tan, "Reynolds number effects of flow/acoustic mechanisms in spherical windscreens," *J. Acoust. Soc. Am.* **113**, 161–166 (2003).
- ⁷G. P. van den Berg, "Wind-induced noise in a screened microphone," *J. Acoust. Soc. Am.* **119**, 824–833 (2006).

Perceiving the speech of multiple concurrent talkers in a combined divided and selective attention task

Valeriy Shafiro

*Rush University Medical Center, Department of Communication Disorders and Sciences, 203 Senn, 1653 W. Congress Parkway, Chicago, Illinois 60612, USA
valeriy_shafiro@rush.edu*

Brian Gygi

*East Bay Institute for Research and Education, 150 Muir Road, Martinez, California 94553, USA
bgygi@ebire.org*

Abstract: Detection and identification of concurrently spoken key words was investigated using the Coordinate Response Measure corpus. On every trial, listeners first had to explicitly detect callsign keywords in a multi-talker stimulus (divided attention), and, if all callsigns were present, identify the color and number words produced by one of the talkers (selective attention). Increasing the number of concurrent talkers and the number of callsigns to be detected each had a marked negative effect on detection and identification performance. These findings indicate that, when memory involvement is limited, listeners cannot reliably detect more than two concurrently spoken words in diotic listening.

© 2007 Acoustical Society of America

PACS numbers: 43.66.Dc, 43.66.Lj, 43.71.Sy [QJF]

Date Received: July 27, 2007 **Date Accepted:** September 14, 2007

1. Introduction

In many communication settings a listener may need to attend to the speech of multiple concurrent talkers for potentially relevant information (e.g., an air traffic controller listening to the calls of several pilots at once). Alternatively, at times, a listener may need to attend to the speech of only one of several concurrent talkers. Despite a long-standing research interest in the factors involved in these two tasks, known as divided and selective attention tasks, respectively (e.g., in [Spelke *et al.*, 1976](#); [Broadbent, 1958](#)), their full understanding is still lacking.

In recent years, divided and selective attention have been increasingly studied using the Coordinate Response Measure (CRM) corpus ([Bolia *et al.*, 2000](#); [Kidd *et al.*, 2005](#), [Humes *et al.*, 2006](#), [Gallun *et al.*, 2007](#), [Shinn-Cunningham and Ihlefeld, 2004](#)). This corpus combines a large number of naturally spoken utterances with easily measurable performance levels. It has become a useful tool for investigating a variety of practical and theoretical questions in situations closely approximating many real-life tasks. The corpus consists of utterances in the form of “Ready (callsign) go to (color) (number) now.” There are eight callsigns, four colors, and eight numbers. All 256 possible utterances in the corpus are spoken by eight talkers: four males and four females. Listeners are usually asked to identify the color and number spoken by one or more target talkers, typically identified by a specific callsign keyword.

In CRM-based studies, selective attention typically refers to the ability to identify the color and number key words spoken by a single target talker. There are, however, differences in the methods used to assess divided attention. In [Brungart *et al.* \(2001\)](#) divided attention referred to listeners’ abilities to identify a target talker in a mixture of concurrent talkers based on a single callsign key word. [Shinn-Cunningham and Ihlefeld \(2004\)](#) defined divided attention as listeners’ ability to report the color and number key words from two concurrent talkers, regardless of the callsign. Others ([Kidd *et al.*, 2005](#); [Humes *et al.*, 2006](#), [Gallun *et al.*, 2007](#)) examined

divided attention in a condition where listeners were asked to report the color-number coordinates of a single talker in a mixture of talkers, but received the cue needed to identify the target talker only after the offset of the stimulus.

Performance in these and most other CRM-based studies is measured based, almost exclusively, on the color-number identification accuracy. However, this obliterates the distinction between the ability to detect the target callsign in a mixture and the ability to identify color and number key words, even though the success of the latter is based on the success of the former (unless the target talker is identified by location or vocal characteristics). In addition, the previous CRM-based studies of divided attention differ substantially in the load placed on the working memory, with those used in [Brungart *et al.* \(2001\)](#) being the least and [Kidd *et al.* \(2005\)](#) and [Humes *et al.* \(2006\)](#) the most demanding. Thus, performance on a divided attention task may vary based on the capacity and availability of a memory buffer which is used to temporarily store the information ([Cowan, 2001](#); [Gallun *et al.*, 2007](#)).

The present study introduces a task in which both divided and selective attention components of a CRM-based task could be examined on the same trials. It explicitly differentiates between the ability to detect a target callsign from the ability to identify color and number key words, making it possible to manipulate divided attention separately from selective attention. This provides an opportunity to assess performance costs associated with changing the divided attention load on both callsign detection and color-number identification. Furthermore, the rapid nature of callsign detection in the course of a single trial limits the listener's ability to store callsign information for later processing and minimizes memory involvement.

2. Method

2.1 Stimuli and design

The stimuli were based on the CRM corpus described above. The divided attention component of CRM was measured as sensitivity to the callsign target words (d'), estimated from listeners' ability to distinguish trials that contained one or more target callsigns from those that did not. The divided attention load was manipulated by changing the number of callsign target words listeners had to detect in a stimulus from one to three. Selective attention was measured using overall accuracy of identifying color and number key words spoken by a single target talker.

There were three separate test sessions that differed in the number of callsign targets that listeners were required to detect. In each of the three sessions one-half of the stimuli contained the target callsign(s), while the other half of the stimuli did not. In all three sessions listeners were instructed to identify the color and number key words spoken by the talker producing the callsign 'Baron,' but only on those trials when they could hear every callsign specified as a target callsign for that session. In the first test session (1T), listeners were asked to detect only a single callsign target ('Baron') in each stimulus. One half of the stimuli contained one talker saying 'Baron,' while the other half of the stimuli did not contain any utterances with this callsign. In the second session (2T), the target callsigns were 'Baron' and 'Hopper.' Listeners were instructed to identify the color and number spoken by the talker saying 'Baron,' but only on those trials when they could hear both 'Baron' and 'Hopper.' One half of the stimuli in that session contained both target callsigns, while the other half of the stimuli always contained one utterance with the callsign 'Baron,' but no utterance with the callsign 'Hopper.' In the third session (3T), listeners were again instructed to identify the color and number spoken by the talker saying 'Baron,' but to do it only when they could hear all three callsigns: 'Baron,' 'Hopper' and 'Tiger.' Foil stimuli in this session always contained one utterance with the target callsign 'Baron,' and one other utterance with either 'Hopper' or 'Tiger,' but not both. This composition of foil stimuli was implemented to ensure that listeners detect all target callsigns. The masking utterances in all three sessions contained any of the five remaining callsigns in the corpus.

The number of talkers and talker sex configuration were also manipulated in the stimuli for each test session. There were eight different talker configuration conditions in the 1T and 2T sessions and six talker configuration conditions in the 3T session, which did not contain

Table 1. Talker sex configurations used in the study. Legend: 1T, 2T, and 3T are the first, second and third experimental sessions, that had one, two, or three target callsigns, respectively; T=target ('Baron') talker; S/D =same/different sex talker than the target talker; parentheses mark the talkers producing callsign targets.

No. of Talkers	2		3		4		3		4		2		3		4		
Session	2	3	4	3	4	2	3	4	2	3	4	2	3	4	2	3	4
1T	(T)D	(T)DD	(T)DDD	(T)SD	(T)SDD	(T)S	(T)SS	(T)SSS									
2T	(TD)	(TD)D	(TD)DD	(TS)D	(TS)DD	(TS)	(TS)S	(TS)SS									
3T	...	(TDD)	(TDD)D	(TSD)	(TSD)D	...	(TSS)	(TSS)S									

two-utterance conditions. The number of talkers in each stimulus was varied across conditions from two to four. The differences among the talker sex configuration conditions were classified based on the coding system developed by Brungart *et al.* (2001), as shown in Table 1. The letter "T" represents the target ('Baron') talker whose color and number words listeners had to identify in each of the three sessions. On half of the trials this talker was male and on half of the trials female. The letter "S" represents a different talker of the same sex as the target talker, and the letter "D" represents a different talker of a different sex than the target talker. The sex configuration of the talkers producing the target callsigns in each session is indicated by parenthesis. For example, a three-utterance stimulus in session 1T with a single target callsign 'Baron,' in which one masking talker was of the same sex and one was of a different sex than the 'Baron' talker, was coded (T)SD. However, a stimulus with a similar overall talker sex configuration in the double-target session 2T, was coded as (TS)D, indicating that the talker producing the second target callsign 'Hopper' was of the same sex as the 'Baron' talker, while a single masking talker was of a different sex than the 'Baron' talker. Male and female talkers were equally represented across all talker-configuration conditions. Each talker-configuration condition contained the total of 160 stimuli: 80 stimuli with the target callsigns and 80 foil stimuli. All utterances in a stimulus had equal root mean square (rms) energy. To eliminate any differences among talkers that may result from spatial separation, diotic stimulus presentation was used, as in Brungart *et al.*, 2001. During presentation, the overall level of each stimulus was also randomly roved in a 6 dB range in 1 dB steps.

2.2 Listeners and procedure

Seven paid female university students (mean age 22 years, range 18–25) served as listeners. All were highly familiar with the task and procedure and had about 10 h of prior experience with this task, such that an asymptotic performance was reached for both callsign detection and color-number identification with single target callsign stimuli. All had passed a hearing screening and had thresholds lower than 25 dB hearing level at 250, 500, 1000, 2000, and 4000 Hz. Each session lasted approximately 2–2.5 h.

The listeners were seated in front of a 17 in. liquid crystal display monitor in an acoustically treated sound booth. Stimuli were presented diotically through circumaural Sennheiser HD250II linear headphones at 60–70 dB sound pressure level (the range is due to the roving stimulus level). They responded to each stimulus using a graphical user interface that consisted of an 8 × 4 matrix of colors and numbers displayed on the screen where each row represented a number and each column a color. The numbers were colored according to their appropriate column. The corresponding color was also written next to each number. The listeners used a mouse to select the color-number response corresponding to that of the target utterance. Stimulus presentation was listener paced. In the beginning of each session, each listener performed 40 practice trials that were not scored. Prior to each session, listeners were informed about the percentage of foil stimuli included and were instructed to make a color-number response only when they heard all target callsigns in the stimulus, and press a button labeled "No Target Present" otherwise. A response was scored as a 'hit' when a listener made a color-number iden-

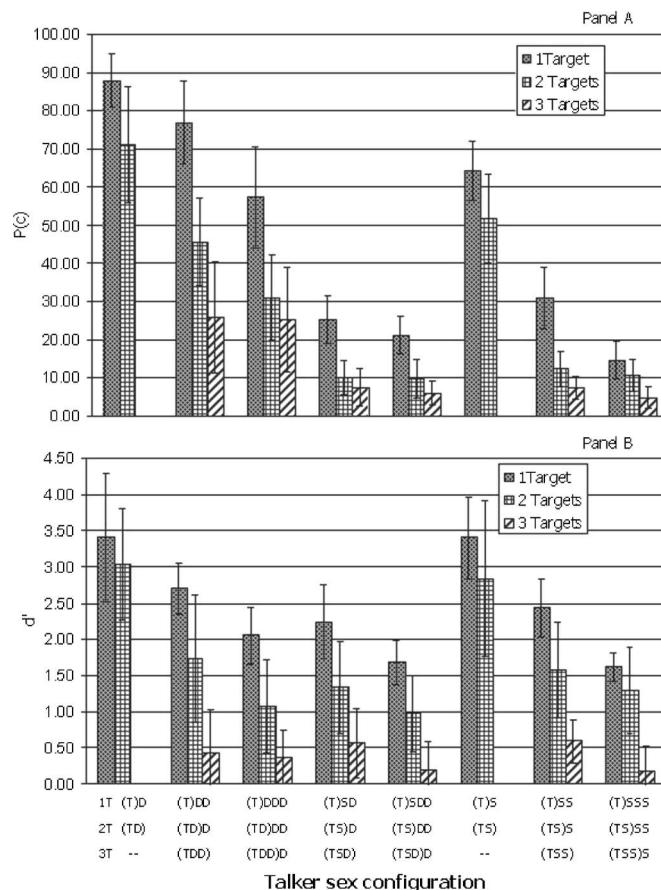


Fig. 1. The mean identification accuracy (Panel A) and target callsign sensitivity (Panel B) in each talker-sex configuration condition for each test session.

tification response (whether correct or incorrect) on a trial that contained all target callsigns for that session. A response was scored as a ‘false alarm’ when a color-number response was made on a trial that did not contain all target callsigns. Detection sensitivity (d') was computed based on the proportion of hits and false alarms, converted to z scores. The value of d' was measured as the difference between the z scores for hits and the z scores for false alarms in each talker configuration condition within each listener (Macmillan and Creelman, 2005).

3. Results

Listener color-number identification accuracy ($P(c)$) and callsign detection sensitivity (d') were compared across the three sessions of the present experiment (Fig. 1). Analysis of variance on arcsine-transformed percent correct color-number identification responses revealed a main effect of the number of target callsigns ($F(2, 12) = 68.19, p < 0.001$) and talker configuration conditions ($F(5, 30) = 154.09, p < 0.001$). There was also an interaction between these factors ($F(10, 60) = 13.348, p < 0.001$). Mean accuracy of color-number identification declined in each of the talker configuration conditions as the number of callsign targets increased (mean $P(c)$ of 50%, 34%, 11% correct for one-, two-, and three-target callsign conditions, respectively). The identification accuracy also declined as the number of talkers in the stimulus increased (mean $P(c)$ of 69%, 27%, 19% correct, for two-, three-, and four-talker conditions, respectively). Thus, both of these factors had a marked affect on identification accuracy. Furthermore, an increase in

the number of talkers was associated with a decrease in the accuracy range. Across the six talker configuration conditions with three or four talkers, the accuracy range decreased from 15% to 77% in 1T, 10%–46% in 2T, and 5%–26% in 3T. This finding indicates that number of talkers and talker sex configuration that differentiated these conditions had a consistently smaller influence on the overall performance as the number of target callsigns that had to be detected increased. Nevertheless, the general pattern of effects of talker sex and number were similar across the test sessions. When the sex of the ‘Baron’ talker was different from all other talkers in the stimulus, color-number identification performance was the highest in every session, regardless of the number of the callsign targets.

The pattern of listener sensitivity to callsign target(s) across talker configuration conditions was similar to the pattern of identification accuracy for the single- and the double-target sessions (Pearson product-moment $r=0.75$, and $r=0.54$, respectively, $p<0.01$). However, no significant association between accuracy and sensitivity was found in the third 3T session, possibly due to floor effects. Similar to identification accuracy, there was a main effect of the number of target callsigns, i.e., listener sensitivity to target callsigns significantly declined ($F(2, 12)=54.65$, $p<0.001$) as the number of callsign targets increased. The mean d' across the six comparable talker configuration conditions (with three and four talkers only) were 2.12, 1.33 and 0.39 in 1T, 2T and 3T, respectively. There was also a main effect of talker sex configuration ($F(1, 6)=50.90$; $p<0.001$), and an interaction between talker sex configuration and the number of target callsigns ($F(10, 60)=3.50$, $p<0.001$). Similar to identification accuracy, callsign detection consistently decreased as the number of talkers in the stimulus increased (i.e., mean d' of 3.18, 1.52, and 1.03 for two-, three-, and four-talker conditions, respectively).

The results also indicate some cross-listener variability in the ability to process stimuli with an increased divided attention load. Overall, there was a greater variability in the range of callsign detection performance in the double and triple-target sessions (individual listeners' mean d' ranged 0.59–2.19 and 0.05–1.06, respectively) than in the single-target session (1.86–2.64). In the double-target session, two of the seven listeners had mean d' values below one (0.82, 0.59, or z score of 1.2 and 1.6, respectively), even though their color-number identification accuracy was within the mid range of the other participants. This disparity between their identification accuracy and detection performance may indicate that they continued to detect primarily only a single target callsign ‘Baron’ when making their responses, and paid less attention to other target callsigns. On the other hand, in the triple-target session one listener's detection sensitivity was above one ($d'=1.06$, or z score of 2), while the rest of the participants' d' were less than 0.45. The same listener also had the highest callsign detection sensitivity in the double-target session ($d'=2.19$, or z score of 1), while her identification accuracy remained close to the mean for that session.

4. Discussion

These results demonstrate that increasing the divided attention load by varying the number of callsign target words listeners need to detect in the stimulus has a twofold effect. First, it results in a progressive decline in listeners' ability to detect concurrently presented target callsigns, with performance reaching near chance level with three callsign targets. Second, it leads to a significant decrease in listeners' ability to identify the color and number keywords spoken by one talker. This second result is likely to be the direct consequence of the first because accurate color-number identification requires accurate detection of the voice of the target talker based on the callsign information. Thus, having to detect additional callsigns in the beginning of the stimulus severely limits listeners' abilities to selectively attend to the voice of a single (‘Baron’) talker.

Several current attention models may be able to explain this finding. First, callsign detection involves higher order cognitive processes because it requires speech processing and lexical access. Thus, concurrently presented multiple callsigns may be simultaneously competing for the limited processing resources required for accurate detection (Bonnell and Hafter, 1998; Gallun *et al.*, 2007). As the number of callsigns increases to three the processing re-

sources are depleted and performance falls to near chance levels. One implication of this hypothesis is that when the divided attention task does not involve the same processing resources (e.g., if listeners were required to detect a specific talker rather than an additional callsign), there should be little if any decrease in performance.

Second, the results may indicate working memory limitations. Due to the rapid online nature of the task, listeners' attention had to be switched to the color and number words spoken by the target ('Baron') talker immediately after making a decision about the presence of the target callsigns in the stimulus. Thus, a decline in sensitivity with increasing number of callsign targets may be indicative of a processing bottleneck that results from the inability to perceive the color and number words of one talker, while simultaneously examining a stored image of the target callsign part of the stimulus. Alternatively, this can also be explained as an inability to obtain and store an accurate image of the whole stimulus (either in sensory or processed form) sufficiently long enough for a correct response. That is, the drop in target sensitivity with an increasing number of callsign targets may indicate that the listeners did not have sufficient time or processing resources to search the signal for all of the target callsigns and to determine the vocal characteristics of the 'Baron' talker, or that the information in the memory image was more degraded as the number of talkers in the stimulus increased. On the other hand, if listeners could perceive the vocal characteristic of the target 'Baron' talker, they might not have been able to detect the other target callsigns and switch their attentional focus to the voice of that talker in time to be able to detect the color and number words.

The present experiment has quantitatively demonstrated the performance costs associated with increasing divided attention load for the divided and selective attention components of a single CRM-based task. It further confirmed the roles of several factors such as the number of talkers and their physical similarity (sex configuration) that have been shown to affect divided and selective attention (Kidd *et al.*, 2005, Humes *et al.*, 2006, Brungart *et al.*, 2001, Shinn-Cunningham and Ihlefeld, 2004). However, in the present design, increasing the number of talkers also led to the progressive decrease in the signal-to-noise ratio (SNR) because of equal rms energy across all utterances in each stimulus. For example, the SNR for each target utterance in two-talker stimuli was 0 dB, while the SNR for each target utterance in four-talker stimuli was -9.5 dB. The additional masking energy associated with additional talkers could also contribute to the decrease in performance. Overall, the number of talkers (and associated SNR changes) and the number of targets had a similar negative effect on performance accuracy and sensitivity. However, as Fig. 1 illustrates, performance consistently decreased with an increased number of target callsigns even for the stimuli with the same number of talkers and talker-sex configuration, suggesting some independence of the number of target callsign effects from the other factors.

Strong correlations between target callsign sensitivity and identification accuracy suggest that both divided and selective attention may be affected by similar stimulus parameters. However, these correlations may also reflect the structure of the CRM task in which selective attention performance (color-number identification) is dependent on the divided attention performance (target callsign detection). To explore this possibility further, the divided and selective attention components of the task should be examined independently in future studies. The results also show cross-listener variation in performing the divided attention task, which became more pronounced as the divided attention load increased to three callsigns. This is consistent with previous findings on auditory divided attention tasks (Yost *et al.*, 1996; Humes *et al.*, 2006), as well as with the larger body of individual differences literature which shows that individual differences become more pronounced with increasing task difficulty (Dillon and Watson, 1996). Finally, diotic listening conditions employed in the current study prevented listeners from using spatial cues that have been demonstrated to aid the listener in perceptually segregating each talker's information (Cherry, 1953; Yost *et al.*, 1996; Shinn-Cunningham and Ihlefeld, 2004; Kidd *et al.*, 2005). The use of such cues may be expected to have a positive effect on listeners' ability to divide attention when attempting to perceive multiple concurrently spoken key words. The effects of these factors need to be explored further before a more general model of divided and selective attention costs can be established.

Acknowledgments

We wish to thank Dr. Hillary Snapp and Blythe Holmes for their help with listener testing, Dr. Barbara Shinn-Cunningham for helpful comments on a previous version of the manuscript, and Drs. Neil Macmillan and Jim Jenkins for their advice on several design and analysis issues in this study.

References and links

- 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.
- Bonnel, A. M., and Hafter, E. R. (1998). "Divided attention between simultaneous auditory and visual signals," *Percept. Psychophys.* **60**, 179–190.
- Broadbent, D. E. (1958). *Perception and Communication* (Pergamon, New York).
- 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.
- Cherry, C. (1953). "Some experiments on the recognition of speech with one and with two ears," *J. Acoust. Soc. Am.* **26**, 975–979.
- Cowan, N. (2001). "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," *Behav. Brain Sci.* **24**, 87–185.
- Dillon, A., and Watson, C. (1996). "User analysis in HCI: The historical lesson from individual differences research," *Int. J. Hum.-Comput. Stud.* **45**, 619–638.
- Gallun, F. J., Mason, C. R., and Kidd, G. R. J. (2007). "Task-dependent costs in processing two simultaneous auditory stimuli," *Percept. Psychophys.* **69**, 757–771.
- 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. R. J., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (2005). "The advantage of knowing where to listen," *J. Acoust. Soc. Am.* **118**, 3804–3815.
- Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide*, 2nd ed. (Erlbaum, Mahwah, NJ).
- Shinn-Cunningham, B. G., and Ihlefeld, A. (2004). "Selective and divided attention: Extracting information from simultaneous sound sources," in *Proceedings of the 10th Annual International Conference on Auditory Display (ICAD)*, Sidney Australia.
- Spelke, E. S., Hirst, W., and Neisser, U. (1976). "Skills of divided attention," *Cognition* **4**, 215–230.
- Yost, W. A., Dye, R. H. J., and Sheft, S. (1996). "A simulated 'cocktail party' with up to three sound sources," *Percept. Psychophys.* **58**, 1026–1036.

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.

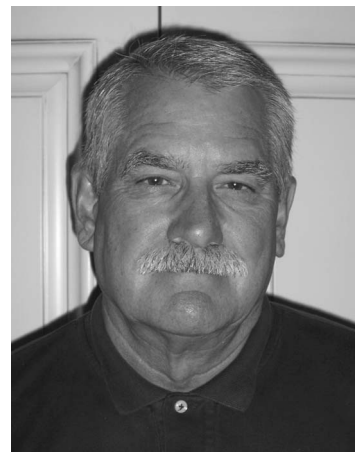
New Fellows of the Acoustical Society of America



William A. Ahroon—For contributions to communications in hostile acoustic environments.



Dimitry Donskoy—For contributions to physical, structural, and underwater acoustics.



William T. Eillison—For contributions to the effects of sound on marine organisms.



Ronald L. Eshleman—For contributions to standards in shock and vibration.



Anthony P. Lyons—For contributions to the effects of fine scale seafloor properties.



Masauuki Morimoto—For contributions to human spatial hearing.

Jennifer Cornish is the recipient of the 2006 Stetson Scholarship



ASA member Jennifer Cornish was selected the recipient of the 2006–07 Raymond H. Stetson Scholarship in Phonetics and Speech Production. At the time she was selected, Jennifer Cornish was a graduate student in the Department of Linguistics at the University at Buffalo.

She received a B.A. from Oberlin College and an M.A. from the University at Buffalo. Her current work deals with linguistic issues related to speech production.

The Stetson Scholarship, which was established in 1998, honors the memory of Professor Raymond H. Stetson, a pioneer investigator in phonetics and speech science. Its purpose is to facilitate the research efforts of promising graduate students and postgraduates. The scholarship includes a \$3000 stipend for one academic year.

Applications for the award are due in March each year. For further information about the award, please contact the Acoustical Society of America, Suite 1N01, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; Web: http://asa.aip.org.

neer investigator in phonetics and speech science. Its purpose is to facilitate the research efforts of promising graduate students and postgraduates. The scholarship includes a \$3000 stipend for one academic year.

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

- 29 June–4 July Joint meeting of the Acoustical Society of America, European Acoustics Association and the Acoustical Society of France, Paris, France [Acoustical Society of America, Suite 1N01, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
- 28 July–1 Aug 9th International Congress on Noise as a Public Health Problem (Quintennial meeting of ICBEN, the International Commission on Biological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, Post Office Box 1609, Groton, CT 06340-1609, Tel: 860-572-0680; Web: www.icben.org. E-mail: icben2008@att.net.
- 10–14 Nov 156th Meeting of the Acoustical Society of America, Miami, Florida [Acoustical Society of America, Suite 1N01, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; Web: http://asa.aip.org].

2009

- 18–22 May 157th Meeting of the Acoustical Society of America, Miami, Florida [Acoustical Society of America, Suite 1N01, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; Web: http://asa.aip.org].

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 1N01, 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

New Associates

- Akal, Tuncay, SUASIS-Sualti Sistemleri, Research and Development, TUBITAK-MAM Teknoloji Serbest, GEBZE, 41470 Kocaeli, Turkey
- Alam, Sheikh Mahbub, Blk. 631, #14-268, Jurong West St. 65, Singapore 640631, Singapore
- Alpha, Curtis I., General Dynamics Electric Boat, 462, 7 Aberdeen Ct, East Lyme, CT 06333
- Anderson, Jim N., New York Univ., Clive Davis., Dept. of Recorded, 194 Mercer St., Rm. 501, New York, NY 10012
- Andruske, Linda L., 3799 S. Banana River Blvd., # 906, Cocoa Beach, FL 32931
- Antunes, Jose V., Inst. Technologico Nuclear, Applied Dynamics Lab., Estrada Nacional 10 Sacavem 2686, Portugal
- Arango, Renzo, 5/749 Pittwater Rd., Dee Why, NSW 2099, Australia
- Backun, Morrie S., Backun Musical Services, 6875 Royal Oak Ave., Burnaby, BC V5J 4J3, Canada
- Baillie, Cameron J., Shiner and Associates, Inc., 225 West Washington St., Suite 1625, Chicago, IL 60177
- Bauer, Juergen, Waterford I.T., Dept. of Architecture, Cork Rd. Campus, Waterford, Ireland
- Brenneman, Robert M., McKay Conant Brook, Inc., 7435 East Stetson Dr., Suite A, Scottsdale, AZ 85251
- Bull, Chris M.S., 601 Westminster Green, 8 Dean Ryle St., London, SW1P 4DA, United Kingdom
- Cammarata, John, 9830 Britinay Ln., Baltimore, MD 21234
- Campbell, Paul A., 19 St. Ternans Rd., Newtonhill, Aberdeen, AB39 3PF, Scotland
- Danworaphong, Sorasak, Walailak Univ., Science, 222 Thaiburi, Thasala, Nakhon Si Thammarat 80160, Thailand

- DeCoensel, Bert, Acoustics Group, Dept. of Information Technology, Ghent Univ., St. Pietersnieuwstraat 41, Ghent, Flanders 9000, Belgium
- Desart, Eric, Columbiastraat 173/1, 2030 Antwerp, Belgium
- Desmet, Wim, Katholieke Universiteit Leuven, Mechanical Engineering, Celestijnenlaan 300B, 3001 Leuven, Belgium
- Diaquila, Samuel V., Sound Solutions Const. Services, BASWAphon Acoustical Finish, 3900 Ben Hur Ave., Ste. 10, Willoughby, OH 44094-6398
- Dillion, Kevin, Miltec Research and Technology, NCPA, Coliseum Dr., Rm. 1069, University, MS 38677
- Drake, Robert M., P.O. Box 5029, Newport, RI 02841
- Drysdale, Dale T., Rimkus Consulting Group, Inc., Environmental, 8 Greenway Plaza, Ste. 500, Houston, TX 77046
- Emelianov, Stanislav, Univ. of Texas at Austin, Biomedical Engineering, 1 University Station C0800, Austin, TX 78712
- Ferk, Heinz, Graz Univ. of Technology, Laboratory of Bldg. Physics, Inffeldgasse 24, A-8010 Graz, Austria
- Frizzell-Makowski, Linda J., National Security Technology Dept., Johns Hopkins Univ., Applied Physics Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723
- Fulcher, Lewis P., Physics and Astronomy, Bowling Green State Univ., 104 Overman Hall, Bowling Green, OH 43403
- Gagesch, Jennifer B., 509 Newberry Ave., La Grange Park, IL 60526
- Gilstrap, Buddie P., 12778 Rancho Penasquitos Blvd, Apt. 131, San Diego, CA 92129
- Gudgel, Judson M., 11729 S. Braden Ave., Tulsa, OK 74137
- Harwell, Ross, 5 Smithfield Dr., Toronto, ON M8Y 3L8, Canada
- Johnson, Roger, 6137 Fullerton Rd., Stevensville, MI 49127
- Jones, Benjamin A., 118 S. Kainalu Dr., Kailua, HI 96734
- Kemp, Jamie, 4490 S. Wright Ct., Morrison, CO 80465
- Krane, Michael, Fluids Research, Applied Research Lab., Pennsylvania State Univ., P.O. Box 30, State College, PA 16804
- Kurtz, Russell, Physical Optics Corporation, 20600 Gramercy Pl., Bldg 100, Torrance, CA 90501
- Madison, Ted K., 3M Company, Occ Health & Environ. Safety, 3M Center, Bldg. 0235-02-E-91, Saint Paul, MN 55144-1000
- Magee, Joshua A., 44 Arnold Way, West Hartford, CT 06119
- Mariappan, Vivek, AMEC, Air and Noise Group, 4810 93 St., Edmonton, AB T6E 5M4, Canada
- Masoy, Svein-Erik, NTNU, Circulation and Medical Imagin, Olav Kyrresgate 9, N-7489 Trondheim, Norway
- Mazuka, Reiko, RIKEN Brain Science Inst., Lab. for Language Development, 2-1 Hirosawa, Wako-shi, Saitama, 351-0198, Japan
- Miller, David K., AVT, Inc., 1031 McLaughlin Run Rd., Bridgeville, PA 15017
- Mimran, Sarah, UltraShape Ltd., Research and Development, POB 80, Kochav Yokneam, 20692, Israel
- Monroe, John N., P.O. Box 6508, Charlottesville, VA 22906
- Mozley, Edward C., 5060 Bella Collina, Oceanside, CA 92056
- Neill, John M., U.S. Greenfiber LLC, Fire and Sound Products, 2500 Distribution St., Ste. 200, Charlotte, NC 28203
- Newton, Jr., Gary F., Bruel and Kjaer, Automotive Research Center, 6855 Commerce Dr., Canton, MI 48187
- Ollos, Ryan J., Applied Physics Lab, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698
- Orfali, Wasim A., Nordmannzeile, Berlin 12157, Germany
- Osborn, Alvadore J., Electronic Sound Company, 691-39th St., Des Moines, IA 50312
- Pluymers, Bert, Katholieke Universiteit Leuven, Mechanical Engineering, Celestijnenlaan 300B, 3001 Leuven, Belgium
- Quesenberry, Damon C., 7650 Corporate Center Dr., Ste. 401, Miami, FL 33126
- Rice, John M., Rice Associates Inc., 31505 Grand River Ave., Ste. 8, Farmington, MI 48336
- Smith, Nicholas A., Boys Town National Research Hospital, 555 N 30th St, Omaha, NE 68131
- Styles, Jon, JPStyles Limited, P.O. Box 37 857, Parnell, Auckland 1151, New Zealand
- Thomas, David J., 224 Harvard Ave., E #301, Seattle, WA 98102
- Tsurugaya, Yoshiaki, 11-14 Ohi 5-Chome, Shinagawa, Tokyo 180-0014, Japan
- Wanderley, Marcelo M., Music Research, McGill Univ., 555 Sherbrooke St., West Montreal QC H3A 1E3, Canada
- Weiss, Susan F., Peabody Conservatory of JHU, Musicology, 1 E. Mt. Vernon Pl., Baltimore, MD 21202
- Wong, Jasper K., Xerox, 26600 SW Parkway, MS 7060-431, Wilsonville, OR 97202
- Wong, Lloyds, Good Connection Co., Ltd., Administration, 10F, No. 36, Song De Rd., Taipei 110, Taiwan
- Yazdannivaz, Amir M., Environmental and Architectural Acoustics, PCR Services Corporation, 233 Wilshire Blvd., Ste. 130, Santa Monica, CA 90403

New Students

- Agius, Rudi, 'PATAGONIA' Vjal il-Qalbiena, Mostin, Mosta, MST05, Malta
- Anderson, Phillip A, 57 Oak Square Ave., Brighton, MA 02135
- Baese, Melissa M., Linguistics, Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208
- Baker, Rachel E., Linguistics, Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208-4090
- Burd, David R. 108 S. Pine St., Rolla, MO 65401
- Burnett, William A. P.O. Box 1481, Castle Rock, CO 80104
- Chang, Charles B., Dept. of Linguistics, Univ. of California, Berkeley, UC Berkeley-Linguistics, 1203 Dwinelle Hall, Berkeley, CA 94720-2650
- Coutinho, Agostinho S., R. do Molhe, 575 R/C, Porto, Portugal
- Dall'Osto, David, 914 N. 38th St., Seattle, WA 98103
- Davis, Ben, 1321 New Castle Rd., Apt. H15, Durham, NC 27704
- Di Gioia, Antonio, 98 Howard Rd., London, E17 4SQ, United Kingdom
- Dunnachie, Matthew E., iSLI, The Alba Centre, Alba Campus, Livingston, EH54 7EG, Scotland
- Falou, Omar, 3134 Jessica Ct., Mississauga, ON L5C 1X6, Canada
- Fenush, Chelsea D., 5200 Henderson Rd., Erie, PA 16509
- Glaeser, Sharon S., 6523 SE 18th Ave., Portland, OR 97202
- Guldiken, Rasim O., Georgia Inst. of Technology, 251 10th St. NW, F305, Atlanta, GA 30318
- Gundy, Dakota, 714 N. Inst., Colorado Springs, CO 80903
- Guthrie, Anne E., 1926 5th Ave., Troy, NY 12180
- Henderson, Elizabeth, UCSD, SIO, 9500 Gilman Dr., La Jolla, CA 92093-0205
- Herman, Jeffrey L., 2152 Carol Ave., Mountainview, CA 94040
- Hough, Ian R., 311 6th St., Downers Grove, IL 60515
- Johnson, Aaron M., 6 Stone Crest Cir, Madison, WI 53717-1320
- Kaczkowski, Stephen V., 400 McChesney Ave. Ext., Apartment 12-7, Troy, NY 12180
- Kanekama, Yori, Wichita State Univ., Communication Sciences and Disorders, 1845 Fairmount St., Wichita, KS 67260-0075
- Kannaikkal-John, Thomas, 296 Wood St., Preston, Melbourne, VIC 3072, Australia
- Lavender, Ashley L., Bioscience and Biotechnology, Drexel Univ., 3141 Chestnut St., Philadelphia, PA 19104
- Lee, Wu-Jung, MIT/WHOI Joint Program, Applied Ocean Science and Eng., Co-op 202, MS#16, Woods Hole, MA 02543
- Lew, Phoitack, 130 Farms Dr., Burlington, MA 01803
- Liu, Yu Cheng, Mechanical and Aeronautical Eng., Feng Chia Univ., No. 100 Wenhwa Rd., Seatwen, Taichung 40724, Taiwan
- Machado, Daniel A., CALLE 151 #117-61, Apartamento 202, Bogota D.C., Colombia
- Mahon, Matthew G., 2932 Pine Haven Dr., Birmingham, AL 35223
- Manzi, Nicholas J., 20 Concord Square, Apt. 1, Boston, MA 02218
- Meredith, Morgan L., 409 W. Gorham St., Apt. 801, Madison, WI 53703
- Metzler, Adam M., 3 Woodlake Rd. Apt. 3, Albany, NY 12203
- Mishra, Taniya, OGI School of Science and Eng., CSLU, 20000 NW Walker Rd., Beaverton, OR 97006
- Misiak, Gregory M., 625 Cooper Rd., West Berlin, NJ 08091
- Murphy, Molly F., NAU PO Box 20253, Flagstaff, AZ 86011
- Nash, Grant, GETRAG, NVH, Axle Development, 1848 Getrag Parkway, Newton, NC 28658
- Oberti, Stefano, ETH Zurich, Center of Mechanics, ETH-Zentrum - CLAH23.2, Tannenstrasse 3, 8092 Zurich, Switzerland
- Peng, Zhao, 3175 JFK Blvd., Apt. 707, Philadelphia, PA 19104

Pope, David M., 2419 Woodhurst Dr., Huntsville, AL 35803
 Reeve, Lora L., HIMB, Marine Mammal Research, P.O. Box 1346, Kanohe, HI 96744
 Reindl, Klaus J., Unterfeld 2, 84085 Niederleierndorf, Germany
 Richardson, Phillip G.M., 8 Ditton Fields, Cambridge, CB58QN, United Kingdom
 Schmidt, Matthew J., 8651 Ventana Dr., Apt. 3718, Oak Creek, WI 53154
 Schneider, Toby E., 25 Cassway Rd., Woodbridge, CT 06525
 Sherman, Bradley S., 200 Bloomfield Ave., Box 1755, West Hartford, CT 06117
 Smadi, Ahmad T., Authentic Subtitles, Translation, 6599 Sabado Tarde #2, Goleta, CA 93117
 Song, Jae Yung, Brown Univ., Box 1978, Providence, RI 02912
 Sukara, Elizabeth A., 509 Commissioners Rd. West, Unit 433, London ON N6J 1Y5, Canada
 White, Andrew, Earth and Space Sciences, 5515 Wallingford Ave. N., Seattle, WA 98103-5935
 White, Charles E., P.O. Box 1841, North Kingstown, RI 02852
 Wilson, E. Ctenay, MIT, RLE/SHBT, 36-761, 50 Vassar St., Cambridge, MA 02139
 Wright, Lindsay, Box 91473, Wesleyan Station, Middletown, CT 06459
 Yao, Yao, 1203 Dwinelle Hall, UC Berkeley, Berkeley, CA 94720-2650

New Electronic Associates

Arzoumanian, Sevag H., Alion Science, ETC/CAA, 84 Sherman St., Cambridge, MA 02140
 Coleman, Robert J., Electro Acoustic Professionals, 935 NW 31st Ave., Ste. B, Pompano Beach, FL 33069
 Dakin, Tom, Applied Microsystems Ltd., R&D, 2071 Malaview Ave., Sidney, BC V8L1Y4, Canada
 Gruver, Joseph, 3801 14th st. #105, Plano, TX 75074
 Hanrahan, Kevin, Univ. of Nebraska-Lincoln, Music, 239 Westbrook Music Bldg., Lincoln, NE 68588
 Hansen, Rolf K., CodaOctopus, Bergen, Sandviksboder 77C, 5841 Bergen, Norway
 Iannace, Gino, via fievo, 82010 San Leucio Del Sanni, Italy
 Kay, Marshall, Audio Artistry, Inc., 5504 Fallen Leaf Ct., Raleigh, NC 27606
 Koga, Takashi, Kajima Corp., Technical Research Inst., 19-1 Tobia-kyu, 2-chome, Chofu, 182-0036, Japan
 Lagergren, Peter, Xtero Datacom, 8505 Eastlake Dr., Burnaby, BC V5A 4T7, Canada
 Monitto, John, Meyer Sound Laboratories Inc., 2832 San Pablo Ave., San Rafael, CA 94702
 Norman, John, NTSU, 121 Brandon St., Kensington WA 6151, Australia
 Oldenhof, Rob W., Beukenlaan 16, 3911 XK 3911 Rhenen, The Netherlands
 Ryden, Nils, Lund Inst. of Technology, Engineering Geology, Box 118, SE-22100 Lund, Sweden
 Saberi, Kouros, Cognitive Sciences, Univ. of California, Irvine, CA 92697-5100
 Sandrik, Timothy E., Harman/Becker, Car Lab 3, 1201 South Ohio St., Martinsville, IN 46151

Song, Limin, ExxonMobil Research, 1545 Route 22 East, Annandale, NJ 08801
 Succi, George, 10 62nd St., Newburyport, MA 01950
 Sun, Dequn, Rue de la Maladiere 20, CH2000 Neuchatel, Switzerland
 Talcott, John B., 180 W. 200 N., Springville, UT 84663
 Weisert, Lee N., 361 E. Main St., Manasquan, NJ 08736

New Corresponding Electronic Associates

Alkhairy, Ashraf, P.O. Box 93746, Riyadh 11683, Saudi Arabia
 Barker, Mike, MBA, P.O. Box 16068, Dowerglen, Republic South Africa
 Dintrans, Alejandro, Las Dalias 2873, Providencia, Santiago, Chile
 Escudeiro, Roberto L., Rua Maria Carolina, 354 Apt. 902, Boa Viagem, Recife, Pe, 51020220, Brazil
 Oliveira, Aluizio, Rua Joaquim Tavora 181 501, Icarai, Niteroi RJ, Brazil
 Tai, Joseph L., Perunding HTA, 33 ACD, Menara C, Pandan Kapital, P. Indah, 55100 Kuala Lumpur, 55100, Malaysia
 Wan, Yi, Lanzhou Univ., Information School, 222 S. Tianshui Rd., 730000 Lanzhou, P.R. China

Associates Elected Members

N. S. Antomanzas-Barroso, J. J. Bielecki, C. Fullgrabe, T. B. Johnson, D. N. LaForgia, L. Morales, S. A. Moyes, S. Nissen, O. A. Sapozhnikov, B. Wen, R. Willis

Student to Electronic Associate

K. A. Pestka

Resigned

R. Piazza—Fellow
 G. Clark, G. L. Ossipov, M. Rossi—Members
 E. Fischer-Jorgensen, S. Takano—Associates

Deceased

S. A. Elder, J. M. Pickett—Fellows
 J. K. Cullen, Jr., R. E. Francois, J. Frarey—Members

Fellows	900
Members	2176
Associates	2672
Electronic Associates	696
Students	1093
Total	7537

ACOUSTICAL NEWS—INTERNATIONAL

Walter G. Mayer

Physics Department, Georgetown University, Washington, DC 20057

International Meetings Calendar

Below are announcements of meetings and conferences to be held abroad. Entries preceded by an * are new or updated listings.

December 2007

6–9 **International Symposium on Sonochemistry and Sonoprocessing (ISSS2007)**, Kyoto, Japan (Web: www.j-sonochem.org/ISSS2007).

March 2008

3–10 * **34th Meeting of the German Association for Acoustics (DAGA2008)**, Dresden, Germany (Web: 2008.daga-tagung.de).

17–19 * **Spring Meeting of the Acoustical Society of Japan**, Narashino, Japan (Web: www.asj.gr.jp/index-en.html).

30–1 * **SAE Brasil Noise and Vibration Conference-NVH**, Florianopolis, SC, Brazil (Web: www.saebrasil.org.br/eventos/secao_parana_sc/nvh2008/site).

April 2008

8–11 **Oceans'08**, Kobe, Japan (Web: www.oceans08mstsieekobe-technoocean08.org/index.cfm).

10–11 * **Institute of Acoustics (UK) Spring Conference**, Reading, UK (Web: www.ioa.org.uk/viewupcoming.asp).

17–18 **Spring Meeting of the Swiss Acoustical Society**, Bellinzona (Tessin), Switzerland (Web: www.sga-ssa.ch).

June 2008

29–4 * **Acoustics'08 Paris: 155th ASA Meeting + 5th Forum Acusticum (EAA) + 9th Congr s Franais d' Acoustique (SFA)**, Paris, France (Web: www.acoustics08-paris.org).

July 2008

6–10 * **15th International Congress on Sound and Vibration**, Daejeon, Korea (Web: www.icsv15.org).

7–10 **18th International Symposium on Nonlinear Acoustics (ISNA18)**, Stockholm, Sweden (Web: www.congrex.com/18th_isna).

27–31 **10th Mechanics of Hearing Workshop**, Keele University, UK (Web: www.mechanicsofhearing.com).

August 2008

25–28 * **1st International Conference on Water Side Security**, Lyngby, Denmark (Web: www.wss2008.org).

25–29 **10th International Conference on Music Perception and Cognition (ICMPC 10)**, Sapporo, Japan (Web: icmpc10.typepad.jp).

September 2008

8–12 **International Symposium on Underwater Reverberation and Clutter**, Lerici, Italy (Web: isurc2008.org).

10–12 * **Autumn Meeting of the Acoustical Society of Japan**, Fukuoka, Japan (Web: www.asj.gr.jp/index-en.html).

15–17 **International Conference on Noise and Vibration Engineering (ISMA2008)**, Leuven, Belgium (Web: www.isma-isaac.be).

23–25 * **Underwater Noise Measurement**, Southampton, UK (Web: www.ioa.org.uk/viewupcoming.asp).

22–26 **INTERSPEECH 2008 — 10th ICSLP**, Brisbane, Australia (Web: www.interspeech2008.org).

October 2008

21–23 * **International Conference on Low Frequency Noise and Vibration**, Tokyo, Japan (Web: www.lowfrequency2008.org).

21–24 **ac stica 2008**, Coimbra, Portugal (Web: www.spacustica.pt).

26–29 **Inter-noise 2008**, Shanghai, China (Web: www.internoise2008.org).

November 2008

2–5 **IEEE International Ultrasonics Symposium**, Beijing, China (Web: www.ieee-uffc.org/ulmain.asp?page=symposia).

24–26 * **Australian Acoustical Society National Conference**, Geelong, Vic, Australia (Web: www.acoustics.asn.au).

April 2009

19–24 * **International Conference on Acoustics, Speech, and Signal Processing**, Taipei, R.O.C. (Web: icassp09.com).

August 2009

23–28 * **Inter-noise 2009**, Ottawa, Ontario, Canada (Contact: TBA).

September 2009

6–10 **InterSpeech 2009**, Brighton, UK (Web: www.interspeech2009.org).

August 2010

23–27

20th International Congress on Acoustics (ICA2010), Sydney Australia (Web: www.ica2010sydney.org).

September 2010

26–30

Interspeech 2010, Makuhari, Japan (Web: www.interspeech2010.org).

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice

11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, Perception, Incorporated, Box 39536, Los Angeles, California 90039

ANGELO CAMPANELLA, 3201 Ridgewood Drive, Hilliard Ohio 43026-2453

JEROME A. HELFFRICH, Southwest Research Institute, San Antonio, Texas 78228

DAVID PREVES, Starkey Laboratories, 6600 Washington Ave. S. Eden Prarie, Minnesota 55344

NEIL A. SHAW, Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290

KEVIN P. SHEPHERD, Mail Stop 463, NASA Langley Research Center, Hampton, Virginia 23681

ERIC E. UNGAR, Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138

ROBERT C. WAAG, Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627

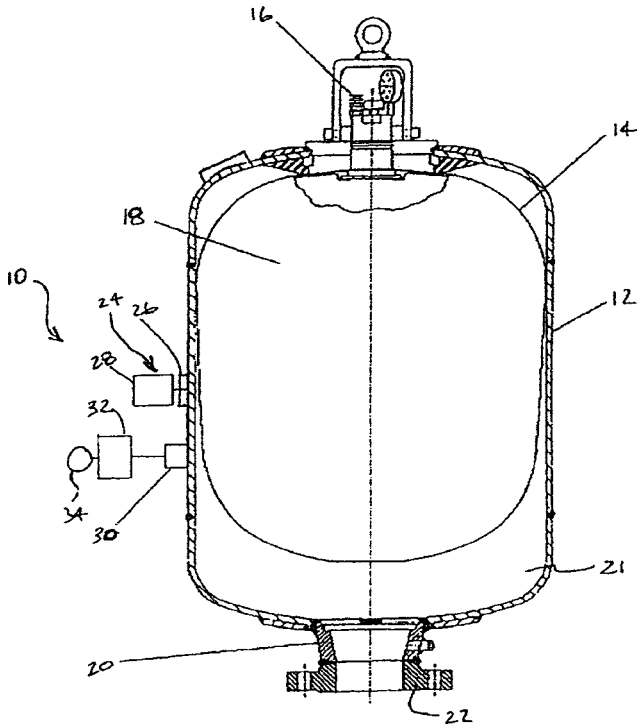
7,216,536

43.28.Py ACOUSTIC VOLUME INDICATOR

Winston B. Young and Huey Wai, assignors to Young Manufacturing & Engineering, Incorporated

15 May 2007 (Class 73/290 V); filed 7 October 2005

A fill-volume sensor is claimed for gas-charged hydraulic accumulators. Sound from tapper 24, 26 resonates in bladder 14 filled with gas 18, the remaining part of tank 12 volume 21 being hydraulic fluid. The subsequent



resonant sound is detected by sensor 30 whose output signal is fed to processor 32. The processor executes a fast Fourier transform on this signal to identify the frequency of the dominant sound resonance, to be compared to

previously determined values according to the volume of the gas filled bladder. The volume quantity result is presented by display 34.—AJC

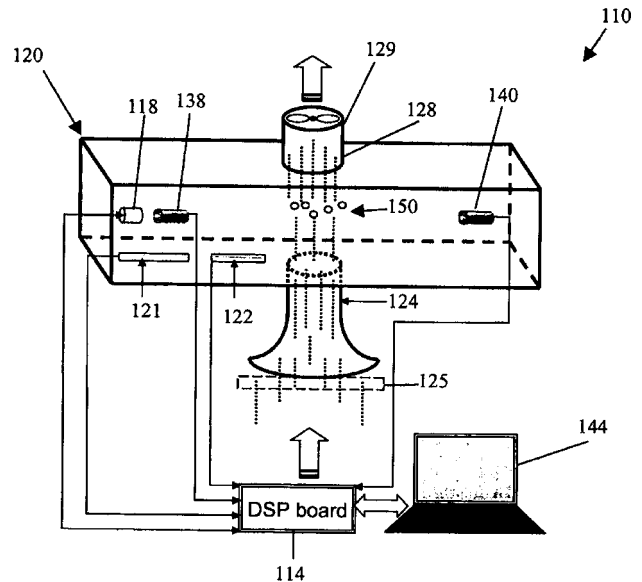
7,213,445

43.28.Vd ACOUSTIC PARTICULATES DENSITY SENSOR

Sean F. Wu and Ravinder S. Beniwal, assignors to Wayne State University

8 May 2007 (Class 73/24.03); filed 23 September 2004

Apparatus 110 and method 144 determine the airborne particle density by measuring the time of flight of a sound pulse across chamber 120, as compared to the theoretical time of flight based on clean air temperature,



pressure, and humidity calculated by mathematical formulas. No evidence is offered that the device has been built or evaluated for accuracy.—AJC

7,234,519

43.28.We FLEXIBLE PIEZOELECTRIC FOR DOWNHOLE SENSING, ACTUATION AND HEALTH MONITORING

Michael L. Fripp *et al.*, assignors to Halliburton Energy Services, Incorporated
26 June 2007 (Class 166/250.01); filed 8 April 2003

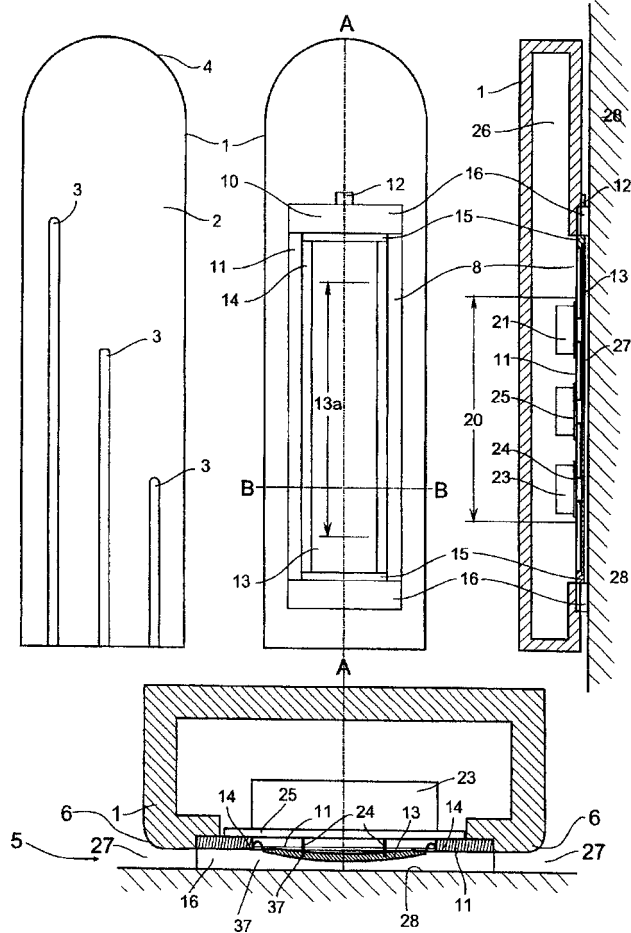
Sound transmission along drill strings has long been used in the oil drilling industry as a means of transmitting data from down-hole sensors to drill operators on the surface. This patent describes a method for making and bonding lead-zirconate-titanate-based transducers to the surface of a drill string and electroding the resulting structure for good electromechanical coupling. The patent clearly describes the problem being addressed and the merits of the solution in a few pages. There is no description of what the design compromises are and why the dimensions are chosen to be what they are, so do not look to this patent for guidance on building your own.—JAH

7,218,749

43.38.Dv METHOD FOR SOUND REPRODUCTION AND PILLAR LOUDSPEAKER

Sisko Noponen, legal representative and Tapani Hintsala, assignors to Anturilaakso Oy
15 May 2007 (Class 381/431); filed in Finland 17 September 1998

A columnar transducer assembly 10 featuring a single rectangular diaphragm 13 with multiple “operating devices” 21 is described, as are various



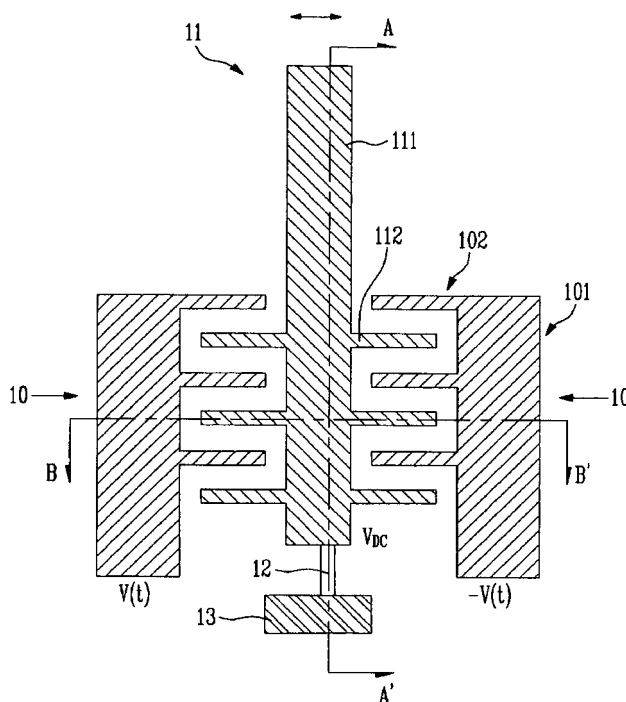
and sundry implementations.—NAS

7,242,129

43.38.Fx PIEZOELECTRIC AND ELECTROSTATIC MICROELECTROMECHANICAL SYSTEM ACTUATOR

Ki Chul Kim *et al.*, assignors to Electronics and Telecommunications Research Institute
10 July 2007 (Class 310/309); filed in Republic of Korea 16 December 2004

The authors describe a combination comb-drive and cantilever assembly that is able to perform two-dimensional positioning motions. The stated objective of this approach is to provide for tracking motions in an optical disk drive, but at a stated range capability of 250 μm, the range of motions



available would appear to be less than needed for such a task. It is an interesting combination of actuator designs that probably came about by accident, as anyone who has ever made a comb drive will attest to the fact that unintended out-of-plane motion is very hard to control.—JAH

6,727,808

43.38.Hz SONIC BLIND SPOT MONITORING SYSTEM

David J. Uselmann and Linda M. Uselmann, both of Madison, Wisconsin
27 April 2004 (Class 340/436); filed 14 December 2001

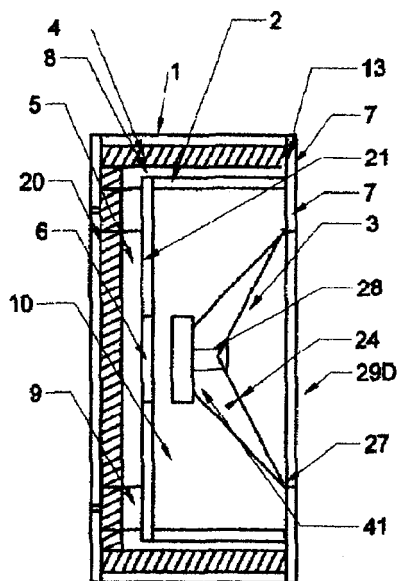
A means to warn a driver of the presence of a vehicle in his “blind spot” consists of sonar emitters and receptors arranged on both sides of the vehicle. Lights are integrated into the rearview mirror to indicate the presence of a vehicle.—KPS

7,207,413

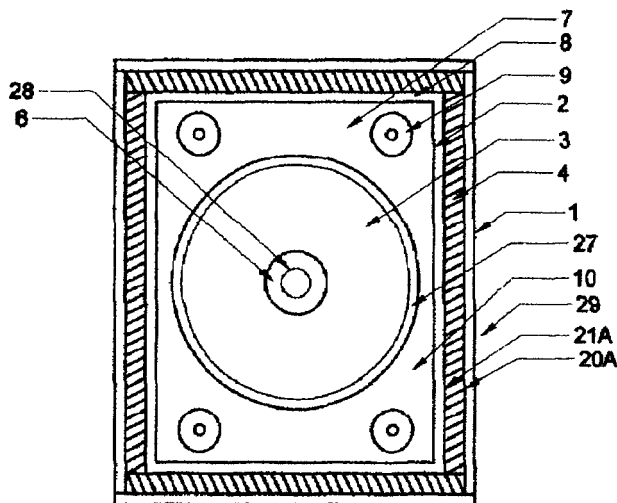
43.38.Ja CLOSED LOOP EMBEDDED AUDIO TRANSMISSION LINE TECHNOLOGY FOR LOUDSPEAKER ENCLOSURES AND SYSTEMS

Jan Princeton Plummer, assignor to TBI Audio Systems LLC
24 April 2007 (Class 181/199); filed 12 May 2004

“The enclosure favors a narrow range of frequencies and for others reacts violently, creating a plethora of incoherent internal standing waves that modulate the diaphragm with nonsymmetrical vibration patterns.” By adding an internal bafflelike device to the enclosure, described as an indirect



A



B

direct-coupled embedded acoustic transmission line (IDC EATL) 5, these violent reactions are vanquished, notwithstanding the size of the IDC EATL with respect to the wavelengths involved and the closed loop topology of the device.—NAS

6,726,557

43.38.Ja VEHICLE HAVING A SOUND-RADIATING ELEMENT

Jürgen Dreyer *et al.*, assignors to Siemens Aktiengesellschaft
27 April 2004 (Class 454/152); filed in Germany 11 September 2001

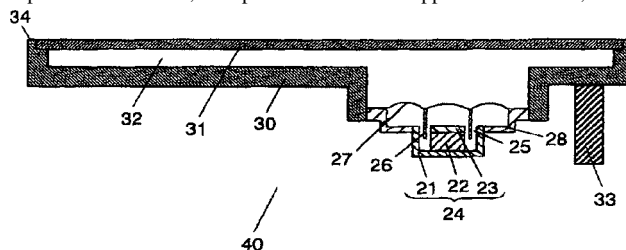
Placement of loudspeakers in automobile interiors is limited by available space. It is proposed that the inner surface of the air conditioning duct be mechanically driven in order to introduce sound into the passenger compartment. This is a space- and weight-efficient solution, but no details are provided on acoustical performance.—KPS

7,218,743

43.38.Ja SPEAKER MODULE AND APPARATUS USING THIS

Masahide Sumiyama *et al.*, assignors to Matsushita Electric Industrial Company, Limited
15 May 2007 (Class 381/186); filed in Japan 3 October 2003

LED assembly 33 is affixed to speaker 40. Except for 33, the device is identical to prior art. “Light from the light emitting section passes through the panel . As a result, the speaker module can appeal its existence, so that



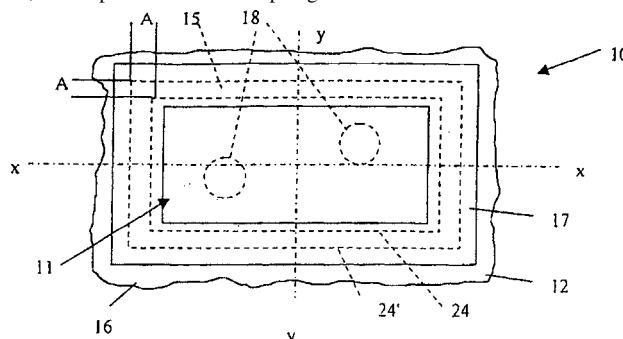
outward design, which makes a high impact, can be realized.” The device appears to have utility in boom boxes and automobiles.—NAS

7,236,601

43.38.Ja PANEL LOUDSPEAKER

Wolfgang Bachmann, Grevenbroich, Germany, *et al.*
26 June 2007 (Class 381/152); filed in Germany 15 May 1998

The diaphragm of a panel-type loudspeaker is suspended in an outer baffle 12 by flexible strips 17, which are “under mechanical tension.” Moreover, some portions of the diaphragm itself “are also under mechanical



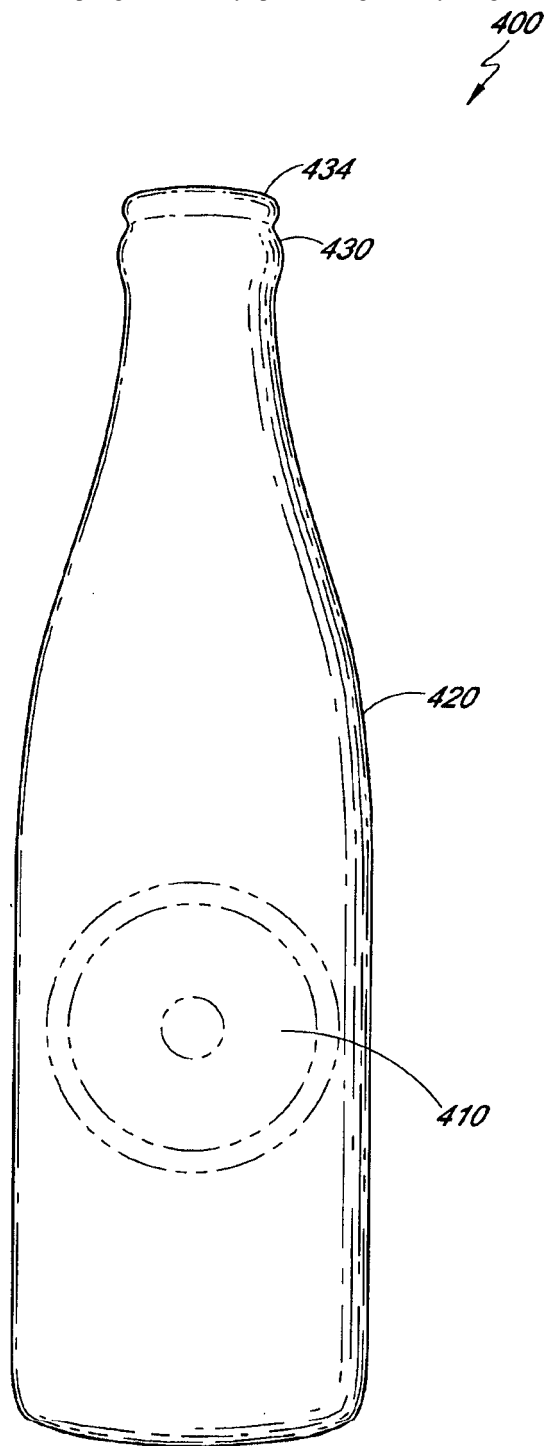
tension.” Since variants of both features can be found in prior art, the novelty of this design seems to hinge on the definition of “mechanical tension.”—GLA

7,218,747

43.38.Ja EXTERNALLY PORTED LOUDSPEAKER ENCLOSURE

Nick Huffman, Simi Valley, California
15 May 2007 (Class 381/345); filed 5 December 2003

Helmholtz, Chebychev, Thiele, Small, Butterworth, and Jensen are referenced in this patent, in which the port for a vented enclosure is described as external. Topologists and many speaker designers may disagree with the



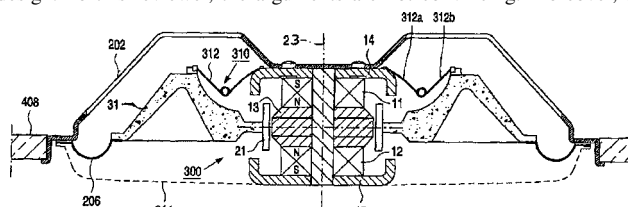
novelty of the device, but it does include all the parts—in this embodiment, a driver 410, external to the “enclosure,” an enclosure 420, and a port section 430 and 434.—NAS

7,242,787

43.38.Ja ELECTROMAGNETIC DRIVING UNIT FOR A LOUDSPEAKER ASSEMBLY

Guido Odilon Maurits D’Hoogh, assignor to PSS Belgium, N.V.
10 July 2007 (Class 381/412); filed in the European Patent Office
25 April 2002

In this unusual push-pull motor assembly, voice coil 21 drives the cone 31 from its midpoint, allowing long excursions in a very shallow assembly. No mention is made of the reduction in distortion afforded by the symmetrical layout. Instead, the patent argues that greater efficiency and longer linear travel can be achieved in comparison to a more conventional overhung coil design. To this reviewer, the arguments are not convincing. Moreover, the



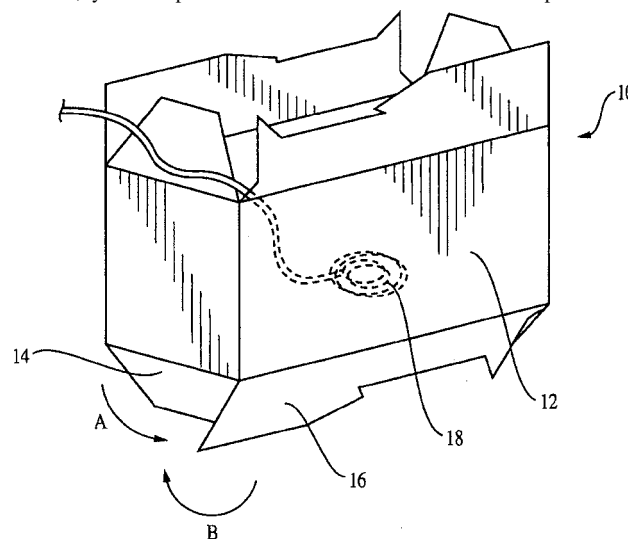
actual construction of such a voice coil assembly is left to the imagination. It might be possible to wind two coils on a conventional tubular former—one above and one below the attachment point—but such an arrangement is prohibited by the patent claims, which clearly specify a “single coil.”—GLA

7,245,729

43.38.Ja LOUDSPEAKER

Andrew D. Bank *et al.*, assignors to New Transducers Limited
17 July 2007 (Class 381/152); filed in United Kingdom 5 April 2001

Yes, you can put an inertia transducer inside a fold-up cardboard



carton to make a talking box. This is new?—GLA

6,771,166

43.38.Kb AIR BAG DEPLOYMENT SENSING APPARATUS AND METHOD

Gary E. Mastenbrook, assignor to Motorola, Incorporated
3 August 2004 (Class 340/425.5); filed 14 November 2001

Automated emergency-call response systems in vehicles respond to a crash event by notifying appropriate authorities. The systems are usually

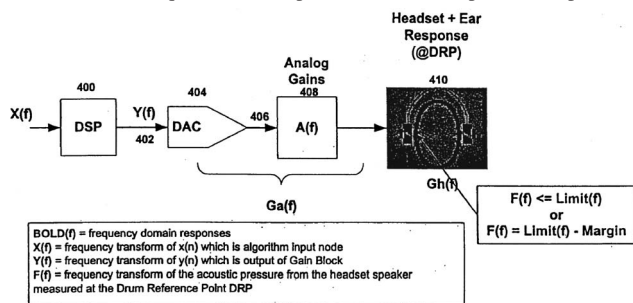
tied into the car's electrical system and are triggered by activation of the air bag. For after-market applications, an acoustical method is proposed that will identify air bag deployment and provide the needed trigger to the emergency call system. The acoustical method relies on both sound level and spectral shape to recognize the acoustical event associated with air bag deployment.—KPS

7,242,783

43.38.Lc AUDIO LIMITING CIRCUIT

William A. Weeks and William R. Morrell, assignors to Plantronics, Incorporated
10 July 2007 (Class 381/98); filed 25 January 2006

This is a continuation of United States Patent 7,013,011 [reviewed in J. Acoust. Soc. Am. 120(4), 1762 (2006)]. Both patents deal with the practical digital implementation of sophisticated audio signal limiting in which sound levels, bands of frequencies, time periods, and the response of a particular



transducer all must be taken into account. The patent describes a practical application in which protective limiting circuitry is included with a plug-in headset that might be used with a telephone.—GLA

6,731,204

43.38.Md OBJECT DETECTION SYSTEM PROVIDING DRIVER INFORMATION THROUGH SOUND

Kurt S. Lehmann, assignor to Continental Teves, Incorporated
4 May 2004 (Class 340/435); filed 19 February 2002

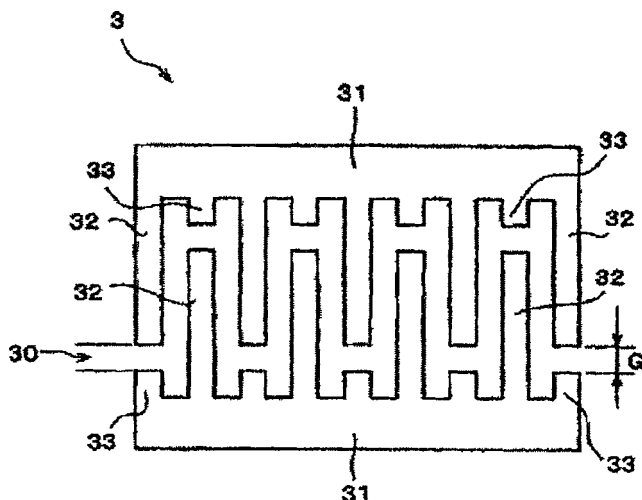
Several automobile manufacturers provide systems to warn a driver of the presence of objects that are behind their vehicle when reversing. This patent aims to extend such a scheme to include objects in all directions. The warning sounds will be presented via the automobile's multiloudspeaker sound system to convey the direction and distance of said objects. Other than "an object detection system" and a "processor" to determine if objects meet "certain criteria," the 20 claims provide no indication of how such a system might actually be constructed.—KPS

7,215,224

43.38.Rh SURFACE ACOUSTIC WAVE FILTER, SURFACE ACOUSTIC WAVE DEVICE AND COMMUNICATION DEVICE

Kenichi Noto, assignor to Murata Manufacturing Company, Limited
8 May 2007 (Class 333/202); filed in Japan 9 April 2001

The author argues that reductions in spurious frequency resonances and in acoustic wave radiation from gap surface areas result from adding



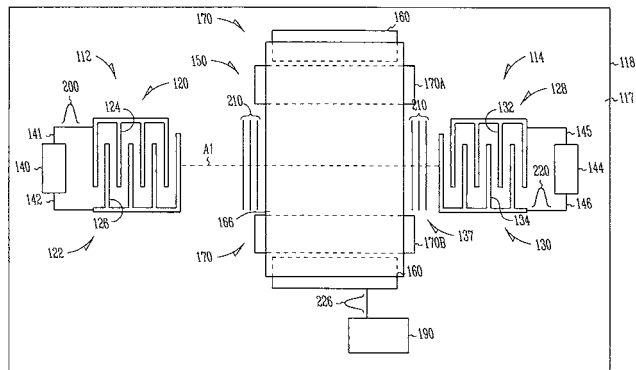
short buss extensions 33 to reduce the width G of gap 30.—AJC

7,218,188

43.38.Rh MICROELECTROMECHANICAL APPARATUS AND METHODS FOR SURFACE ACOUSTIC WAVE SWITCHING

Qing Ma and Dong S. Shim, assignors to Intel Corporation
15 May 2007 (Class 333/193); filed 18 January 2005

A microelectromechanical system (MEMS) acoustical switch is claimed where a controller 190 sends voltage 226 into a deformable cover apparatus 166 over substrate 117 that is carrying a surface acoustic wave (SAW) 210 along path 137 from SAW source 120 to SAW receiver 128. Cover 166 is covered on its side facing substrate 117 with, for instance, a



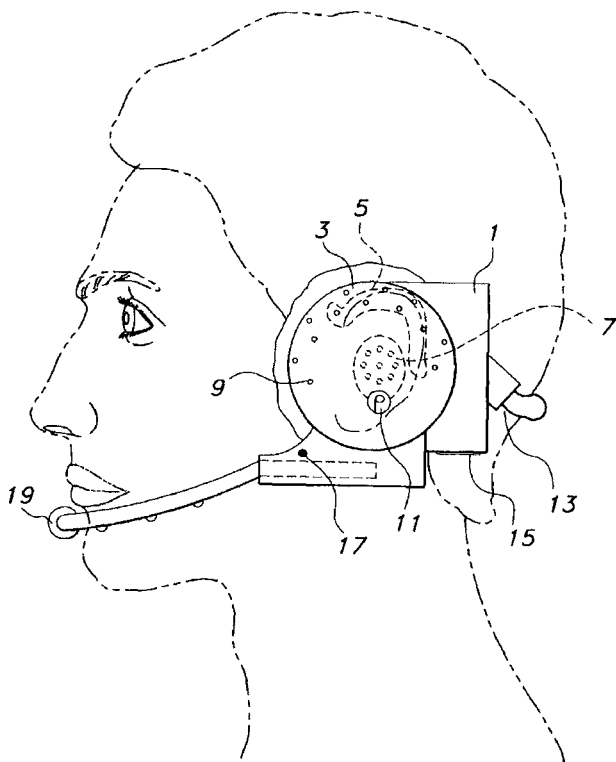
grating or a sound absorber. The cover is driven toward the substrate by voltage 226 on electrodes 160, causing an electromagnetic force to deflect the cover. When the cover contacts substrate surface 117, the transmitted SAW is attenuated.—AJC

7,242,765

43.38.Si HEADSET CELLULAR TELEPHONES

Tommy Lee Hairston, Ringgold, Georgia
10 July 2007 (Class 379/419); filed 28 June 2002

The invention is an all-in-one cellular telephone that fits over the user's ear and is controlled through speech recognition circuitry. According to the first (and only independent) patent claim, the earpiece includes "an internal speaker for delivering sound to the user's ear, and an external



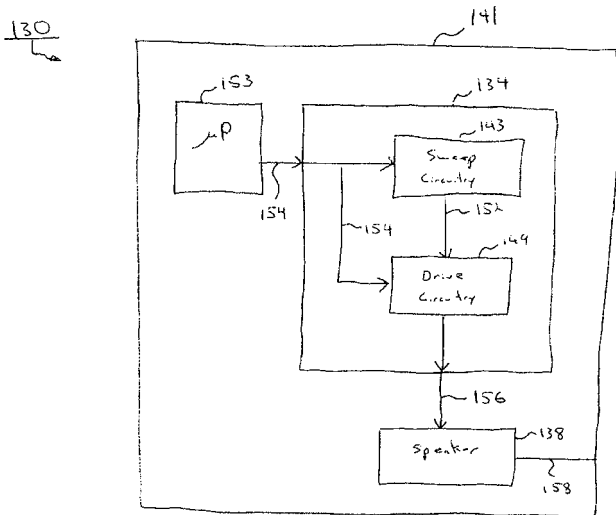
speaker electrically coupled to the internal speaker for allowing the user to hear ambient noise through the internal speaker, and adapted to allow a ringtone to be heard externally of the device.” Whether this represents a significant departure from prior art is an exercise best left to patent attorneys.—GLA

7,245,891

43.38.Si WIRELESS MOBILE COMMUNICATION DEVICE HAVING A SPEAKER VIBRATION ALERT AND METHOD OF USING SAME

Derick Hugunin, assignor to Kyocera Wireless Corporation
17 July 2007 (Class 455/148); filed 21 October 2003

To reduce the size and cost of a cellular telephone, a single transducer with a single actuating coil can be used for both audio reproduction and



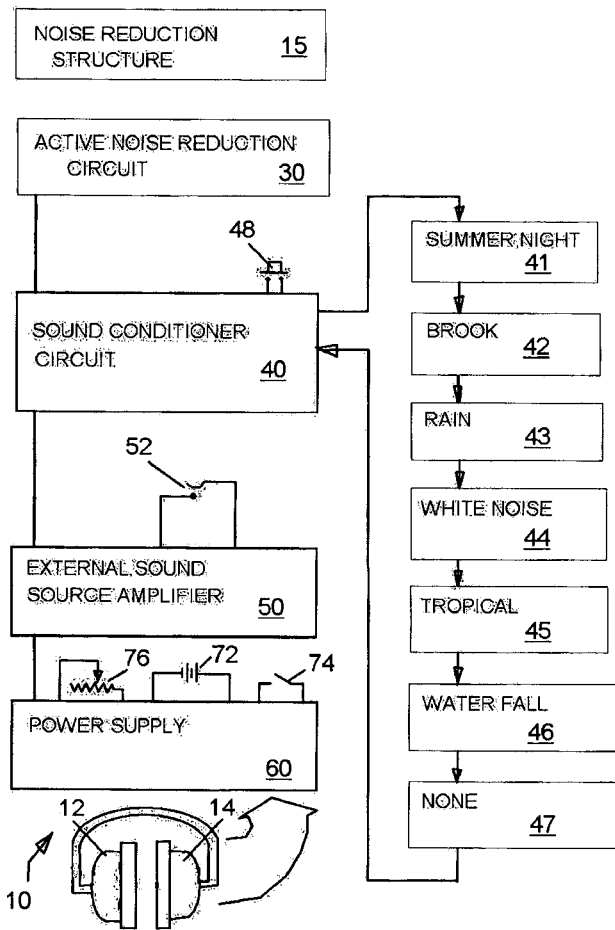
tactile signaling. Such a loudspeaker has a second mechanical resonant frequency well below its normal cone resonant frequency. Thus, when driven by a speech signal it functions like an ordinary loudspeaker, but when driven at its mechanical resonance it generates inaudible vibrations. For this scheme to work, however, the driving frequency and the resonant frequency must be almost identical, whereas normal manufacturing tolerances result in a fairly wide range of vibration resonant frequencies. Several approaches to the problem have already been patented. The solution suggested here is to drive the transducer with a swept sine wave signal.—GLA

7,248,705

43.38.Si NOISE REDUCING HEADPHONES WITH SOUND CONDITIONING

Steven Mishan, assignor to Van Hauser LLC
24 July 2007 (Class 381/71.6); filed 29 December 2005

This headphone assembly includes built-in electronics and (hopefully lightweight) batteries. As indicated in the diagram, modes of operation



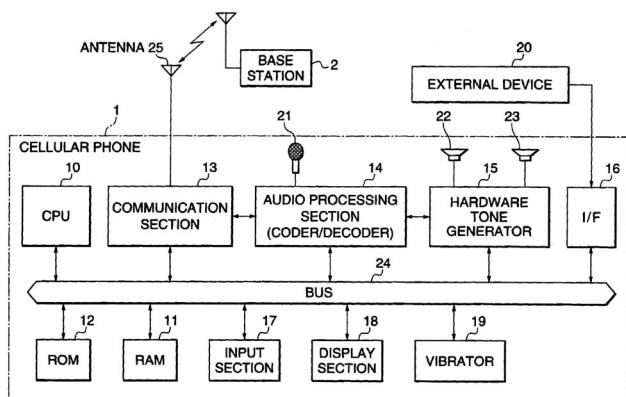
include active noise reduction and the reproduction of stored background sounds as well as external signals.—GLA

7,247,784

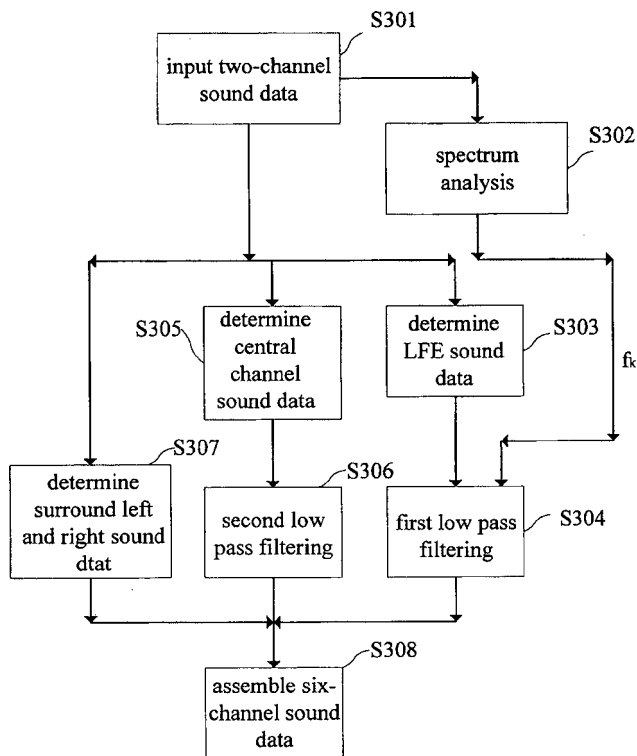
43.38.Si MUSICAL SOUND GENERATOR, PORTABLE TERMINAL, MUSICAL SOUND GENERATING METHOD, AND STORAGE MEDIUM

Yasuyuki Muraki and Ichiro Futohashi, assignors to Yamaha Corporation
 24 July 2007 (Class 84/609); filed in Japan 18 August 2000

This is another long, complicated patent dealing with the generation of custom ringtones for cellular telephones. A computational method is described that is said to minimize specialized hardware requirements without



imposing a heavy load on the embedded CPU. One cannot help but wonder what the world would be like if just half of the engineering time devoted to custom ringtones was diverted to, say, stem cell research.—GLA



below 3 kHz are summed and fed to a center channel. Left and right signals are also band-limited somehow and fed to left and right surround channels.—GLA

7,243,073

43.38.Vk METHOD FOR REALIZING VIRTUAL MULTI-CHANNEL OUTPUT BY SPECTRUM ANALYSIS

Ta-Jung Yeh *et al.*, assignors to VIA Technologies, Incorporated
 10 July 2007 (Class 704/500); filed in Taiwan 23 August 2002

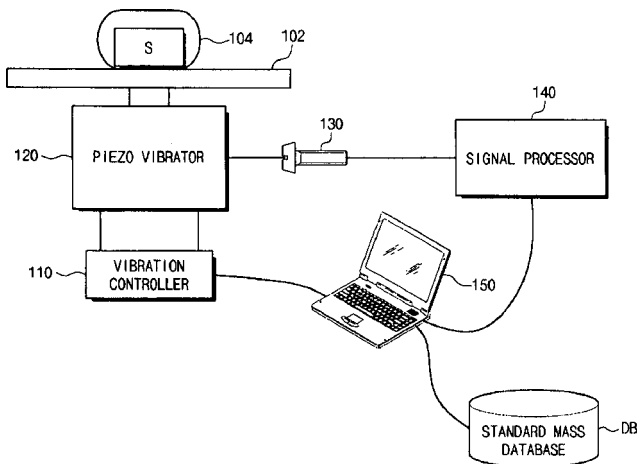
This patent describes a method for deriving 5.1 audio playback from a two-channel stereo signal. Unfortunately, the text is no better than the flow diagram in explaining exactly what has been patented. It appears that left and right signals are fed directly to front-left and front-right channels. Low frequencies are summed and fed to a common bass channel. Frequencies

7,216,543

43.40.Le MASS MEASUREMENT SYSTEM AND METHOD USING MEASUREMENT OF FREQUENCY SHIFT OF VIBRATOR

Hong Yul Paik *et al.*, assignors to Korea Aerospace Research Institute
 15 May 2007 (Class 73/579); filed in Republic of Korea 15 July 2004

A zero-gravity mass sensor is claimed where piezo-vibrator 120 causes the mechanical system 120-102-104-S to vibrate at a frequency that depends on the overall system mass, including the unknown mass S of the sample.



The relationship of this frequency to the sample mass *S* is previously determined via reference masses and is stored in database **DB**. A frequency signal is generated by sensor **130** and fed into processor **140** to be calculated into a displayed mass *S* value by **150**, for instance.—AJC

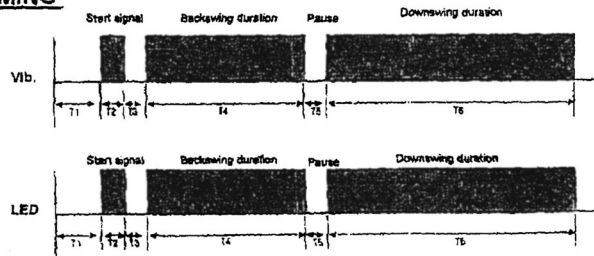
7,217,197

43.40.Le GOLF SWING TRAINING DEVICE

Glenn Park, Rancho Palos Verdes, California
15 May 2007 (Class 473/221); filed 30 October 2003

This golfer's aide allows for programming the device to provide audio, visual, or tactile signals to the golfer to help improve the timing of the

TIMING



various parts of a golf swing (see figure). It is said to be usable on the links as well as in more controlled circumstances.—NAS

7,248,958

43.40.Le ROAD SURFACE STATE ESTIMATING APPARATUS, ROAD SURFACE FRICTION STATE ESTIMATING APPARATUS, ROAD SURFACE STATE PHYSICAL QUANTITY CALCULATING APPARATUS, AND ROAD SURFACE STATE ANNOUNCING APPARATUS

Yoshitoshi Watanabe *et al.*, assignors to Kabushiki Kaisha Toyota Chuo Kenkyusho
24 July 2007 (Class 701/70); filed in Japan 8 November 2002

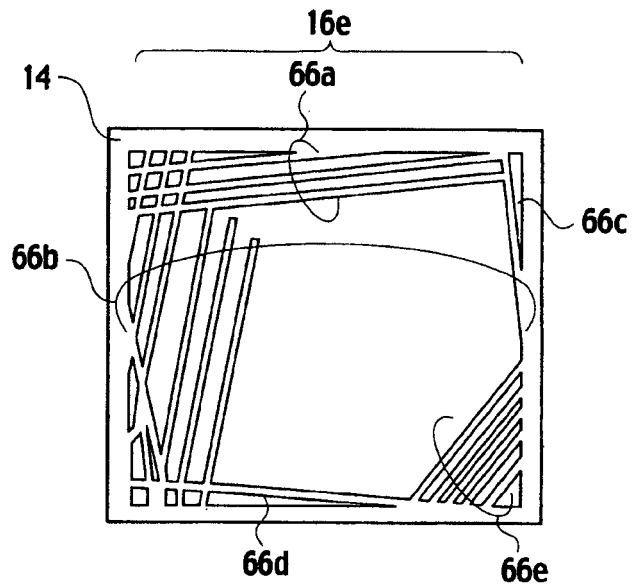
A feature vector is generated by a processor on the basis of a signal that characterizes the sound produced by a tire running over a road surface and a signal from a wheel rotation speed detector. This vector is compared to a judgment map that is stored in a memory so as to produce an estimate of the state of the road surface. Information obtained from luminance sensors, hue sensors, etc., may be used to assist in development of the estimate.—EEU

7,236,066

43.40.Sk FILM BULK ACOUSTIC RESONATOR AND FILTER CIRCUIT INCLUDING A PLURALITY OF FILM BULK ACOUSTIC RESONATORS

Yasuo Ebuchi, assignor to Kabushiki Kaisha Toshiba
26 June 2007 (Class 333/187); filed in Japan 13 May 2005

The authors describe a method of fabricating film bulk acoustic wave resonators (FBARs) with an additional layer of material patterned so as to break up lateral (transverse) modes that are interfering with the desired longitudinal-mode behavior of the FBAR device. The patent is clearly written and describes the approach, which includes some interesting electrode



variants like the one shown in the figure. This sort of modification should not be new to most persons well versed in the art, but nevertheless it gives food for thought. It seems to have been designed by simple heuristics and trial and error.—JAH

7,234,379

43.40.Tm DEVICE AND A METHOD FOR PREVENTING OR REDUCING VIBRATIONS IN A CUTTING TOOL

Ingvar Claesson, Dalby, Sweden *et al.*
26 June 2007 (Class 82/1.11); filed 28 June 2005

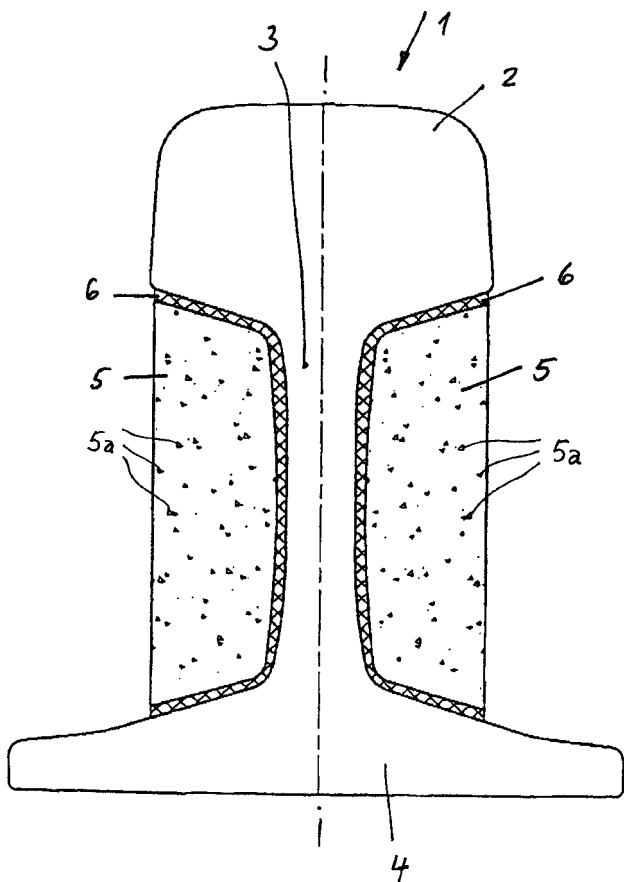
The holder of a cutting tool, such as that of a lathe, is provided with a dynamic absorber and an active control system. The dynamic absorber is located in a cavity in the tool holder near the cutting tool and consists of a mass that is supported on rubber rings or the like. These resilient elements may have different stiffnesses in the axial and in the radial directions, thus providing different absorber resonant frequencies in the different directions. In the active system, a signal derived from an accelerometer near the cutting tool is fed to a controller, which causes actuators near the base of the tool holder to provide forces to oppose its motion. During transients that occur as the tool first makes contact with the work piece, the controller may cause the piezoelectric actuators to be loaded by a resistor, so that these actuators act like passive dampers.—EEU

7,234,647

43.40.Tm DAMPING DEVICE FOR RAILS

Udo Wirthwein *et al.*, assignors to Wirthwein AG
26 June 2007 (Class 238/382); filed in Germany 7 April 2002

Damping of the noise-radiating vibrations of rails is accomplished by a constrained-layer arrangement added to a rail **1** in the area of its web **3**. Each form-fitting constraining element **5** is made of an injection-molded or extruded thermoplastic, such as polyethylene or polypropylene, filled with



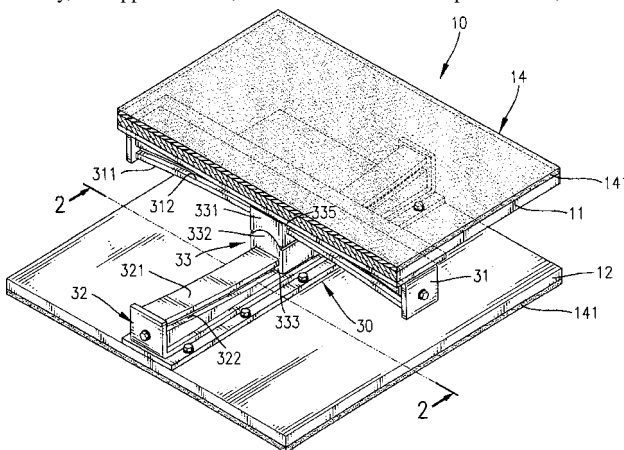
metal particles **5a** to add mass. The adhesive that holds the constraining element in place serves as the damping layer **6**.—EEU

7,237,364

43.40.Tm FOUNDATION SHOCK ELIMINATOR

Chong-Shien Tsai, Taichung, Taiwan
3 July 2007 (Class 52/167.6); filed 2 July 2004

An arrangement, intended to protect a payload from earthquake shocks and the like, consists of a base **12** and a payload-support platform **14**, which are interconnected by a sliding rail assembly. The sliding block **332** can move along the length of the curved rail **321**, which is fastened to the base. Similarly, the upper rail **312**, which is attached to the platform **14**, can slide



along block **331** in the direction perpendicular to rail **321**. The lower rail is curved “upward” and the upper rail is curved “downward,” so that the supported weight causes the system to return to its center. Resilient coatings **141** on the bottom and top surfaces serve to isolate vertical vibrations.—EEU

7,241,837

43.40.Tm VIBRATION DAMPER COMPOSITION

Shigeru Yaguchi et al., assignors to Kaneka Corporation
10 July 2007 (Class 525/95); filed in Japan 5 April 2000

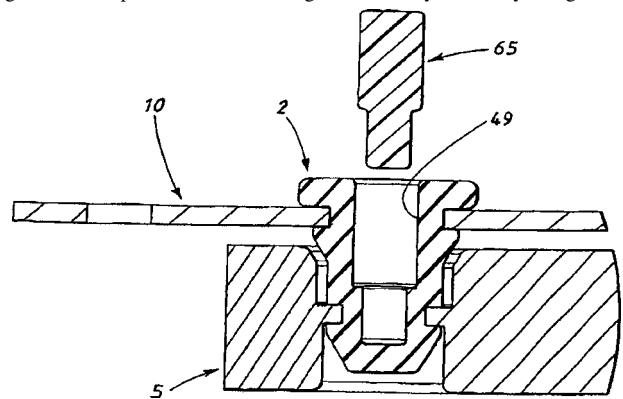
This patent describes the composition of a viscoelastic compound with a high damping capacity and near-constant rigidity at and near room temperature. The material is said to have hot-melt processability and the capacity to be formed into double-sided adhesive tape for use in a shear-damping assembly. Such assemblies are intended, for example, to protect buildings from earthquakes and typhoons.—EEU

7,246,797

43.40.Tm METHOD FOR DAMPING VIBRATIONS AND A METHOD FOR MOUNTING THE DEVICE

Bengt-Göran Gustavsson, assignor to Trelleborg Forsheda Sweden AB
24 July 2007 (Class 267/141.4); filed in Sweden 2 June 2000

A dynamic absorber to be attached to a plate element **10** consists of a rectangular mass **5** (a section through one corner of which is shown in the figure) and four elastomeric connectors **2**. These connectors and their matching holes and protrusions are configured for easy assembly. Plugs **65** of



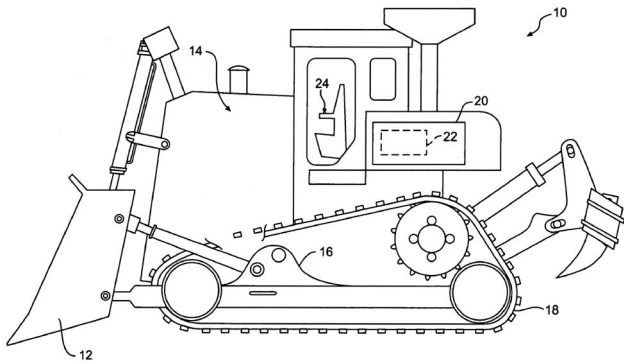
various hardnesses and shapes can be inserted in the central hole to adjust the stiffness of the connectors in the axial and radial directions and thus tune the absorber to different frequencies.—EEU

7,206,681

43.40.Vn ADAPTIVE VIBRATION MANAGEMENT SYSTEM FOR A WORK MACHINE

Kent Allen Casey et al., assignors to Caterpillar Incorporated
17 April 2007 (Class 701/50); filed 20 December 2004

An adaptive vibration management system (VMS) **20** is claimed. Machine **10** comprises an operator manipulating an input **24** (e.g., joystick), a power source **14**, and a work device (e.g., blade). Vibration sensors (not shown) are attached to machine **10** to report the vibration on all three axes (X, Y, Z) to which the operator is subjected during a work shift. The VMS includes a memory to record operator identification, vibration levels logically due to machine operation (accelerate, turn, blade position), and joystick **24** inputs. The active VMS feeds this information to a display for the



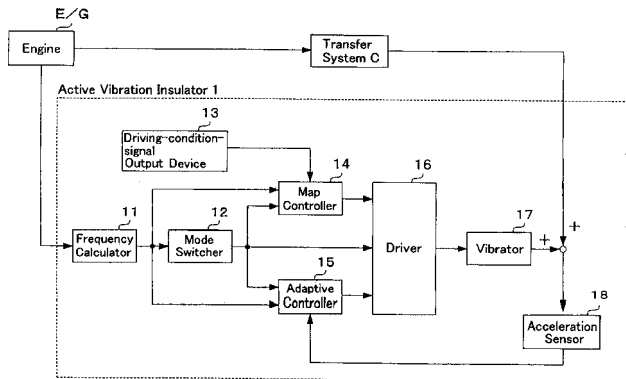
operator and to controller 22, which can modify joystick commands to electrical or hydraulic actuators when machine vibration exceeds a preset threshold level. This mitigates vibrations during operations found to produce excess operator vibration. The VMS may include an active machine suspension system to mitigate travel vibrations, to stabilize the machine platform, and to stabilize the operator seat.—AJC

7,209,338

43.40.Vn ACTIVE VIBRATION INSULATOR

Hiroyuki Ichikawa and Takehiko Fushimi, assignors to Tokai Rubber Industries, Limited
24 April 2007 (Class 361/139); filed in Japan 30 July 2004

A vehicle engine active vibration isolator 1 (“insulator”) is claimed which has inputs of engine speed 11, vehicle driver choices via a transmission drive selector, driving conditions 13, and accelerometer 18. The map controller 14 has parameters that are preset according to the vehicle and engine type, with inputs being engine firing frequency and the driving condition. An



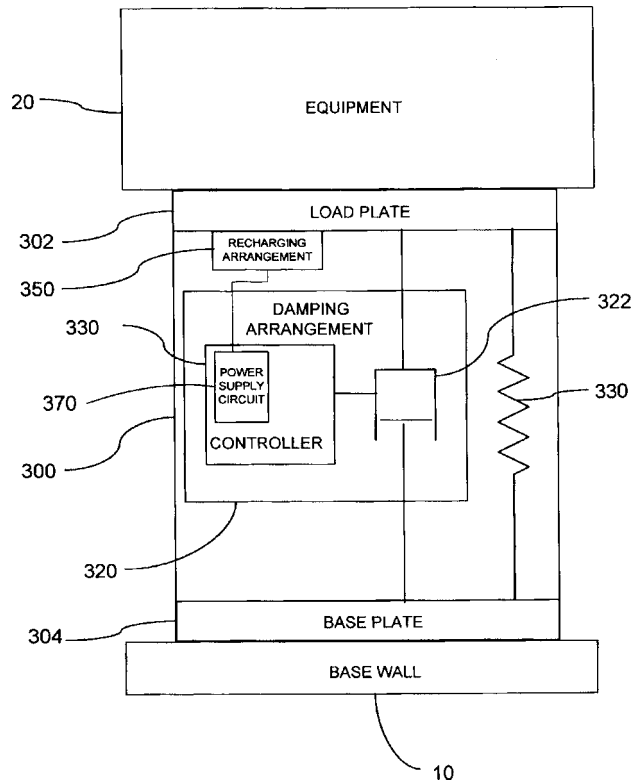
automatic mode switch 15 changes actions from the map controller when engine firing frequency rises above 50 Hz. The controller (“driver” 16) uses a retarded harmonic synthesizer least-mean-square (DXHS-LMS) filter and pulse width modulation to create the actuator (“vibrator” 17) drive current.—AJC

7,213,690

43.40.Vn SELF-POWERING SHOCK, VIBRATION AND ACOUSTIC ISOLATION SYSTEM

Edward T. Tanner, assignor to Newport News Shipbuilding
8 May 2007 (Class 188/287); filed 15 September 2003

A semiactive shock and vibration isolator 300 for heavy shipboard equipment 20 is claimed where, along with a shock and vibration isolator spring 330 and damper 322, a controller 330 is provided that will sense a large shock and will change the damping force provided by damper 322. For



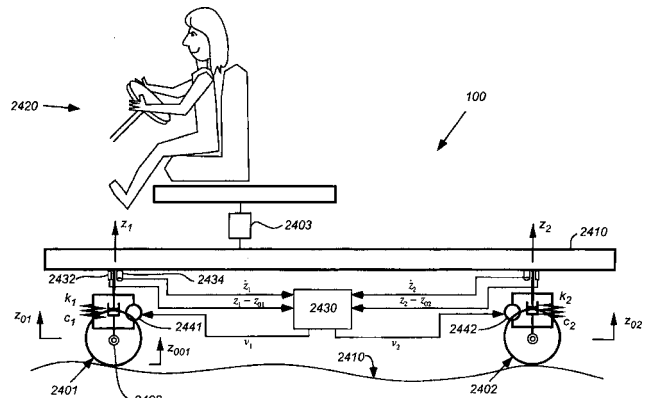
remotely mounted equipment, a controller power supply 350, 370 may be provided that derives energy from shocks and ship vibration via a moving coil or piezoelectric electrical generator charging a battery.—AJC

7,216,018

43.40.Vn ACTIVE CONTROL VIBRATION ISOLATION USING DYNAMIC MANIFOLD

Lei Zuo and Jean-Jacques E. Slotine, assignors to Massachusetts Institute of Technology
8 May 2007 (Class 700/280); filed 28 June 2004

The authors claim a robust active-vibration control plant (system) with multiple degrees of freedom that employs sliding (“moving target”) control having frequency-domain performance. For example, moving vehicle 100 comprises payload 2420 coupled to chassis 2410 by suspension 2403. The typical wheel 2401 is coupled to suspension 2403 via an appropriate coupler

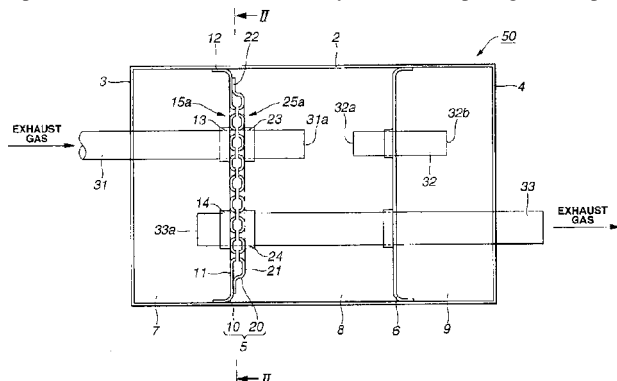


2408. Sensors for displacement 2432 and velocity 2434 are connected to processor 2430, which drives actuators 2441, 2442. Modal decomposition is claimed that handles multiple inputs and outputs to provide vibration isolation of payload 2420.—AJC

43.50.Gf MUFFLER

Yoichi Kaku *et al.*, assignors to Nissan Motor Company, Limited
16 March 2004 (Class 181/227); filed in Japan 18 February 2002

A muffler for an automobile engine has inlet 31 and exhaust 33. A notable aspect of the design is two, closely spaced, perforated plates that comprise element 5. Methods to carefully control the spacing of said plates



are given. By selection of the thickness of the plates and the size and density of the perforations, this arrangement can be reliably tuned to a specific frequency.—KPS

43.50.Gf REDUCTION IN THE NOISE PRODUCED BY A ROTOR BLADE OF A WIND TURBINE

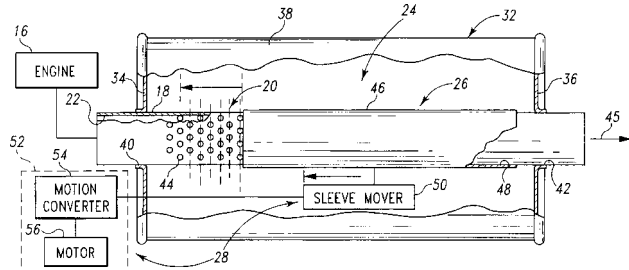
Aloys Wobben, Aurich, Germany
4 May 2004 (Class 416/241 A); filed in Germany 9 December 1998

This patent asserts that a kind of paint (“Lotusan”) that reduces ice accretion can be applied to the blades of a wind turbine to yield noise reduction. No information is provided on either the mechanism or the effectiveness of this noise reduction scheme.—KPS

43.50.Gf EXHAUST PROCESSOR WITH VARIABLE TUNING SYSTEM

Mehmet S. Ciray, assignor to Arvin Technologies, Incorporated
11 May 2004 (Class 60/312); filed 6 February 2002

A muffler that includes a tunable Helmholtz resonator is described.

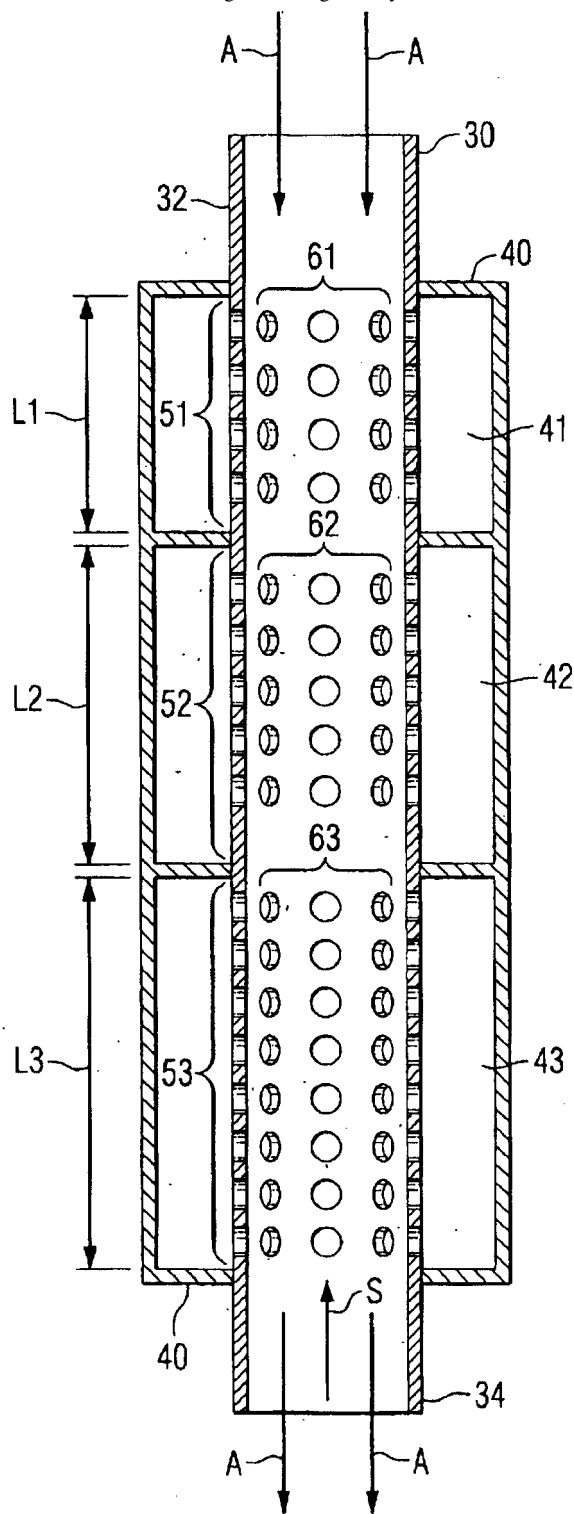


The tuning is performed by means of a variable throat. Extensive details of various alternate designs are provided.—KPS

43.50.Gf SOUND ATTENUATOR FOR A SUPERCHARGED MARINE PROPULSION DEVICE

Daniel J. Schlagenhaft, assignor to Brunswick Corporation
22 June 2004 (Class 181/249); filed 5 November 2002

This muffler is downstream of the compressor providing intake air to a marine engine. It consists of concentric tube resonators 41–43. The length of each resonator L1–L3, along with the geometry of the holes 61–63, are



chosen so that each of the three resonators is tuned to a different range of frequencies. Details of the muffler geometry and its acoustical performance are provided.—KPS

6,742,623

43.50.Gf MUFFLER FOR INTERNAL COMBUSTION ENGINE

Katsuya Tajima *et al.*, assignors to Kioritz Corporation
1 June 2004 (Class 181/269); filed in Japan 28 September 2001

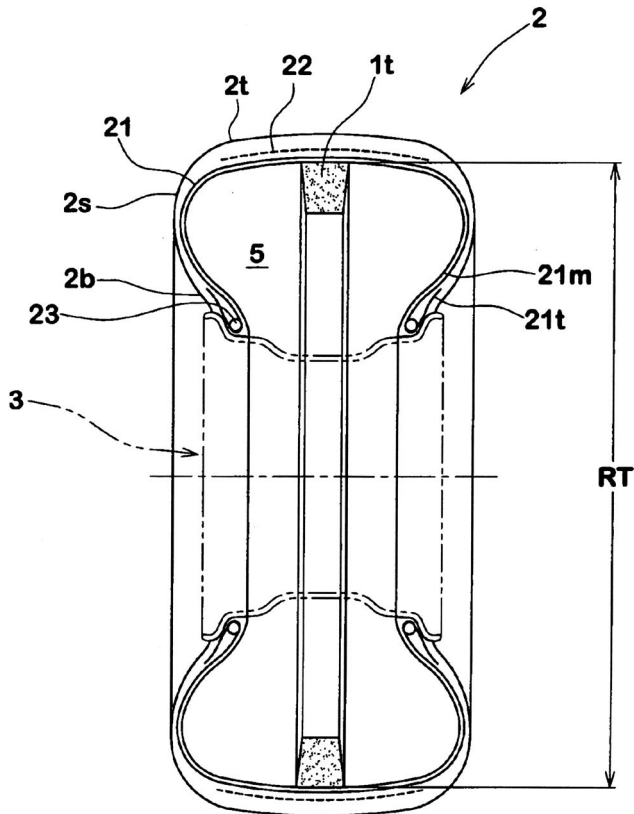
A muffler suitable for portable engines such as found in chainsaws has a lightweight, compact design. The exhaust gases enter an expansion chamber where the flow is divided into two flow paths. These two gas flows are subsequently brought together so that they impinge on one another before exiting the muffler.—KPS

7,213,624

43.50.Gf NOISE DAMPER AND PNEUMATIC TIRE WITH COMPENSATION TREAD GROOVE

Naoki Yukawa and Hidehiko Hino, assignors to Sumitomo Rubber Industries, Limited
8 May 2007 (Class 152/157); filed in Japan 10 March 2004

Passenger-car tire noise and vibration damping is claimed. Foam ring **1t** of diameter **RT** is placed on the inside surface opposite the tire tread. The preferred diameter of the preinstallation foam ring is 1.1 **RT** for the tread damper. An 8 dB reduction of the 240 Hz tire annular cavity resonance



(60 kph) at the driver's ear was measured for the **1t** tread damper ring. To prevent excess heat buildup under the **1t** foam, an annular groove is cast into the tread wear surface.—AJC

6,758,304

43.50.Gf TUNED HELMHOLTZ RESONATOR USING CAVITY FORCING

Ian R. McLean, assignor to Siemens VDO Automotive Incorporated
6 July 2004 (Class 181/206); filed 7 September 2000

A tunable Helmholtz resonator is described in which a loudspeaker is placed within the resonator cavity and driven by a signal related to the speed of the engine.—KPS

6,769,511

43.50.Gf VARIABLE TUNED EXHAUST SYSTEM

Norman Brooks and Doru N. Serban, assignors to General Motors of Canada Limited
3 August 2004 (Class 181/241); filed 18 February 2003

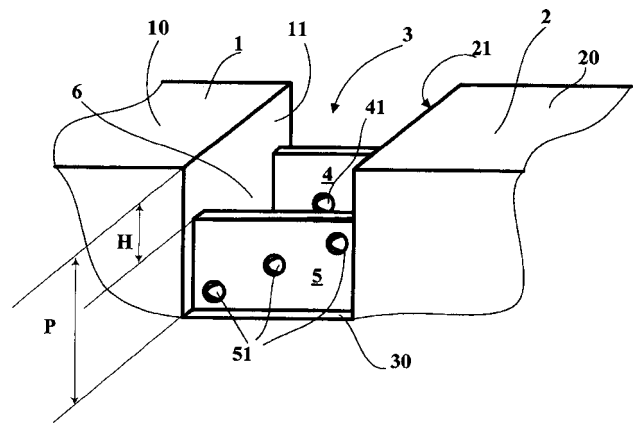
A muffler for an automobile is described that contains an adjustable element that slides and is actuated by the pressure of the exhaust gas. The sliding results in changes to the length of a resonant chamber and to the flow path, thus providing variable tuning.—KPS

7,213,625

43.50.Gf TREAD WHICH REDUCES RUNNING NOISE

Didier Martin *et al.*, assignors to Michelin Recherche et Technique S.A.
8 May 2007 (Class 152/209.17); filed in France 29 January 2001

Tire tread surface grooves **P**, lateral to the direction of travel, are required for heavy vehicles to ensure gripping of unsure surfaces. The remaining treads **2**, **10** require stabilizing ribs **4**, **5**, also recessed by **H**, to prevent noise generation when traveling on smooth roads. Cavity closure



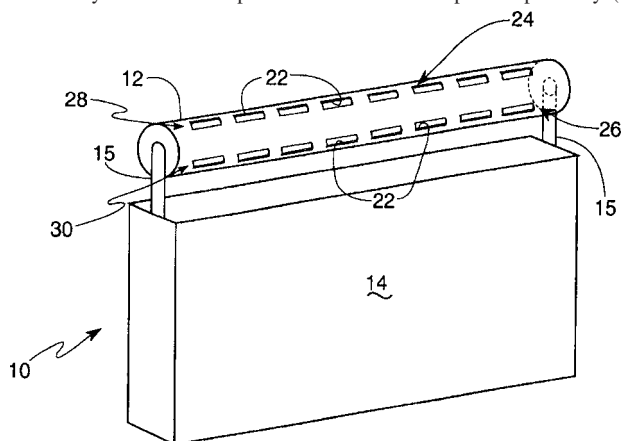
contact will produce noise when treads **10**, **20** are worn down by **H**. The author asserts that vent holes **41**, **51** cast at the time of manufacture will prevent that noise increase.—AJC

6,739,554

43.50.Nm AIRCRAFT WEAPONS BAY ACOUSTIC RESONANCE SUPPRESSION SYSTEM

Michael J. Stanek, assignor to The United States of America as represented by the Secretary of the Air Force
25 May 2004 (Class 244/137.1); filed 2 June 2003

A system is described for suppressing acoustic resonances caused by flow over a cavity such as the weapons bay of an aircraft. It consists of a perforated rod deployed near the upstream edge of the cavity that is designed to generate high-frequency vortices that suppress acoustic resonances in the cavity. At low flow speeds this can be accomplished passively (see



United States Patent 6,446,904). An active system for high-speed flow consists of two rows of holes, driven by synthetic jets that induce Karmanlike alternating vortex shedding to suppress cavity resonances.—KPS

7,234,663

43.55.Ti BUBBLE CLOUD ACOUSTIC DAMPING FOR LAUNCH VEHICLE FAIRING

John E. Higgins, assignor to The United States of America as represented by the Secretary of the Air Force
26 June 2007 (Class 244/1 N); filed 14 June 2005

Launch vehicle fairings often consist of an inner and an outer skin. Because a simple double-skin arrangement provides minimal protection of the payload from exterior noise, the space between the skins is often filled with water. Air bubbles are introduced into the water in order to increase the impedance mismatch between the water and the skins, thereby increasing a fairing's noise reduction. The water can be ejected after the greatest acoustic loads on the fairing have subsided, so as to improve the payload lift capacity.—EEU

6,778,672

43.60.Rw AUDIO RECEPTION CONTROL ARRANGEMENT AND METHOD FOR A VEHICLE

David S. Breed *et al.*, assignors to Automotive Technologies International Incorporated
17 August 2004 (Class 381/86); filed 20 April 2001

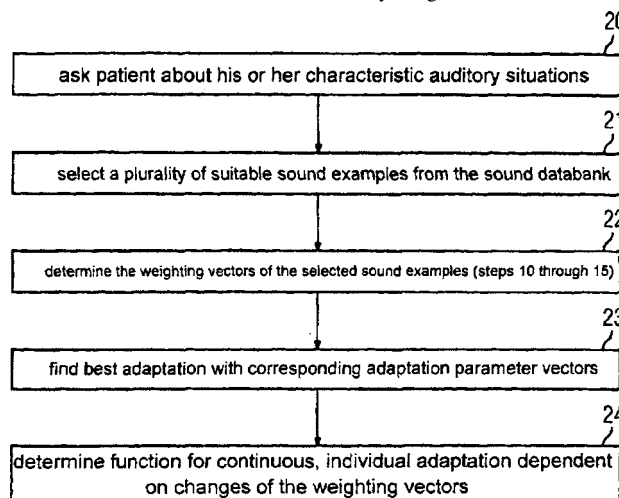
This massive patent describes a system that uses ultrasonic emitters and sensors to determine, among other things, the presence and positions of a vehicle's occupants. Such a system can be used in combination with the air bag system to adjust the bag inflation. The performance of the sound system can be optimized for the known occupants. Furthermore, such ultrasonic systems can be used for numerous other purposes, ranging from determining seat and window positions and driver's eye positions to recognizing a particular driver. This information can then interact with other systems in the vehicle. The breadth of this patent is truly astounding.—KPS

7,236,603

43.66.Ts DEVICE AND METHOD TO ADAPT A HEARING DEVICE

Patrick Mergell, assignor to Siemens Audiologische Technik GmbH
26 June 2007 (Class 381/312); filed in Germany 30 September 2002

The performance of hearing aids may be adjusted to make listening easier, depending on the auditory environment surrounding the wearer. To reduce errors in detecting the listening environment, stored evaluation data are provided for audio signals characteristic of predetermined auditory situations. These evaluation data are continuously weighted in accordance with



the listening requirements of individual hearing-device users. Preference in weighting is given to the most frequent auditory situations encountered by the wearer.—DAP

7,236,604

43.66.Ts HEARING AID DEVICE

Der-Yang Tien, Taipei, Taiwan
26 June 2007 (Class 381/312); filed in Taiwan 15 October 2003

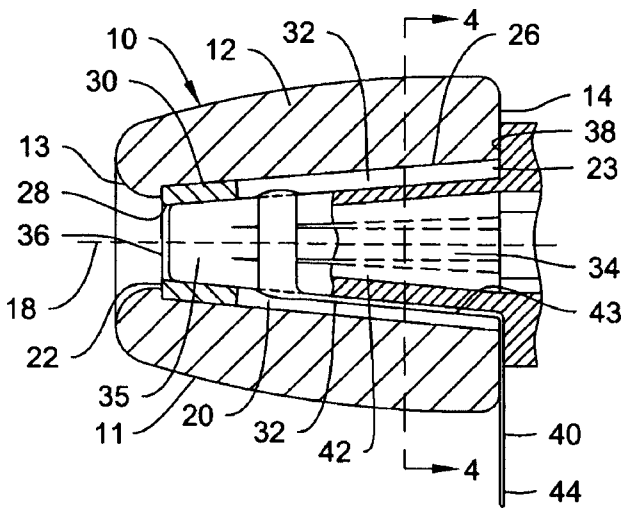
A remote input device allows the hearing aid wearer to enter data and commands via either wired or wireless means to the digital hearing aid. The commands adjust the volume, speed, and tone of the output sound from the digital sound processing module. A translation system is also described for converting the received digital sound data to other languages.—DAP

7,236,605

43.66.Ts USER DISPOSABLE SLEEVE FOR USE WITHIN THE EAR CANAL

Robert J. Oliveira *et al.*, assignors to Hearing Components, Incorporated
26 June 2007 (Class 381/328); filed 5 December 2003

A user-disposable sleeve contains an inner portion that is releasably



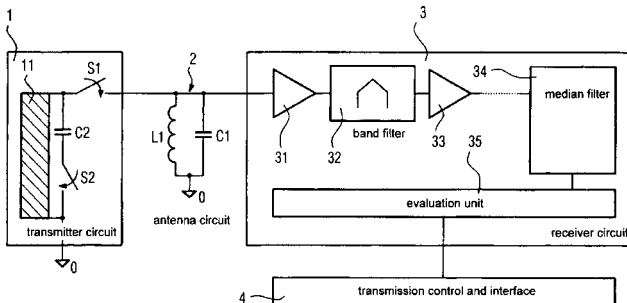
attached to elongated sound controlling devices and a compressible foam outer portion adapted to fit within a user's ear canal.—DAP

7,239,713

43.66.Ts WIRELESS TRANSMISSION SYSTEM FOR HEARING DEVICES

Torsten Niederdränk, assignor to Siemens Audiologische Technik GmbH
 3 July 2007 (Class 381/317); filed in Germany 30 September 2002

A radio for a hearing device uses an antenna with a self-exciting inductance/capacitance oscillator circuit that performs communication with other hearing devices containing similar componentry. The advantage is that



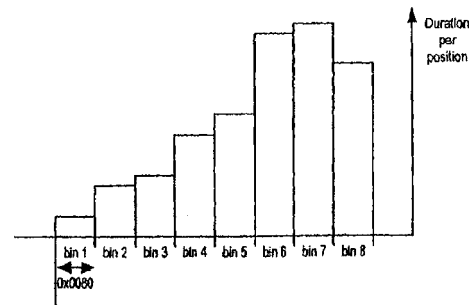
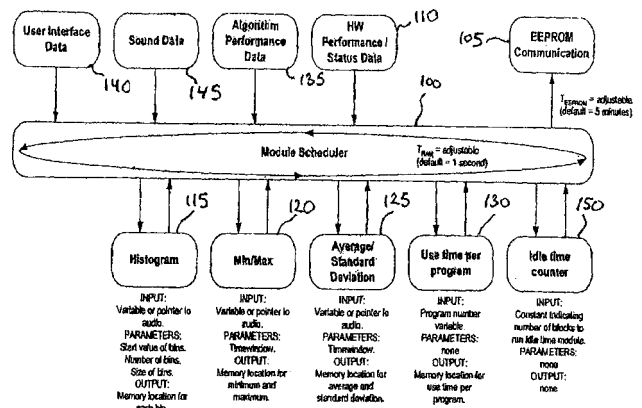
a crystal oscillator, which would occupy significant space, is not required. The system preferably implements a low-frequency, inductive, half-duplex wireless link.—DAP

7,242,777

43.66.Ts DATA LOGGING METHOD FOR HEARING PROSTHESIS

Joseph Renier Gerardus Maria Leenen and Rudie Adriaan Landman, assignors to GN ReSound A/S
 10 July 2007 (Class 381/60); filed in Denmark 30 May 2002

This patent covers basic event-driven data logging methodology, including monitoring an input signal and comparing it to a predetermined signal criterion. At least one hearing prosthetic variable and data derived from the input signal are recorded in nonvolatile memory if the input signal



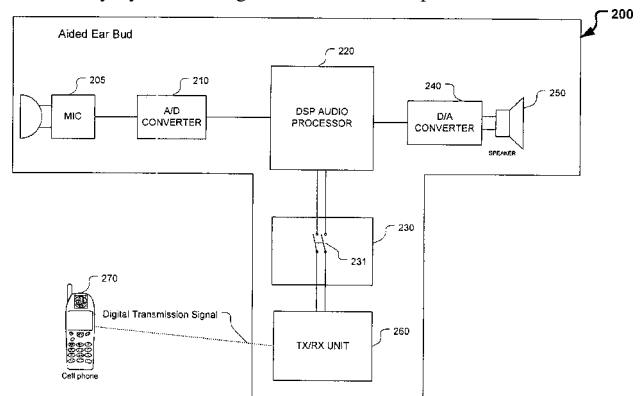
data matches a predetermined signal criterion or at least one hearing prosthetic variable matches a predetermined variable criterion. The variable(s) may be associated with adaptive feedback cancellation, multiband dynamic range compression, noise-reduction, and beam-forming algorithms.—DAP

7,245,730

43.66.Ts AIDED EAR BUD

Mark Enzmann, assignor to Cingular Wireless II, LLC
 17 July 2007 (Class 381/315); filed 13 January 2003

A wireless earbud operates in default mode as a hearing aid while also optionally providing means for a hearing aid wearer to communicate with a mobile phone. The operating mode is changed from hearing aid to earbud automatically by a switching command in the input from a mobile device.



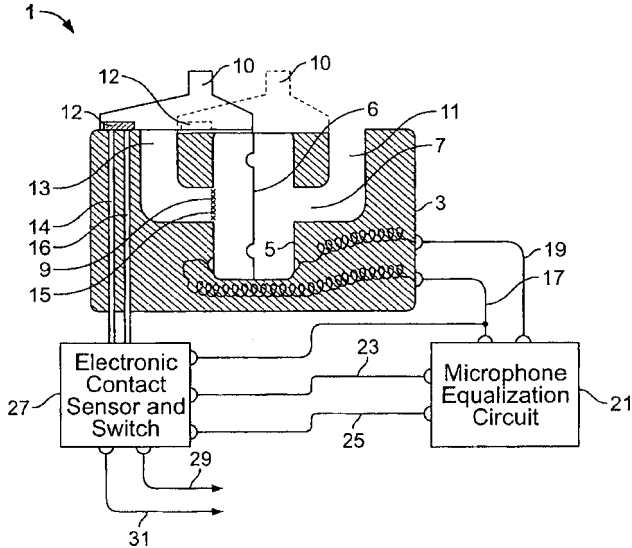
The microphone in the earbud picks up the wearer's voice, which is transmitted to the mobile device. The signal from the mobile device is mixed with the earbud microphone and compensation is applied as dictated by the hearing loss of the wearer before routing the combined signal to the earbud speaker.—DAP

7,245,728

43.66.Ts MICROPHONE FOR HEARING AID AND COMMUNICATIONS APPLICATIONS HAVING SWITCHABLE POLAR AND FREQUENCY RESPONSE CHARACTERISTICS

Mead C. Killion *et al.*, assignors to Etymotic Research, Incorporated
 17 July 2007 (Class 381/122); filed 5 April 2005

A hearing aid microphone assembly operates in either directional mode or omnidirectional mode. The omnidirectional mode is selected by means of a switch that causes the rear inlet tube of the microphone system to be plugged. When the switch leaves the rear inlet tube uncovered, the



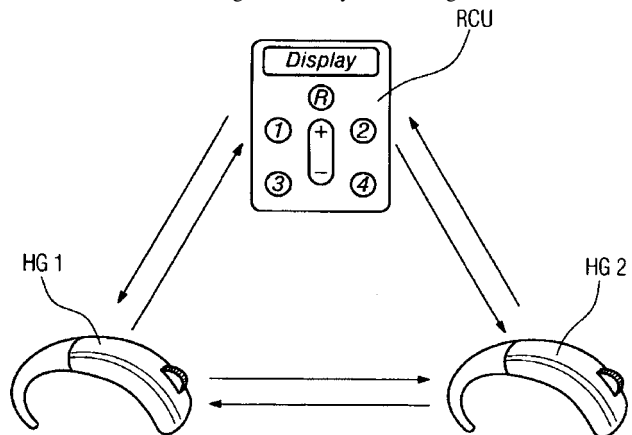
directional mode is selected. A circuit senses the switch position so that the frequency responses in the directional and omnidirectional mode are equalized.—DAP

7,245,731

43.66.Ts DEVICE AND METHOD FOR COMMUNICATION OF HEARING AIDS

Torsten Niederdränk and Gottfried Rückerl, assignors to Siemens Audiologische Technik GmbH
 17 July 2007 (Class 381/315); filed in Germany 5 February 2003

A low-power communication protocol in a wireless binaural hearing aid fitting gives higher priority to transmissions from a remote control over those from the other hearing aid. The system distinguishes remote control



transmissions by their longer-length data packets. The ear-to-ear communication between the two hearing aids is temporarily interrupted if higher-priority transmissions are received from the remote control.—DAP

7,245,732

43.66.Ts HEARING AID

Mie Ø. Jørgensen and Lars Bramsløw, assignors to Oticon A/S
 17 July 2007 (Class 381/317); filed in Denmark 17 October 2001

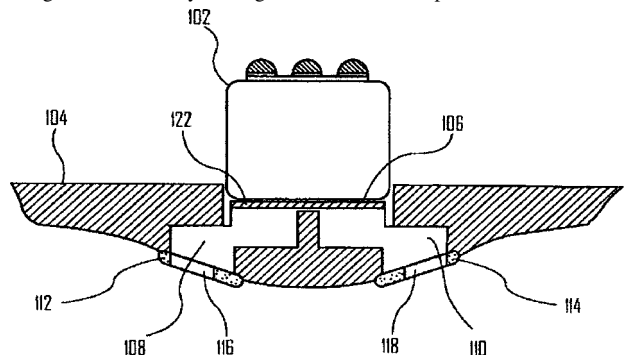
A digital hearing aid system includes an open vent channel of at least 3 mm in diameter, an adaptive feedback-cancellation system, and a processor that provides increased gain in a low-frequency region below 1000 Hz, and primarily below 500 Hz, to make up for the gain lost through the vent. To avoid echo effects, the design is configured so that the signal processing delay is less than 15 ms and preferably even less than 5 ms.—DAP

7,245,733

43.66.Ts HEARING INSTRUMENT MICROPHONE ARRANGEMENT WITH IMPROVED SENSITIVITY

Oleg Saltykov, assignor to Siemens Hearing Instruments, Incorporated
 17 July 2007 (Class 381/322); filed 3 February 2003

To increase sensitivity, an artificially longer port spacing between the front and rear openings for a directional microphone system in a custom hearing aid is created by closing off with a screen a portion of at least one of



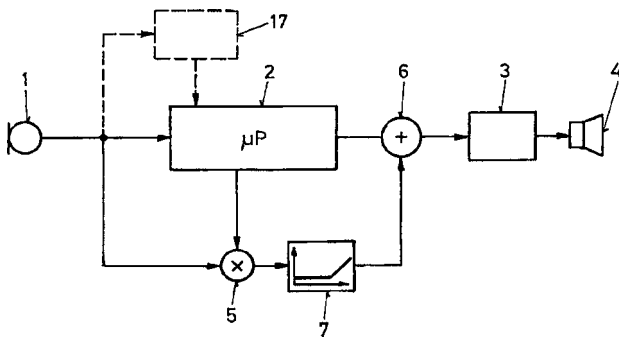
the front and rear ports. The microphone system may consist of several microphone cartridges and an omnidirectional and directional microphone.—DAP

7,248,710

43.66.Ts EMBEDDED INTERNET FOR HEARING AIDS

Hans-Ueli Roeck, assignor to Phonak AG
 24 July 2007 (Class 381/312); filed 5 February 2004

The input signal to a hearing aid is routed to at least two paths—one or more low-group-delay side paths, each containing a simple filter having frequency-dependent gain, which are then summed with the output of the longer-group-delay, fast Fourier transform-based main multiband signal processing path. The low group delay in the side signal paths is said to improve



the wearer's ability to localize sounds by improved processing of transient signals and better utilization of the precedence effect. A silence detector may be used to switch off the entire main signal path and leave only the side paths.—DAP

7,248,711

43.66.Ts METHOD FOR FREQUENCY TRANSPOSITION AND USE OF THE METHOD IN A HEARING DEVICE AND A COMMUNICATION DEVICE

Silvia Allegro *et al.*, assignors to Phonak AG
24 July 2007 (Class 381/316); filed 5 March 2004

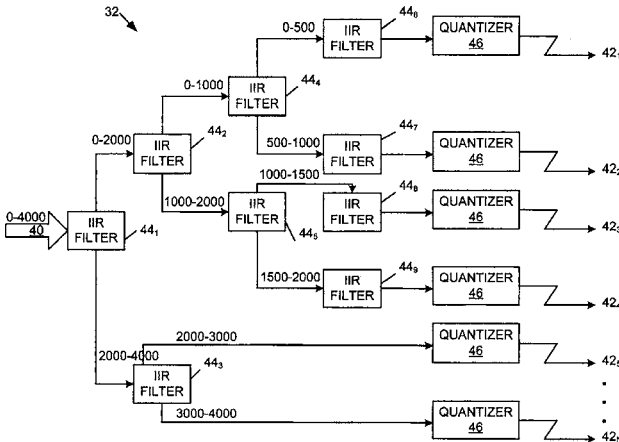
A system processes lower frequencies nearly linearly to preserve harmonic relationships, while a perception-based nonlinear frequency transposition algorithm also moves higher-frequency components to a lower-frequency region that is audible to the hearing device user. Claims are provided that relate to wireless transmission of a reduced bandwidth signal between two hearing devices in a binaural hearing aid fitting. The compressed received signal can be used directly or restored to the original.—DAP

7,239,253

43.72.Gy CODEC SYSTEM AND METHOD

Karl Denninghoff, assignor to Intel Corporation
3 July 2007 (Class 341/50); filed 16 September 2004

A software-based subband audio codec has two banks of filters that produce more subbands for lower-frequency signals than for higher-frequency signals. After several quantizers receive a particular subband from a particular filter, the quantized subbands are combined and the compressed signal is transmitted over a communication link, together with the filter state



at the end of each sound packet. At the receiving side, decompression is performed via reverse quantization and a bank of filters reconstructs the subband signals. Each filter receives a filter state from the compressed data stream to reconstruct the data packet sequence.—DAP

7,107,215

43.72.Ja DETERMINING A COMPACT MODEL TO TRANSCRIBE THE ARABIC LANGUAGE ACOUSTICALLY IN A WELL DEFINED BASIC PHONETIC STUDY

Mikhail E. Ghali, assignor to Sakhr Software Company
12 September 2006 (Class 704/257); filed 16 April 2001

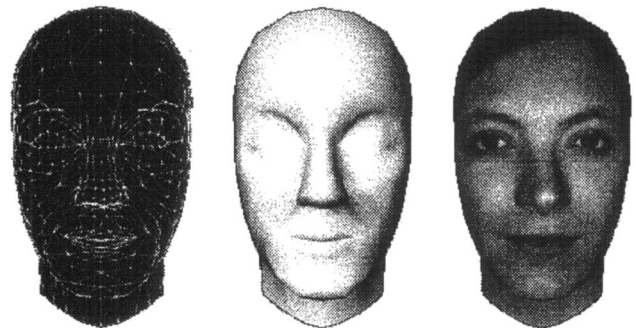
Applicable to the field of phonetic transcription, this patent describes a process for studying a language and for developing a character set for that language to be used in computer processing. The general procedure presented here is to first study the language from both written and phonetic points of view to develop a comprehensive maximal phonetic character set to include all dialects and variations of the language. Choosing a specific subset, such as modern standard Arabic as described here, the maximal set is then reduced to a working phonetic symbol set suitable for computer speech processing, including both recognition and dictation. Phoneticians who expect that the results of their work may be applied to a commercial product would do well to note the generality of the process patented here.—DLR

7,123,262

43.72.Ne METHOD OF ANIMATING A SYNTHESIZED MODEL OF A HUMAN FACE DRIVEN BY AN ACOUSTIC SIGNAL

Gianluca Francini *et al.*, assignors to Telecom Italia Lab S.p.A.
17 October 2006 (Class 345/473); filed in Italy 31 March 2000

Several systems have been patented for the generation of animated video images of a talking face or head based on the analysis of an audio signal. The patent includes references to several such systems that do not use the viseme set as described in the MPEG-4 video coding standard. The



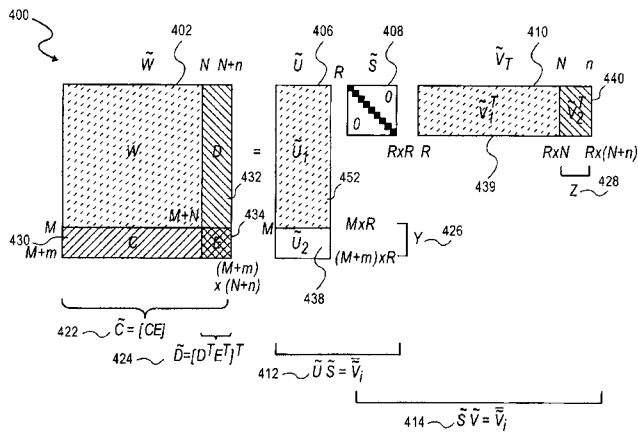
patent describes the process of associating the phonetic sequences with facial animation sequences, covering the animation process in some considerable detail. Of the two dependent claim sets, one set (seven claims) involves compatibility with Standard ISO/IEC 14496 VER. 1. The other two claims refer to active shape model (ASM) parameters.—DLR

7,124,081

43.72.Ne METHOD AND APPARATUS FOR SPEECH RECOGNITION USING LATENT SEMANTIC ADAPTATION

Jerome R. Bellegarda, assignor to Apple Computer, Incorporated
17 October 2006 (Class 704/255); filed 28 September 2001

This patent discloses the application of the latent semantic analysis (LSA) model, covered by United States Patents 4,839,853 and 6,208,971, among many others, to the semantic analysis aspects of speech recognition. Specifically, the patent deals with methods for adapting the LSA model to



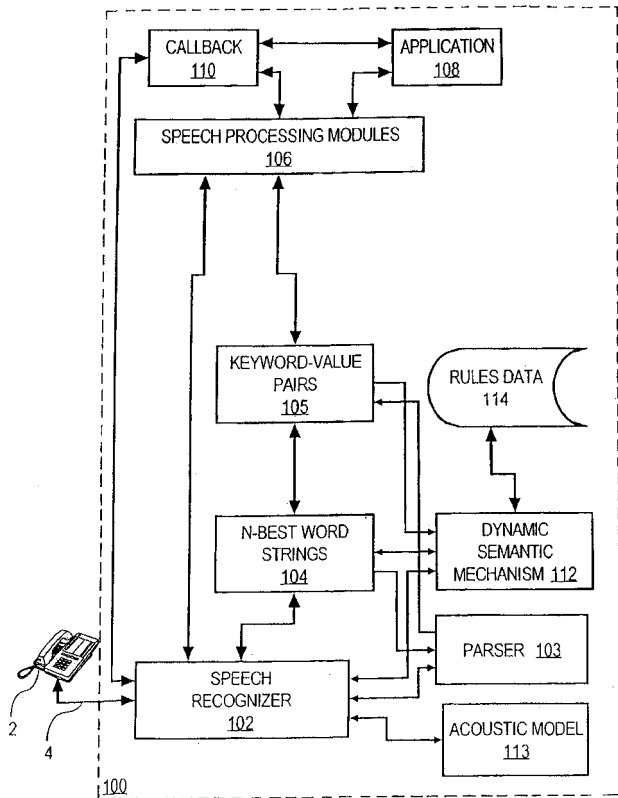
cover novel contexts or semantic styles, while keeping down the amount of computation required. Methods and matrix operations are described for adapting the LSA space in a two-sided manner, taking into account both new words and new documents (contexts), or in a one-sided manner, considering only new documents, but discarding any newly encountered words.—DLR

7,127,393

43.72.Ne DYNAMIC SEMANTIC CONTROL OF A SPEECH RECOGNITION SYSTEM

Michael S. Phillips *et al.*, assignors to Speech Works International, Incorporated
24 October 2006 (Class 704/240); filed 10 February 2003

The speech recognition system presented here is intended as an improvement over prior systems based on *n*-gram word-statistics models. The problem with such prior art models, as depicted here, is that the word-sequence statistics are not context dependent, and thus fail to make use of any gains to be had by better context sensitivity. A method is presented for adjusting the word-sequence probabilities for an utterance as computed



7,127,394

43.72.Ne ASSIGNING MEANINGS TO UTTERANCES IN A SPEECH RECOGNITION SYSTEM

Robert Don Strong, assignor to Apple Computer, Incorporated
24 October 2006 (Class 704/255); filed 18 February 2004

This speech recognition system uses a dynamically constructed set of finite-state grammar models to improve the recognition accuracy. As soon as the user begins to speak, the recognizer determines what application software is currently running and constructs a language model appropriate to that application. The language model consists of a set of rules, each containing a finite-state grammar and a semantic structure identifying the action to be taken if that grammar is satisfied by an incoming utterance. As soon as the language model is complete, processing of the incoming utterance proceeds, using the grammar models in the newly constructed rule set to identify words and phrases in the incoming speech. As soon as the incoming utterance ends, the semantic structures used for parsing the utterance may then immediately be applied by performing the specified actions.—DLR

7,149,695

43.72.Ne METHOD AND APPARATUS FOR SPEECH RECOGNITION USING SEMANTIC INFERENCE AND WORD AGGLOMERATION

Jerome R. Bellegarda, assignor to Apple Computer, Incorporated
12 December 2006 (Class 704/275); filed 13 October 2000

This is another application of the latent semantic analysis (LSA) method described in United States Patent 7,124,081, reviewed above, to speech recognition. Here, the application is a voice command module, allowing the user to operate the computer by voice commands. The LSA process is followed to construct a matrix of command-specific word sets versus the various software applications that are set up to use the voice command capability. However, the standard LSA model makes no use of word order. As modified here, that model is augmented by recalling a set of *n*-gram word sequences corresponding to each application. To that end, the command-specific word sets used to form the LSA matrix consist of the *n*-gram word sequences rather than the words themselves.—DLR

7,240,002

43.72.Ne SPEECH RECOGNITION APPARATUS

Katsuki Minamino *et al.*, assignors to Sony Corporation
3 July 2007 (Class 704/253); filed in Japan 7 November 2000

The likelihood of a correct speech recognition result is calculated using extracted feature values and is used to improve speech recognition accuracy. A word series is produced as a candidate for the speech recognition result, and the connection relationships in the word series are stored, evaluated, and corrected for word boundary locations.—DAP

7,235,052

43.80.Vj ULTRASOUND DETECTABLE INSTRUMENT WITH BUBBLE GENERATOR

Ewen James Crawford Kellar and Seyed Mehdi Tavakoli, assignors to BTG International Limited
26 June 2007 (Class 600/458); filed in United Kingdom 28 April 1999

This instrument contains two contacting elements within a fluid-permeable carrier material. The elements produce gas bubbles in the presence of a fluid. The bubbles enhance ultrasound visualization of the instrument.—RCW

7,238,157

43.80.Vj BROAD-BEAM IMAGING METHODS

Glen McLaughlin *et al.*, assignors to Zonare Medical Systems, Incorporated
3 July 2007 (Class 600/443); filed 16 January 2004

Waveforms resulting from broad-beam transmission and reception of ultrasound are processed using phase and magnitude to produce echolocation data. Since the approach does not depend on using narrow transmit or receive beams, a multidimensional area can be imaged using a single transmission. The approach allows an increase in the imaging frame rate and a reduction in the ultrasound energy required to produce an image.—RCW

7,238,158

43.80.Vj PULSE INTERLEAVING IN DOPPLER ULTRASOUND IMAGING

Kenneth Abend, assignor to Allez Physionix, Limited
3 July 2007 (Class 600/454); filed 26 January 2004

A two-dimensional array of piezoelectric transducers is used for three-dimensional imaging and monitoring of blood flow. Array thinning, large elements, and limited scanning are also used. The three-dimensional veloc-

ity of blood flow rather than only one component of velocity is determined. Post-Doppler processing is used to achieve high-resolution vessel tracking and flow mapping.—RCW

7,242,793

43.80.Vj PHYSICALLY-BASED, PROBABILISTIC MODEL FOR ULTRASONIC IMAGES INCORPORATING SHAPE, MICROSTRUCTURE AND SYSTEM CHARACTERISTICS

Jason W. Trobaugh and R. Martin Arthur, assignors to Washington University
10 July 2007 (Class 382/128); filed 23 April 2001

The microstructure of a gross shape is modeled by multiple discrete scatterers distributed across the surface of the shape. The model is formed using the mean, variance, and ratio of the mean to standard deviation of the amplitude computed at each pixel in an image. Based on the results of the computations, each pixel is assigned a Raleigh or Gaussian density function. A product of the assigned density functions is used to define likelihood values from which an image is derived.—RCW

7,248,725

43.80.Vj METHODS AND APPARATUS FOR ANALYZING ULTRASOUND IMAGES

Gil Zwirn and Solange Akselrod, assignors to Ramot At Tel Avia University Limited
24 July 2007 (Class 382/128); filed 1 November 2005

An ultrasound image of a volume is analyzed by dividing a region of interest in the image into segments. For each segment, representative intensity values are computed and corrected for attenuation to represent different tissues. A representative intensity for each segment is used to map the region of interest into different tissue types.—RCW

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.

Enhanced sound transmission from water to air at low frequencies (L)

B. Edward McDonald^{a)} and David C. Calvo
Naval Research Laboratory, Washington DC 20375

(Received 30 March 2007; revised 1 August 2007; accepted 11 September 2007)

Excitation of acoustic radiation into the air from a low-frequency point source under water is investigated using plane wave expansion of the source spectrum and Rayleigh reflection/transmission coefficients. Expressions are derived for the acoustic power radiated into air and water as a function of source depth and given to lowest order in the air/water density ratio. Near zero source depth, the radiation into the water is quenched by the source's acoustic image, while the power radiated into air reaches about 1% of the power that would be radiated into unbounded water. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2793709]

PACS number(s): 43.30.Dr, 43.30.Jx [JGM]

Pages: 3159–3161

I. INTRODUCTION

A recent article¹ has pointed out that a low frequency monopole at a small depth underwater can radiate the majority of its energy into the air. While this is technically true, one must realize that the energy radiated into the water falls almost to zero as the source depth approaches zero because of the source's negative acoustic image above the water. This letter derives expressions for the energy radiated from water into air to lowest order in the air-to-water density ratio. As the source depth approaches zero, the energy radiated into air is shown to be about 1% of the energy that would be radiated in unbounded water. Unless one compares the energy radiated into water and air with the unbounded water case, it can be misleading to describe the air-water interface as *anomalously transparent* as implied in Ref. 1.

II. THEORY

The theoretical approach to be taken is similar to that in Ref. 1: The point source is resolved into plane waves, and then Rayleigh reflection and transmission coefficients are used to synthesize the acoustic Poynting vector² for energy radiated into the air and to depth in water. Unlike Ref. 1, however, we give asymptotic expressions for the energy radiated into air rather than going directly from Poynting's theorem to numerical results.

In unbounded water, the pressure field of a point source with time dependence $\exp(-i\omega t)$ normalized to pressure p_0 at radius r_0 and its spatial Fourier transform are given by

$$p = p_0 r_0 \frac{\exp(ik_w r)}{r} = \frac{p_0 r_0}{2\pi^2} \int \exp(i\mathbf{q} \cdot \mathbf{r}) \times \left[(q^2 - k_w^2)^{-1} + \frac{i\pi}{2k_w} \delta(q - k_w) \right] d^3\mathbf{q}, \quad (1)$$

where $k_w = \omega/c_w$, c_w is sound speed in water, and $q = |\mathbf{q}|$. The integral in Eq. (1) is the Cauchy principal value, resulting from a contour integration as described by Chew.³

We adopt a Cartesian coordinate system and place the point source at $(x, y, z) = (0, 0, -D)$ in anticipation of adding an air-water interface at $z=0$. Carrying out the q_z integration in Eq. (1), the acoustic field without the interface is

$$p = \frac{p_0 r_0}{2\pi^2} \int d^2\mathbf{q}_\perp \exp(i\mathbf{q}_\perp \cdot \mathbf{r}) F(z, q_\perp), \quad (2)$$

$$F(z, q_\perp) = \begin{cases} i\pi u^{-1} \exp i|z+D|u, & q_\perp^2 < k_w^2 \\ \pi u^{-1} \exp -|z+D|u, & q_\perp^2 > k_w^2, \end{cases} \quad (3)$$

where $u = \sqrt{k_w^2 - q_\perp^2}$ and $\mathbf{q}_\perp = (q_x, q_y)$.

A. Addition of an air-water interface

An air-water interface at $z=0$ is added by use of appropriate plane wave reflection coefficients. Snell's law dictates constancy of the horizontal wave number \mathbf{q}_\perp across the interface. Three horizontal wave number regimes are considered depending on the nature of the waves involved: (A) waves which are oscillatory in air and water; (B) waves which are evanescent in water but oscillatory in air; and (C)

^{a)}Electronic mail: mcdonald@sonar.nrl.navy.mil

those which are evanescent in both water and air. The vertical dependence of the pressure field is then modified from Eq. (3) for each wave number regime.

(A) For $q_{\perp}^2 < k_w^2$:

$$z < 0: F(z, q_{\perp}) \rightarrow i\pi u^{-1}(\exp(i|D+z|u) + R \exp(i(D-z)u)),$$

$$z > 0: F(z, q_{\perp}) \rightarrow i\pi u^{-1}(1+R)\exp(iDu)\exp(isz), \quad (4)$$

where the Rayleigh reflection coefficient is

$$R = \left(1 - \frac{\rho_w s}{\rho_a u}\right) / \left(1 + \frac{\rho_w s}{\rho_a u}\right),$$

$$s = \sqrt{k_{\alpha}^2 - q_{\perp}^2}, \quad (5)$$

and ρ_w, ρ_{α} are water and air densities.

(B) For $k_{\alpha}^2 > q_{\perp}^2 > k_w^2$:

$$z < 0: F(z, q_{\perp}) \rightarrow \pi u^{-1}(\exp(-|z+D|u) + R \exp(-(D-z)u)),$$

$$z > 0: F(z, q_{\perp}) \rightarrow \pi(1+R)u^{-1}\exp(-Du)\exp(isz), \quad (6)$$

$$R = \left(1 + i\frac{\rho_w s}{\rho_a u}\right) / \left(1 - i\frac{\rho_w s}{\rho_a u}\right). \quad (7)$$

(C) For $q_{\perp}^2 > k_{\alpha}^2$: Energy is not radiated into the air.

B. Acoustic power transmission from water to air

The time averaged acoustic Poynting vector² for a harmonic pressure field is

$$\langle \mathbf{I} \rangle = \text{Im} \frac{1}{2\rho\omega} p^* \nabla p, \quad (8)$$

where the asterisk denotes complex conjugation. Substituting Eq. (2) into Eq. (8) and integrating over the horizontal plane on the air side of the interface yields the acoustic power radiated into the air:

$$\begin{aligned} W_{\alpha} &= \hat{\mathbf{z}} \cdot \int \langle \mathbf{I} \rangle dx dy \\ &= \frac{1}{2\rho_a \omega} \frac{p_0^2 r_0^2}{(2\pi^2)^2} \text{Im} \int dx dy \int d^2 \mathbf{q}_{\perp} \int d^2 \mathbf{q}'_{\perp} \\ &\quad \times \exp(-i(\mathbf{q}_{\perp} - \mathbf{q}'_{\perp}) \cdot \mathbf{r}) F^*(0, q_{\perp}) \partial_z F(0, q'_{\perp}). \end{aligned} \quad (9)$$

It is convenient to normalize (9) to the power radiated by the point source (1) in unbounded water:

$$W_0 = 2\pi \frac{p_0^2 r_0^2}{\rho_w c_w}. \quad (10)$$

The integral $\int dx dy \exp(-i(\mathbf{q}_{\perp} - \mathbf{q}'_{\perp}) \cdot \mathbf{r})$ in Eq. (9) yields $(2\pi)^2 \delta(\mathbf{q}_{\perp} - \mathbf{q}'_{\perp})$, and Eqs. (9) and (10) give

$$\begin{aligned} \frac{W_{\alpha}}{W_0} &= \frac{\rho_w}{4\pi^3 \rho_a k_w} \int d^2 \mathbf{q}_{\perp} \text{Im} F^*(0^+, q_{\perp}) \partial_z F(0^+, q_{\perp}) \\ &= \frac{\rho_w}{4\pi^3 \rho_a k_w} \text{Im} \left[\int_0^{k_w} 2\pi q_{\perp} dq_{\perp} |1+R|^2 \pi^2 is/u^2 \right. \\ &\quad \left. + \int_{k_w}^{k_{\alpha}} 2\pi q_{\perp} dq_{\perp} |1+R|^2 \pi^2 is e^{-2uD}/u^2 \right] \\ &= \frac{\rho_w}{2\rho_a k_w} \left[\int_0^{k_w} q_{\perp} dq_{\perp} \frac{4s}{(u+s\rho_w/\rho_a)^2} \right. \\ &\quad \left. + \int_{k_w}^{k_{\alpha}} q_{\perp} dq_{\perp} \frac{4s}{u^2 + (s\rho_w/\rho_a)^2} e^{-2uD} \right]. \end{aligned} \quad (11)$$

To lowest order in ρ_a/ρ_w , Eq. (11) becomes

$$\begin{aligned} \frac{W_{\alpha}}{W_0} &= \frac{2\rho_a}{k_w \rho_w} \left[\int_0^{k_w} q_{\perp} / \sqrt{k_{\alpha}^2 - q_{\perp}^2} dq_{\perp} + \int_{k_w}^{k_{\alpha}} q_{\perp} \right. \\ &\quad \left. \times \exp(-2D\sqrt{q_{\perp}^2 - k_w^2}) / \sqrt{k_{\alpha}^2 - q_{\perp}^2} dq_{\perp} \right] + O\left(\frac{\rho_a}{\rho_w}\right)^2. \end{aligned} \quad (12)$$

Introducing a new integration variable $v = \phi^{-1} \sqrt{c_w^2/c_a^2 - q_{\perp}^2/k_w^2}$, where $\phi = \sqrt{c_w^2/c_a^2 - 1}$, Eq. (12) becomes

$$\frac{W_{\alpha}}{W_0} \simeq \frac{2\rho_a}{\rho_w} \left[\frac{c_w}{c_a} - \phi + \phi \int_0^1 \exp(-2k_w D \phi \sqrt{1-v^2}) dv \right]. \quad (13)$$

The first two terms in the square brackets come from the oscillatory portion of the point source spectrum, while the integral comes from the evanescent part. Equation (13) is integrated numerically and plotted as the solid curve in Fig. 1 using $c_w = 1500$ m/s, and sea level air values for 20 °C: $\rho_a/\rho_w = 1.20 \times 10^{-3}$ and $c_a = 344$ m/s. For dimensionless source depths $k_w D$ greater than about 2, the power radiated to air is almost entirely due to the oscillatory part of the source spectrum, giving $W_{\alpha}/W_0 \rightarrow (2\rho_a/\rho_w)(c_w/c_a - \phi) \simeq 3 \times 10^{-4}$. As the depth approaches zero, the evanescent portion of the source spectrum dominates the radiation into air, with $W_{\alpha}/W_0 \rightarrow 2\rho_a c_w / (\rho_w c_a) \simeq 0.0105$ (see Fig. 1). For comparison, the fraction of the energy flux of plane waves at normal incidence to the interface transmitted to the air is $4c_a \rho_a / (c_w \rho_w) \simeq .0011$.

C. Acoustic power radiated to depth in water

Using Eqs. (4) and (5) then following the procedure in Eqs. (8)–(11) to find the acoustic power W_w radiated downward into the water, one carries out integration over a horizontal plane for $z < -D$ to find

$$\frac{W_w}{W_0} = \frac{1}{2k_w} \int_0^{k_w} q_{\perp} dq_{\perp} u^{-1} |\exp(-iDu) + R \exp(iDu)|^2. \quad (14)$$

To lowest order in ρ_a/ρ_w Eq. (14) becomes

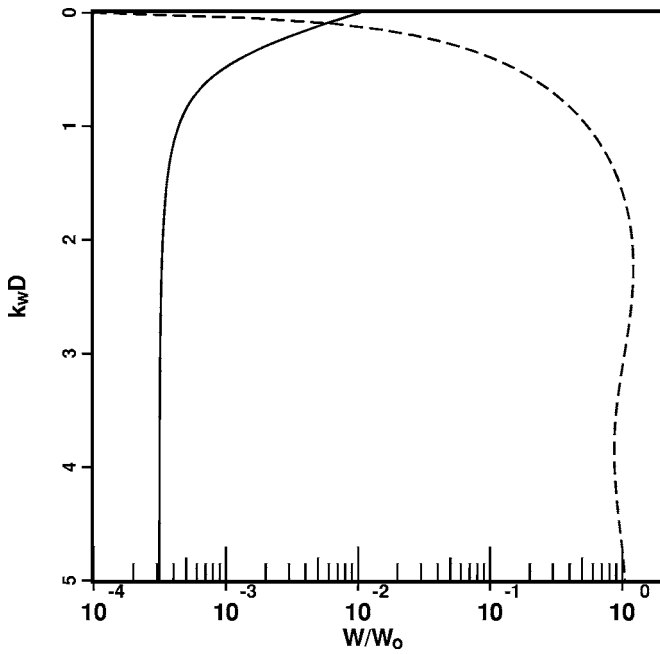


FIG. 1. Power radiated into air (solid) vs dimensionless depth $k_w D$ from Eq. (13) and into water (dashed) from Eq. (15) normalized to the unbounded water value W_0 .

$$\frac{W_w}{W_0} = 1 - \frac{\sin 2k_w D}{2k_w D} - 2 \frac{\rho_a}{\rho_w} \times \int_0^1 v dv (1 - \cos 2k_w D v) / \sqrt{\phi^2 + v^2}. \quad (15)$$

When $D=0$, the lowest order surviving term is quadratic in ρ_a/ρ_w and c_a/c_w so that Eq. (14) leads to

$$\frac{W_w(D=0)}{W_0} \rightarrow \frac{2}{3} \left(\frac{\rho_a c_a}{\rho_w c_w} \right)^2 = 5.0 \times 10^{-8}. \quad (16)$$

The normalized power radiated into water (or equivalently the radiation efficiency of the point source) from Eq. (15) is plotted as a dashed line in Fig. 1. The density ratio term in Eq. (15) has been included for completeness, but its amplitude is so small that its presence or absence makes no discernible difference to Fig. 1.

III. SUMMARY

As a low frequency point source underwater approaches the air-water interface, its efficiency of radiation into water drops nearly to zero because of destructive interference from its image above the water. On the other hand, the power radiated into air increases as the evanescent portion of the source spectrum is able to interact with the interface and excite oscillatory waves in the air. As shown in Fig. 1, the normalized power into water and air become equal at $k_w D \approx 0.093$, corresponding to a depth of 0.015 acoustic wavelengths. For source depths less than this value, the majority of radiated energy is into the air, approaching about 1.1% of the power that would be radiated in unbounded water. We recommend that this phenomenon be described as enhanced transmission into air rather than the possibly misleading phrase anomalous transparency used in Ref. 1.

ACKNOWLEDGMENT

Work supported by the Office of Naval Research.

¹O. A. Godin, "Anomalous transparency of water-air interface for low-frequency sound," *Phys. Rev. Lett.* **97**, 164301 (2006).

²A. D. Pierce, *Acoustics* (Acoustical Society of America, Melville, NY, 1989), Sec. 1-11.

³W. C. Chew, *Waves and Fields in Inhomogeneous Media* (Van Nostrand Reinhold, New York, 1990), Sec. 2.2.2.

Negative axial radiation forces on solid spheres and shells in a Bessel beam (L)

Philip L. Marston^{a)}

Department of Physics and Astronomy, Washington State University, Pullman, Washington 99164-2814.

(Received 23 May 2007; revised 20 September 2007; accepted 21 September 2007)

Prior computations predict that fluid spheres illuminated by an acoustic Bessel beam can be subjected to a radiation force directed opposite the direction of beam propagation. The prediction of negative acoustic radiation force is extended to the cases of a solid poly(methylmethacrylate) PMMA sphere in water and an empty aluminum spherical shell in water. Compared with the angular scattering patterns for plane wave illumination, the scattering into the back hemisphere is suppressed when the radiation force is negative. This investigation may be helpful in the development of acoustic tweezers and in the development of methods for manipulating objects during space flight. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2799501]

PACS number(s): 43.25.Qp, 43.25.Uv, 43.20.Fn, 43.80.Jz [TDM]

Pages: 3162–3165

I. INTRODUCTION

There has been significant interest in the possibility of using the radiation force of sound beams to pull-on or trap small objects.^{1–6} This is in part because ultrasonic beams might provide an alternative to laser beams^{7,8} for trapping objects of biological or medical interest^{1–4} and because of potential applications in reduced gravity associated with space flight.^{5,9} Recently Marston⁵ noticed situations where axial radiation force computed for an acoustic Bessel beam incident on fluid spheres can be opposite the direction of beam propagation. As in the present letter, it was assumed that the center of the sphere lies on the axis of the Bessel beam and that the effects of dissipation can be neglected. The purpose of this letter is to illustrate some situations where the radiation force is computed to be opposite the direction of Bessel beam propagation for solid spheres and empty spherical shells in water. This condition is referred to as a negative radiation force. Predictions of the type shown here may be helpful for the identification of other situations where ultrasonic beams may be used to pull on spheres or shells. It is important to identify conditions applicable to solid spheres or shells since it may be experimentally easier to verify the existence of negative acoustic radiation forces in beams using solid objects. The Bessel beam solution uses the exact scattering.

A Bessel beam is an axisymmetric solution of the linear Helmholtz equation for which the complex acoustic pressure amplitude is given by^{10–13}

$$p_B(R, z) = p_0 \exp(i\kappa z) J_0(\mu R), \quad (1a)$$

$$\mu = k \sin(\beta), \quad (1b)$$

where p_0 is the axial amplitude, z is the axial coordinate, $R = \sqrt{(x^2 + y^2)}$ is the transverse coordinate, J_0 is a Bessel function, $\kappa = \sqrt{(k^2 - \mu^2)}$, and $k = \omega/c_0$, where c_0 is the speed of sound in the fluid. The physical acoustic pressure is given by

the real part of $p_B(R, z) \exp(-i\omega t)$. The cylindrical location of beam nodes is affected by the cone angle β of the Bessel beam shown in Fig. 1 in Marston.⁵

The present analysis of the radiation force is based on a partial wave series solution for the scattering which is exact in the limit of an inviscid surrounding fluid and a low amplitude Bessel beam.^{12,13} Consequently the solution is not as restricted in the sphere radius as the ray theory for fluid spheres that has been used elsewhere.^{2–4} The measured dependence of the radiation force on the radius of a sufficiently large elastic sphere has been found to agree with partial wave calculations based on the assumption of plane wave illumination in an inviscid fluid.¹⁴ Accordingly, though the calculations given here neglect viscosity, the results may be experimentally relevant to demonstrating negative forces on spheres in beams provided sufficient care is taken in the synthesis of the Bessel beam. (The spatial region where the incident beam is described by Eq. (1) lies close to the source.^{10,11}) Even in the absence of experiments these calculations serve to illustrate the importance of the directional dependence of the scattered ultrasound for situations producing negative forces and this will give insight into other cases.

The incident wave in the present analysis is restricted to the case of a pure Bessel beam containing only a single frequency. However, Leung *et al.*¹⁵ and Silva *et al.*¹⁶ have pointed out that nonlinear aspects of propagation in the surrounding fluid can affect acoustic radiation forces. To reduce the resulting anomalous radiation force on a sphere, Leung *et al.* demonstrated that the second harmonic content of applied sound may be reduced by superposing a second harmonic having the appropriate phase.¹⁵ Low frequency sinusoidal oscillations of the radiation pressure may be induced by applying incident waves having closely spaced frequencies.^{16,17}

II. COMPUTATION OF THE RADIATION FORCE

Relevant results from Marston^{5,12,13} are summarized here. The intensity in the z direction in W/m^2 on the axis of the Bessel beam in Eq. (1) is $I_0 = [p_0^2 / (2\rho_0 c_0)] \cos \beta$ where ρ_0 is the density of the surrounding fluid. Notice that I_0 vanishes

^{a)}Electronic mail: marston@wsu.edu

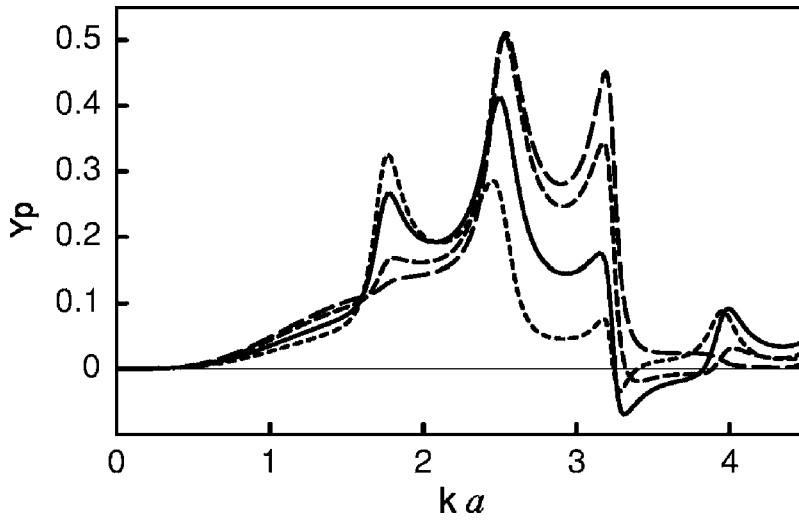


FIG. 1. Dimensionless radiation force Y_p computed from Eq. (3) for an ideal PMMA sphere in water plotted as a function of ka . The values of β for the curves are: 72.5424° (short dashes), 66.4218° (solid curve), 60° (intermediate dashes), and 56.6330° (long dashes). The solid curve has negative Y_p for ka of 3.27–3.81.

when $\beta=90^\circ$ since that limit corresponds to a cylindrical standing wave. The axial radiation force F_z on a sphere of radius a is related to a dimensionless radiation force Y_p by⁵

$$F_z = (\pi a^2)(I_0/c_0)(1/\cos \beta)Y_p(ka, \cos \beta), \quad (2)$$

$$Y_p = - (2/ka)^2 \sum_{n=0}^{\infty} (n+1) [\alpha_n + \alpha_{n+1} + 2(\alpha_n \alpha_{n+1} + \beta_n \beta_{n+1})] P_n(\cos \beta) P_{n+1}(\cos \beta), \quad (3)$$

where P_n is a Legendre polynomial. The coefficients α_n and β_n are related to the dimensionless scattering function s_n for the n th partial wave for plane wave scattering by a sphere by the definitions $\alpha_n = [\text{Re}(s_n) - 1]/2$, and $\beta_n = \text{Im}(s_n)/2$ where Re and Im designate real and imaginary parts. The far field scattered pressure is $p_s(r, \theta) = (a/2r)p_0 F \exp(ikr)$ where r is the radial coordinate and the dependence on the scattering angle θ and the beam parameter β is described by the dimensionless complex form function^{13,14}

$$F(ka, \cos \theta, \cos \beta) = (-i/ka) \sum_{n=0}^{\infty} (2n+1)(s_n - 1)P_n(\cos \theta)P_n(\cos \beta). \quad (4)$$

The normalization of the s_n is defined in such a way that in the absence of absorption by the sphere $|s_n|=1$, however the derivations of Eqs. (3) and (4) are not restricted to that case.^{5,12} In the absence of absorption s_n is related to the denominator D_n in the partial wave series by $s_n = -D_n^*/D_n$ where * denotes complex conjugation. Compact expressions for the D_n for elastic spheres and empty elastic spherical shells used here were reviewed recently by Marston.¹³ The relevant properties of the solid are the density ρ_E and the longitudinal and transverse wave velocities, c_L and c_T . In the numerical evaluations of Eqs. (3) and (4) here, sufficient terms have been included for convergence.¹³ For plane-wave illumination, β vanishes, the function Y_p is non-negative when there is no absorption,⁵ and standard expressions for Y_p and F are recovered.^{5,12,13}

III. NEGATIVE RADIATION FORCE ON A PMMA SPHERE IN WATER

At low ka , a poly(methylmethacrylate) (PMMA) sphere in water exhibits nearly equally spaced resonances associated with a weakly dispersive subsonic wave which circumnavigates the sphere.^{13,18} Each resonance is associated with a specific partial wave in Eq. (4). Inspection of Eq. (4) shows, however, that the cone angle β alters the magnitude and phase of a given partial wave. Marston⁵ found for fluid spheres that negative radiation forces are associated with conditions giving reduced scattering into the back hemisphere. Since it was plausible that this may also be the case for solid spheres, the parameter β was adjusted to search for regions of reduced backscattering or negative Y_p . The parameters of the PMMA and water were the values used by Hefner and Marston:¹⁸ $c_L=2690$ m/s, $c_T=1340$ m/s, and $\rho_E=1.19$ g/cm³ for PMMA, and $c_0=1479$ m/s and $\rho_0=1.00$ g/cm³ for water.

Figure 1 illustrates a region of negative Y_p that was discovered during the aforementioned search where the parameter β was varied. Beginning from the left, the sequence of peaks is associated with the resonance of partial waves having $n=2, 3, 4$, and 5 . The values of $\cos \beta$ for the curves are: 0.3 (short dashes), 0.4 (solid curve), 0.5 (intermediate dashes), and 0.55 (long dashes). The solid curve ($\beta=66.4218^\circ$) has the widest region of negative Y_p . The curve with long dashes ($\beta=56.6330^\circ$) has a non-negative Y_p . The negative extremum in Y_p in the solid curve is at $ka=3.31$.

Figure 2 compares the angular dependence of $|F|$ when $ka=3.31$ for plane wave illumination ($\beta=0$, the solid curve) with the condition for the aforementioned extrema of Y_p ($\beta=66.4218^\circ$, the dashed curve). Inspection of Fig. 2 shows that in agreement with the examples in Marston,⁵ the scattering into the back hemisphere is relatively suppressed at a condition giving a negative Y_p . This is consistent with considerations based on the momentum flux of the scattered wave⁵ and with optical analogies.⁷

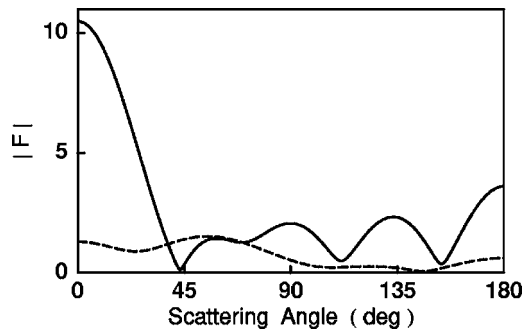


FIG. 2. The angular scattering pattern $|F|$ from Eq. (4) for the sphere considered in Fig. 1 with $ka=3.31$ for plane wave illumination (solid curve) and a Bessel beam with $\beta=66.4218^\circ$ (dashed curve). In the Bessel beam case the scattering into the back hemisphere is suppressed. Inspection of Fig. 1 shows that the Bessel beam case is in a region of negative Y_p .

IV. NEGATIVE RADIATION FORCE ON AN ALUMINUM SHELL IN WATER

In the case of an empty shell the search for conditions for negative Y_p is more complicated because of the presence of an additional parameter: the ratio b/a where b denotes the inner radius of the shell and a denotes the outer radius. Nevertheless, a modest effort found situations where negative Y_p are possible for an empty aluminum shell in water. The parameters used in the computation were: $\rho_E/\rho_0=2.712$, $c_0=1479$ m/s, $c_L=6370$ m/s, and $c_T=3136$ m/s. Negative Y_p are present when $b/a=0.96$, which corresponds to a shell having positive buoyancy. Figure 3 shows two regions having negative Y_p . The solid curve ($\beta=45^\circ$) shows a region of negative Y_p between resonance peaks associated with the $n=2$ and $n=3$ partial waves in Eq. (4). The dashed curve ($\beta=60^\circ$) shows that Y_p is negative below $ka=1.28$. Below ka of 0.65, $|Y_p|$ is less than 0.001.

Figures 4 and 5 illustrate the relationship between the scattering and the region of negative Y_p . Figure 4 is for $ka=1.53$. The solid curve is for plane wave illumination ($\beta=0$). The dashed curve has $\beta=45^\circ$. Comparison shows that the scattering into the back hemisphere is suppressed. The condition $ka=1.53$ and $\beta=45^\circ$ gives $Y_p=-0.02545$. Figure 5 shows an analogous comparison of the backscattering.

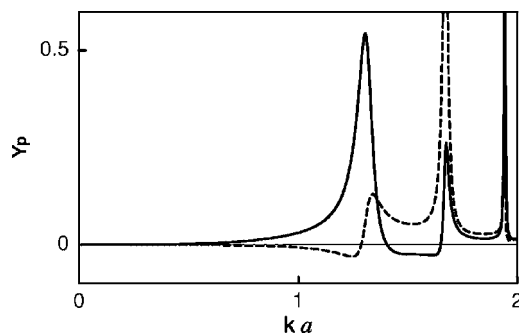


FIG. 3. Radiation force function Y_p for an evacuated aluminum spherical shell in water having an inner-to-outer radius ratio $b/a=0.96$. The Bessel beam parameters are: $\beta=45^\circ$ (solid curve) and $\beta=60^\circ$ (dashed curve). The curves show distinct regions of negative Y_p .

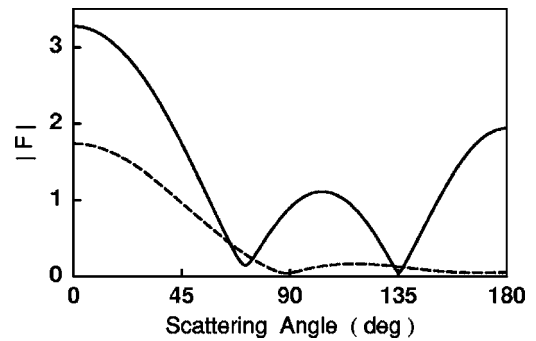


FIG. 4. Angular scattering patterns $|F|$ for the shell considered in Fig. 3 with $ka=1.53$ for plane wave illumination (solid curve) and a Bessel beam with $\beta=45^\circ$ (dashed).

V. DISCUSSION AND CONCLUSIONS

These calculations predict regions of negative radiation force on a solid elastic sphere and an elastic spherical shell illuminated by an ideal acoustic Bessel beam. Though the absorption of acoustic energy in the water and in the sphere has been neglected, ray-based calculations by Lee and Shung⁴ for a fluid sphere illuminated by a focused beam suggest that there are situations where absorption may make the radiation force more negative. In the case of the PMMA sphere considered here, Hefner and Marston¹⁸ observed that absorption by the sphere only weakly modifies the scattering in the ka region of interest. The case of the evacuated aluminum shell may be of greater importance, however, because negative Y_p are predicted to exist for β as small as 45° . The scattering into the back hemisphere is suppressed at situations giving a negative Y_p . A systematic search for regions of negative Y_p for a wide range of material and shell parameters was beyond the scope of the present investigation. The influence of gas content within a shell (relevant to the response of contrast agents¹⁹) has not been considered. Nevertheless, the examples given here are sufficient to show that negative radiation forces for spheres illuminated by Bessel beams are not restricted to fluid sphere examples⁵ and that there is a correlation with reduced back-hemisphere scattering.

Inspection of the structure in Figs. 1, 3, and 5 shows that Y_p is affected by resonances of sphere elastic modes. Consequently, regions of negative force for these elastic spheres are not likely to be describable with models based on re-

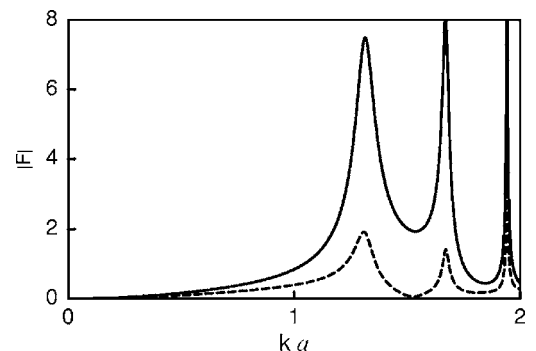


FIG. 5. Backscattering amplitude $|F|$ with $\theta=180^\circ$ of the shell considered in Fig. 3. The conditions on β are as in Fig. 4. The backscattering is small for the dashed curve ($\beta=45^\circ$) in the region of negative Y_p in Fig. 3.

fracted rays. Refracted-ray models have been used to explain negative forces of focused beams on dielectric spheres (in optics^{7,20}) and on fluid spheres (in acoustics²⁻⁴).

ACKNOWLEDGMENTS

This research was supported in part by NASA. The author is grateful to Dr. David B. Thiessen for helpful comments.

- ¹J. R. Wu, "Acoustical tweezers," *J. Acoust. Soc. Am.* **89**, 2140–2143 (1991).
- ²J. W. Lee, K. L. Ha, and K. K. Shung, "A theoretical study of the feasibility of acoustical tweezers: Ray acoustics approach," *J. Acoust. Soc. Am.* **117**, 3273–3280 (2005).
- ³J. W. Lee and K. K. Shung, "Radiation forces exerted on arbitrarily located sphere by acoustic tweezer," *J. Acoust. Soc. Am.* **120**, 1084–1094 (2006).
- ⁴J. W. Lee and K. K. Shung, "Effect of ultrasonic attenuation on the feasibility of acoustical tweezers," *Ultrasound Med. Biol.* **32**, 1575–1583 (2006).
- ⁵P. 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).
- ⁶P. L. Marston, "Axial radiation force of a Bessel beam on a sphere, direction reversal of the force, and solid sphere examples," *J. Acoust. Soc. Am.* **121**, 3109(A) (2006).
- ⁷A. Ashkin, "Forces of a single beam gradient laser trap on a dielectric sphere in the ray optics regime," *Biophys. J.* **61**, 569–582 (1992).
- ⁸T. A. Nieminen, G. Knöner, N. R. Heckenberg, and H. Rubinsztein-Dunlop, "Physics of optical tweezers," *Methods Cell Biol.* **82**, 207–236

(2007).

- ⁹P. L. Marston and D. B. Thiessen, "Manipulation of fluid objects with acoustic radiation pressure," *Ann. N.Y. Acad. Sci.* **1027**, 414–434 (2004).
- ¹⁰J. Durnin, "Exact solutions for nondiffracting beams. I. The scalar theory," *J. Opt. Soc. Am. A* **4**, 651–654 (1987).
- ¹¹J. Durnin, J. J. Miceli, Jr., and J. H. Eberly, "Diffraction-free beams," *Phys. Rev. Lett.* **58**, 1499–1501 (1987).
- ¹²P. L. Marston, "Scattering of a Bessel beam by a sphere," *J. Acoust. Soc. Am.* **121**, 753–758 (2007).
- ¹³P. L. Marston, "Acoustic beam scattering and excitation of sphere resonance: Bessel beam example," *J. Acoust. Soc. Am.* **122**, 247–252 (2007).
- ¹⁴T. Hasegawa and K. Yosioka, "Acoustic radiation force on fused silica spheres, and intensity determination," *J. Acoust. Soc. Am.* **58**, 581–585 (1975).
- ¹⁵E. Leung, N. Jacobi, and T. Wang, "Acoustic radiation force on a rigid sphere in a resonance chamber," *J. Acoust. Soc. Am.* **70**, 1762–1767 (1981).
- ¹⁶G. T. Silva, S. G. Chen, J. F. Greenleaf, and M. Fatemi, "Dynamic ultrasound radiation force in fluids," *Phys. Rev. E* **71**, 056617 (2005).
- ¹⁷P. L. Marston, "Shape oscillation and static deformation of drops and bubbles driven by modulated radiation stresses-theory," *J. Acoust. Soc. Am.* **67**, 15–26 (1980).
- ¹⁸B. T. Hefner and P. L. Marston, "Backscattering enhancements associated with subsonic Rayleigh waves on polymer spheres in water: Observation and modeling for acrylic spheres," *J. Acoust. Soc. Am.* **107**, 1930–1936 (2000).
- ¹⁹J. A. Ketterling, J. Mamou, J. S. Allen III, O. Aristizábal, R. G. Williamson, and D. H. Turnbull, "Excitation of polymer-shelled contrast agents with high-frequency ultrasound," *J. Acoust. Soc. Am.* **121**, EL48–EL53 (2007).
- ²⁰K. Svoboda, and S. M. Block, "Biological applications of optical forces," *Annu. Rev. Biophys. Biomol. Struct.* **23**, 247–285 (1994).

Acoustic diffraction by deformed edges of finite length: Theory and experiment^{a)}

Timothy K. Stanton and Dezhang Chu

Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543-1053

Guy V. Norton

Naval Research Laboratory, Stennis Space Center, Mississippi 39529I

(Received 30 June 2005; revised 16 August 2006; accepted 10 November 2006)

The acoustic diffraction by deformed edges of finite length is described analytically and in the frequency domain through use of an approximate line-integral formulation. The formulation is based on the diffraction per unit length of an infinitely long straight edge, which inherently limits the accuracy of the approach. The line integral is written in terms of the diffraction by a generalized edge, in that the “edge” can be a single edge or multiple closely spaced edges. Predictions based on an exact solution to the impenetrable infinite knife edge are used to estimate diffraction by the edge of a thin disk and compared with calculations based on the T-matrix approach. Predictions are then made for the more complex geometry involving an impenetrable thick disk. These latter predictions are based on an approximate formula for double-edge diffraction [Chu *et al.*, *J. Acoust. Soc. Am.* **122**, 3177 (2007)] and are compared with laboratory data involving individual elastic (aluminum) disks spanning a range of diameters and submerged in water. The results of this study show this approximate line-integral approach to be versatile and applicable over a range of conditions.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2405126]

PACS number(s): 43.20.El, 43.30.Ft [AJMD]

Pages: 3167–3176

I. INTRODUCTION

The subject of acoustic diffraction by finite deformed edges has been studied in previous investigations through analytical, numerical, and experimental methods. While all motivations have included advancing the fundamental understanding of the acoustic scattering processes, the applications were diverse and include understanding acoustic diffraction around noise barriers, seamounts, and seashells. Given the complexity and possible limitations of exact solutions [such as being constrained to certain canonical shapes (Bowman *et al.*, 1987)], approaches are generally approximate and may be based, in part, on formally exact solutions to the wave equation for the infinitely long straight edge. For example, in the work of Medwin and colleagues, the Biot-Tolstoy solution was truncated in the time domain in order to estimate the acoustic diffraction by finite straight wedges (Medwin, 1981; Medwin *et al.*, 1984). More recently, Svensson *et al.* (1999) incorporated the solution in a line integral to develop an approximate analytical time-domain formulation describing the diffraction by arbitrarily deformed edges of finite length. Also, Menounou and colleagues developed an approximate line-integral approach describing diffraction by edges and wedges of half-planes of arbitrary directivity functions (specific to the type of edge) and incident signal for both straight (Menounou *et al.*, 2000) and jagged (Menounou and You, 2004) edges. One extension of the Medwin concept was the

construction of a deformed edge through an assemblage of wedges (Keiffer *et al.*, 1994). Given the inherent inaccuracies of the approximate approaches, others have approached the problem through numerical methods (Kristensson and Waterman, 1982; Norton *et al.*, 1993). In addition to the above cases involving impenetrable bodies, other studies have investigated the interaction of sound with elastic bodies that possess edges (Hefner, 2000; Hefner and Marston, 2001; 2002).

Although there has been a growing body of work involving diffraction by deformed finite-length edges, each work has been limited to either a particular geometry (i.e., shape of edge) or a particular infinite-edge solution as a basis, or both. Furthermore, there are relatively few comparisons of predictions of diffraction by deformed finite-length edges with numerical and experimental data. In this paper a general approximate analytical formula is derived to describe diffraction by deformed, finite-length edges (or multiple closely spaced edges). The formula is applied to the case of disks where the acoustic diffraction by the leading edge of various disks is studied theoretically, numerically, and experimentally. Disks of infinitesimal thickness (knife-edge) and finite thickness (double edge) are investigated for various diameters and orientations. As described below, the formulation is an advancement in terms of generality. Also, the numerical and experimental data are not only useful in comparisons with the theoretical predictions, but also contribute to the limited base of data associated with deformed finite-length edges.

^{a)}Publication of this paper was delayed by request of author so that it would appear in the same issue with “Higher-order acoustic diffraction by edges of finite thickness” by Dezhang Chu, Timothy K. Stanton, and Allan D. Pierce.

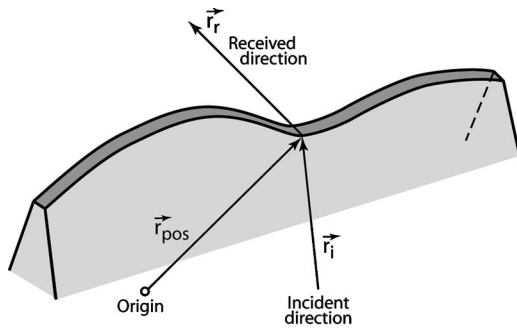


FIG. 1. Deformed truncated wedge of finite length.

The approximate analytical expression describing the diffraction by the deformed edge is derived using a formulation based on the case of infinitely long edges, but employing a line-integral along the length of the edge to account in an approximate manner for deformations of the edge (Fig. 1). Although this particular line-integral method is adapted from the deformed cylinder formulation in Stanton (1989), it is conceptually similar to other approaches such as Svensson *et al.* (1999) and Menounou *et al.* (2000). In fact, line-integral approaches have been used for more than one century and a brief review with comparisons between various approaches is presented in Menounou *et al.* (2000). Although similar conceptually, there are differences, however, involving generality and geometry. For example, the Svensson *et al.* formulas involve use of a particular infinite-edge solution and apply it to predict multiple-order diffraction between opposite edges of a disk. The Menounou *et al.* formulas involve an arbitrary infinite-edge solution and apply it to edges and wedges of half-planes. The approach described here is limited to diffraction from a single edge (or single group of closely spaced edges), but is written in a general form based on an *arbitrary* infinite-edge solution. Furthermore, it is written for an arbitrarily deformed edge of finite length. Note that although the Medwin *et al.* approach involved finite length edges, it incorporated one particular infinite edge formula (Biot-Tolstoy), while the approach in this paper is more general as it involves an arbitrary infinite edge formula. Since this new formulation is based on the diffraction by an infinitely long edge, it is generally applicable for conditions under which deformations in the edge are slowly varying and when the edge is near normal to the directions of the incident and diffracted signals (as in the Svensson *et al.* and Menounou *et al.* approaches).

As mentioned above, the general formula is evaluated for two cases of increasing complexity, each involving impenetrable disks. The first case involves the ideal geometry of the thin (knife-edge) disk which is based on an exact solution for the infinitely long knife edge presented in Morse and Ingard (1968). The predictions are then compared with calculations using the T-matrix approach (Kristensson and Waterman, 1982; Norton *et al.*, 1993) for two diameters and a wide range of orientation angles. This is a relatively pure test of the deformed edge formulation as the only approximation in the comparison involves the fundamental approximation of the line integral being based, in part, on an infinite edge theory. The more complex case is then studied for disks

of finite thickness. In this case, double-edge diffraction must be taken into account. The line-integral formula incorporates a double-edge diffraction formula from Chu *et al.* (2007), which is based on an extension of a formula in Pierce (1974) involving diffraction by an impenetrable infinitely long double edge. The results of these latter predictions are compared with laboratory measurements of broadband diffraction by the leading double edge of individual elastic disks (aluminum, submerged in water) of various diameters and over a wide range of orientation angles. In this latter case, the theory has three approximations. In addition to its inherent approximation of being based on an infinite edge theory, the infinite edge theory only approximately accounts for higher-order diffraction from the double edge. Also, the theory does not account for elastic effects.

This paper is organized as follows: In Sec. II, a general formula is derived describing the diffraction by finite-length deformed edges. The line-integral component of the formula is evaluated for the simple cases of a straight finite-length edge and a disk. In addition, the diffracted field by the leading edge is written explicitly for the case of a thin, impenetrable disk using the exact formula describing diffraction by an infinitely long knife-edge from Morse and Ingard (1968). In Sec. III, numerical calculations are described that are based on the T-matrix approach and are used to predict diffraction by a thin, impenetrable disk. In Sec. IV, a laboratory experiment is described in which the broadband diffraction by the leading double edge of a variety of individual thick, elastic disks submerged in water is measured. In Sec. V, results of the numerical calculations, laboratory experiment, and modeling using the line integral are presented and compared with each other. In Secs. VI and VII, the results are discussed and interpreted, and conclusions are made about the range of validity of the line-integral formulation.

II. THEORY

A. General formulation

The general approach toward describing the diffraction by deformed finite-length edges is based on the line-integral method used in Stanton (1989) to describe the scattering by deformed finite-length cylinders. In that case, the scattering per unit length of an infinitely long cylinder was incorporated into an integral along the length of the deformed cylinder. The integral took into account phase shifts due to deviations of the cylinder from a straight line. In a similar fashion, the diffraction per unit length from an edge is estimated and used to estimate the diffraction from deformed edges. This can apply to a single edge or assemblage of edges such as the double-edge case to be addressed later in this paper.

The line-integral approach begins by examining the radiated pressure from a line source. From Chap. 21 of Skudrzyk (1971), the differential pressure radiated from an infinitesimal section of an infinitely long line is

$$dp_{ls} = \beta q \frac{e^{ikr_r}}{r_r} dz, \quad (1)$$

where

$$\beta \equiv -\frac{ik\rho c}{4\pi}, \quad (2)$$

q is the volume flow per unit length of the line source, k is the acoustic wave number ($=2\pi/\lambda$, where λ is the acoustic wavelength), r_r is the distance between the receiver and an arbitrary point along the line, dz is the differential length of the line which coincides with the z axis, ρ is the density of the surrounding fluid, and c is the speed of sound of the fluid. The total pressure radiated by the line is calculated by integrating along its entire length,

$$p_{ls} = \beta \int_{-\infty}^{\infty} q(z) \frac{e^{ikr_r}}{r_r} dz, \quad (3)$$

where q is assumed to vary as a function of z .

At this point, the radiation from a line source has been formulated in terms of the volume flow per unit length. In this deformed-edge approximation, the radiation will now be treated as the reradiation due to the diffraction by a single, infinitely long straight edge or assemblage of closely spaced, infinitely long straight edges. The diffraction is bistatic and the source and receiver are in the same plane whose normal is the edge, which is still straight at this point in the derivation.

The deformed-edge formulation depends on whether the infinite edge formulation is based on use of a plane-wave source or a point source. A detailed derivation will first be given for the point-source-based infinite edge formulation. At the end of the derivation, details of the plane-wave-based formulation will be briefly summarized. For a point source a distance r_i away from the edge at position z , the volume flow can be written as

$$q(z) = f^{(\text{PS})}(z) \frac{e^{ikr_i}}{r_i}, \quad (4)$$

where now $f^{(\text{PS})}(z)$ is the component of q due to the (point source) diffraction by the edge and the term $r_i^{-1} \exp(ikr_i)$ accounts for the spreading of the incident wave from the source and its phase shift. Integrating Eq. (3) and assuming that $f^{(\text{PS})}(z)$ varies slowly with z compared with kr_i and kr_r allows the stationary phase approximation to be used. Under this condition, the integral is approximately

$$P_{\text{diff}}^{(\infty, \text{PS})} \approx \beta f^{(\text{PS})}(0) (2\pi)^{1/2} [kr_i^{(0)} r_r^{(0)} (r_i^{(0)} + r_r^{(0)})]^{-1/2} \times e^{ik(r_i^{(0)} + r_r^{(0)})} e^{i\pi/4}, \quad (5)$$

where $r_i^{(0)}$ and $r_r^{(0)}$ are the shortest distances between the source and receiver, respectively, and the edge. As mentioned above, the expression for radiation, p_{ls} , in Eq. (3) has been replaced by the reradiation or diffraction $P_{\text{diff}}^{(\infty, \text{PS})}$ by an infinitely long edge and due to a point source.

From this equation, $f^{(\text{PS})}(0)$ is

$$f^{(\text{PS})}(0) = \beta^{-1} (2\pi)^{-1/2} [kr_i^{(0)} r_r^{(0)} (r_i^{(0)} + r_r^{(0)})]^{1/2} \times e^{-ik(r_i^{(0)} + r_r^{(0)})} e^{-i\pi/4} P_{\text{diff}}^{(\infty, \text{PS})}. \quad (6)$$

In order to estimate the diffracted field from a finite-length deformed edge, or assemblage of closely spaced edges, this expression for $f^{(\text{PS})}(0)$ is used to estimate the volume flow in

an expression similar to Eq. (3). However, q is only known (and approximately, at that) for $z=0$. Assuming that the component of the edge that dominates the diffraction is the one involving normal incidence (e.g., $z=0$ for the straight edge), the integral is approximated by replacing $f^{(\text{PS})}(z)$ with $f^{(\text{PS})}(0)$. This approximation restricts the geometries to those deformed edges with at least one section that involves normal incidence (such as at the midpoints of the leading and trailing edges of a disk). The phase shift $\exp(ikr_i)$ in q is still allowed to vary within the integral as the phase shifts play a significant role in the diffracted signal. Replacing $q(z)$ in Eq. (3) with $f^{(\text{PS})}(0) r_i^{-1} \exp(ikr_i)$, and evaluating $\exp(ikr_i)$ and $\exp(ikr_r)$ in the far field, the diffracted field due to a finite-length deformed edge is, approximately,

$$P_{\text{diff}} \approx P_{\text{diff}}^{(\infty, \text{PS})} (2\pi)^{-1/2} [k(r_i^{(T)} + r_r^{(T)}) / (r_i^{(T)} r_r^{(T)})]^{1/2} e^{-i\pi/4} \times \int_{\mathbf{r}_{\text{pos}}} e^{ikr_{\text{pos}}(\hat{r}_i - \hat{r}_r) \cdot \hat{r}_{\text{pos}}} |d\mathbf{r}_{\text{pos}}|, \quad (7)$$

where \hat{r}_i , \hat{r}_r , and \hat{r}_{pos} are unit vectors (indicated by the “ $\hat{}$ ” symbol) for the incident field, diffracted field toward the direction of the receiver, and position of the edges, respectively. For this finite-length geometry, the outer dimensions are small compared with r_i and r_r so that the amplitude dependence upon r_i and r_r could be taken outside the integral and be replaced by the terms $r_i^{(T)}$ and $r_r^{(T)}$, which correspond to the distance from the source and receiver, respectively, to a fixed point on or near the target (i.e., the origin). The term $P_{\text{diff}}^{(\infty, \text{PS})}$ is evaluated using these latter distances. Furthermore, in this far-field approximation, the outer dimensions of the object are smaller than the first Fresnel zone of the transceiver. The term $|d\mathbf{r}_{\text{pos}}|$ replaces dz to allow for an arbitrary deformation of the edge.

Equation (7) is a general approximate expression predicting the far-field diffraction by a deformed finite-length edge or edges. The formula is general both with respect to the facts that the edge is of arbitrary deformation and that the approach is not specific to any particular infinite-edge formulation (i.e., $P_{\text{diff}}^{(\infty, \text{PS})}$), with the exception that the infinite-edge formulation involves use of a point source. For the case in which the infinite-edge formulation involves a plane-wave source, there will be no phase shift ($\exp[ikr_i]$) or spreading (r_i^{-1}) terms in Eq. (4), simplifying that equation to $q(z) = f^{(\text{PW})}(z)$. Here, the phase of the incident plane wave is assumed to be uniformly zero along the straight axis. The only spreading term will involve r_r for the point receiver. Following the same steps in the derivation as for the point-source-based formulation, the far-field diffraction by a deformed finite-length edge or edges is

$$P_{\text{diff}} \approx P_{\text{diff}}^{(\infty, \text{PW})} (k/2\pi r_r^{(T)})^{1/2} e^{-i\pi/4} \int_{\mathbf{r}_{\text{pos}}} e^{ikr_{\text{pos}}(\hat{r}_i - \hat{r}_r) \cdot \hat{r}_{\text{pos}}} \times |d\mathbf{r}_{\text{pos}}|. \quad (8)$$

This plane-wave-based formulation is quite similar to the corresponding point-source-based formulation given in Eq. (7), with the major difference being the dependence of P_{diff} in Eq. (7) on the distance $r_i^{(T)}$ associated with the point source.

The most significant approximation associated with both Eqs. (7) and (8) concerns the fact that the derivations are based on the diffraction from an infinitely long straight edge or edges. As with the deformed cylinder approach, these formulations will only be valid for smooth deformations in the edge and for angles of incidence near normal to the edge. However, for certain objects, this latter restriction is relaxed. As will be seen in later sections, for cases such as the disk, although there are angles of incidence along the perimeter well away from normal, errors associated with those angles do not contribute significantly to the estimated diffracted field since the contributions from the first Fresnel zone, which involve normal and near-normal incidence, dominate the diffraction.

B. Special cases

The general formulation in Eq. (8) is evaluated for two simple geometries, the straight finite-length edge(s) and disk.

1. Straight finite-length edge(s)

In the case of the straight edge(s) of length L , the integral in Eq. (8) is evaluated over the range $-L/2$ to $+L/2$ with the origin at the center of one of the edges. In this case, the approach is valid for scattering geometries where effects due to the ends of the target are not appreciable. Since a finite edge in this formulation is the finite-length portion of a semi-infinite strip, geometries where appreciable diffraction occurs by the edge of the semi-infinite portion of the strip must be avoided for this formulation to apply. The integral in Eq. (8) is equal to the product of the length of the edge(s) and a sinc function which describes the orientation dependence of the diffraction,

$$P_{\text{diff}} = P_{\text{diff}}^{(\infty, \text{PW})} (k/2\pi r_r^{(T)})^{1/2} e^{-i\pi/4} L \frac{\sin \Delta}{\Delta}, \quad (9)$$

where $\Delta = \frac{1}{2}kL(\hat{r}_i - \hat{r}_r) \cdot \hat{r}_{\text{edge}}$ for the bistatic case and $\Delta = kL \cos \gamma$ for the case of backscatter [\hat{r}_{edge} is the unit direction vector of the edge and γ is the angle between the direction of the incident field and the axis of the edge(s)]. This result shows that the diffraction by a finite-length edge(s) is proportional to the length of the edge(s). Furthermore, the orientation dependence of the diffraction is approximated by the sinc function.

2. Disk

In the case of the disk, first the general solution is expressed for the case of backscatter for this geometry, then an explicit expression is written for the case of a thin impenetrable disk. Finally, high-frequency approximations are made for the leading and trailing edges of a disk.

a. General expression for backscatter From the geometry defined in Fig. 2, the vectors in Eq. (8) are given as

$$\mathbf{r}_{\text{pos}} = a \cos \phi \hat{i} + a \sin \phi \hat{j}, \quad (10)$$

$$d\mathbf{r}_{\text{pos}} = (-a \sin \phi \hat{i} + a \cos \phi \hat{j}) d\phi, \quad (11)$$

$$|d\mathbf{r}_{\text{pos}}| = a d\phi, \quad (12)$$

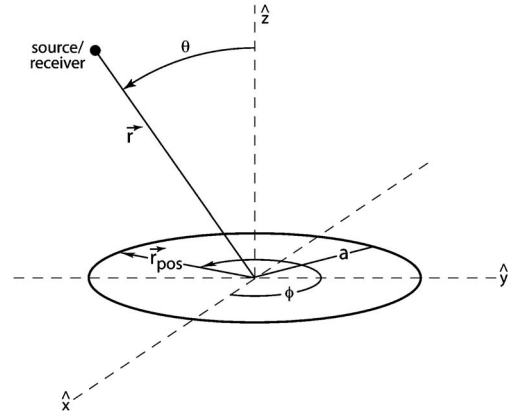


FIG. 2. Diffraction geometry for disk (backscattering). The source/receiver and disk are in the yz and xy planes, respectively.

$$\hat{r}_i = \sin \theta \hat{j} - \cos \theta \hat{z}, \quad (13)$$

$$\hat{r}_r = -\sin \theta \hat{j} + \cos \theta \hat{z} \text{ (backscatter)}, \quad (14)$$

where a is the radius of the disk, \hat{i} , \hat{j} , and \hat{k} are the unit vectors associated with the x , y , and z axes, respectively, and the source/receiver is in the yz plane.

Inserting these quantities into Eq. (8) gives the expression for the diffracted pressure in the back direction,

$$P_{\text{diff}} = P_{\text{diff}}^{(\infty, \text{PW})} (k/2\pi r_r^{(T)})^{1/2} e^{-i\pi/4} a I, \quad (15)$$

where

$$I \equiv \int e^{i2ka \sin \theta \sin \phi} d\phi. \quad (16)$$

The separate contributions to I from the trailing and leading edges are determined by integrating over the ranges $0 \leq \phi \leq \pi$ and $\pi \leq \phi \leq 2\pi$, respectively:

$$I = \pi J_0(2ka \sin \theta) \pm i2s_{0,0}(2ka \sin \theta) \quad (17)$$

where the “+” and “-” signs correspond to the trailing and leading edges, respectively. The terms J_0 and $s_{0,0}$ are the Bessel and Lommel functions of order zero, respectively.

b. Thin impenetrable disk Evaluation of Eq. (15) requires use of an expression for $P_{\text{diff}}^{(\infty, \text{PW})}$ describing the diffraction by an infinitely long edge. Morse and Ingard (1968) present an exact expression for the total field describing the bistatic diffraction by an impenetrable, infinitely long straight knife-edge due to an incident plane-wave source [Eq. (8.4.6) of that work]. For the case of backscatter, their equation simplifies to

$$P_{\text{diff}}^{(\infty, \text{PW})} = A e^{-ikr_r^{(0)}} \{E[\sqrt{2kr_r^{(0)}}] - 1\} + A e^{ikr_r^{(0)} \cos 2\Psi} E[-\sqrt{2kr_r^{(0)}} \sin \Psi], \quad (18)$$

where A is the amplitude of the incident plane wave and Ψ is defined in Morse and Ingard as the angle of the direction of the incident plane wave relative to the normal to the half-plane containing the edge. Here, the incident plane wave, $A \exp(-ikr_r^{(0)})$, was subtracted from the total field to give the scattered field. For an incident wave traveling in a direction

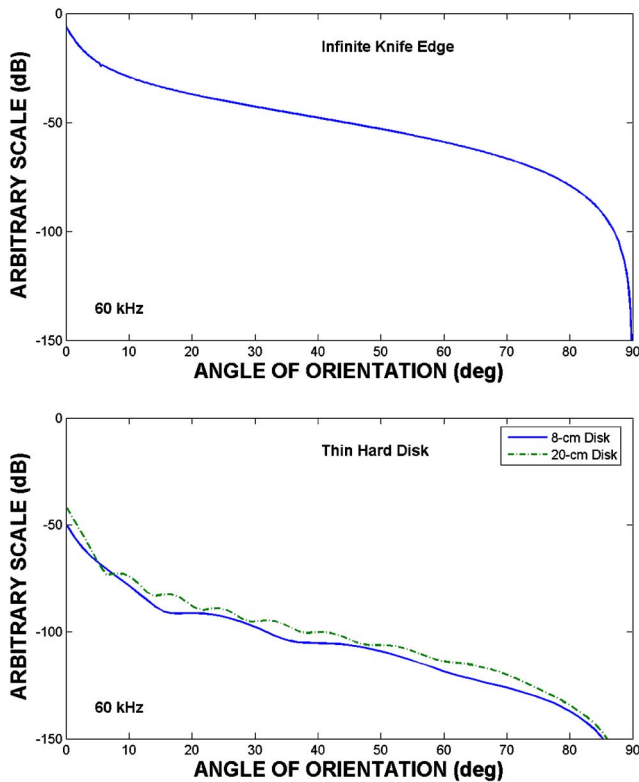


FIG. 3. (Color online) Diffracted echo in backscatter direction from infinitely long, impenetrable straight knife-edge (upper) compared with diffraction in backscatter direction by the leading edge of thin, impenetrable disks (lower). Since the diffracted field spreads differently for the infinitely long edge and the disks, the plots are on arbitrary scales for comparison. The frequency is 60 kHz for all predictions and the diameters of the disks are 8 and 20 cm. The same (exact) expression for $P_{\text{diff}}^{(\infty, \text{PW})}$ [Eq. (18)] was used to produce the upper plot as it was in the lower plots [once integrated through use of Eqs. (15) and (16) to give Eq. (19)], as well as in Fig. 5. The angle $\theta=0^\circ$ is normal incidence to flat surface of disk and half-plane associated with the infinite knife-edge, while $\theta=90^\circ$ corresponds to edge-on incidence.

normal to the plane, $\Psi=0$, and for edge-on incidence, $\Psi = \pi/2$. In the disk geometry defined in Fig. 2 of this paper, where θ is always positive, $\Psi = \theta$ for the leading edge of the disk and $\Psi = -\theta$ for the trailing edge. The function $E[\dots]$ is defined in Eq. (8.4.3) of Morse and Ingard and can be expressed in terms of Fresnel integrals as shown in Eq. (8.4.5) of that work. Note that Eq. (18) is for all $kr_r^{(0)}$ and in the limit of $kr_r^{(0)} \gg 1$, the diffracted field varies as $(kr_r^{(0)})^{-1/2}$.

Inserting Eq. (18) into Eq. (15) gives an approximate explicit expression for the diffraction by the edge of a thin, impenetrable disk,

$$P_{\text{diff}} = A(S_1 + S_2)(k/2\pi r_r^{(T)})^{1/2} e^{-i\pi/4} aI, \quad (19)$$

$$\text{where } S_1 \equiv e^{-ikr_r^{(T)}} \{E[\sqrt{2kr_r^{(T)}}] - 1\}$$

$$\text{and } S_2 \equiv e^{ikr_r^{(T)}} \cos 2\theta E[\pm \sqrt{2kr_r^{(T)}} \sin \theta];$$

where the “+” and “−” signs in S_2 correspond to the trailing and leading edges, respectively. Evaluation of Eq. (19) for the leading edge of two disks shows that the diffraction by the edges is a strong function of orientation of the disk and the dimensionless product, ka (Fig. 3). Also, because of the curvature of the disk, the dependence of the diffraction upon ka and orientation angle θ tends to have an oscillatory component [versus the smoothly varying predictions for the infi-

nite knife-edge from Eq. (18)] because of the varying numbers of Fresnel zones occupying the perimeter of the disk.

c. High-frequency limit to circular disk ($ka \gg 1$). In order to further investigate the diffraction by the edge(s) of a disk, the high-frequency limit is explored. In this case, the frequency is high relative to the radius (i.e., $ka \gg 1$). Using the method of stationary phase and dividing the integral in Eq. (16) into sections corresponding to the leading and trailing edges of the disk (one stationary point per section), the integral becomes

Leading edge: $\pi \leq \phi \leq 2\pi$

$$I = \sqrt{\frac{\pi}{ka \sin \theta}} e^{i\pi/4} e^{-i2ka \sin \theta}, \quad (20)$$

Trailing edge: $0 \leq \phi \leq \pi$

$$I = \sqrt{\frac{\pi}{ka \sin \theta}} e^{-i\pi/4} e^{i2ka \sin \theta}. \quad (21)$$

In each of these cases, the integral is shown to vary inversely with the square root of the product of ka and $\sin \theta$. Note that the factor of $a^{-1/2}$ is offset by the factor of a in the numerator of Eq. (15), resulting in the diffracted field increasing with $a^{1/2}$.

3. First Fresnel zone and effective length

The first Fresnel zone plays an important role in the diffraction by edges. The radius of the first Fresnel zone of a plane-wave source incident upon an infinitely long straight line and received in the backscatter direction by a point receiver is $\sqrt{r_r^{(0)}\lambda}$. Similarly, the radius of the first Fresnel zone of a plane-wave source incident upon a curved finite-length line with a constant radius of curvature a and received in the backscatter direction by a point receiver in the far field of the line and at high frequencies is $\sqrt{a\lambda/2}$ (where the curvature is in the plane containing the direction of the incident field and the line is bent symmetrically toward or away from the source/receiver). By inserting either of those expressions for the radius of the first Fresnel zone into the term L (length) of the diffracted pressure in the backscatter direction due to a straight finite edge at normal incidence [Eq. (9)], the result will be the diffracted pressure in the backscatter direction for those corresponding cases of an infinitely long straight edge or leading or trailing edge of a disk at edge-on incidence (to within a phase factor). Thus, the effective size of the infinitely long straight edge and edge of a disk is the radius of the first Fresnel zone.

III. NUMERICAL CALCULATIONS—THIN IMPENETRABLE DISK

In order to understand the diffraction by a deformed edge under ideal conditions, the T-matrix approach was used to predict the diffraction by thin, impenetrable disks. The T-matrix method is a formally exact numerical solution to the wave equation and is routinely used as a basis for comparison for other approaches. The formulation used is based on the work of Kristensson and Waterman (1982), with details of its implementation given in Norton *et al.* (1993).

The T-matrix calculations were computed in the frequency domain and then converted into the time domain through the use of a fast Fourier transform (FFT). The requirement for sampling in the frequency domain was such that the return from the leading and trailing edge in the time domain is separated by at least one time step. Based on the maximum ka used (64), this resulted in the minimum incident angle, capable of resolving the two returns, of approximately six degrees. This maximum ka was kept the same for both disks. Thus, the maximum frequency and delta frequency were different for each disk.

No special windows were applied to the T-matrix-generated frequency response prior to transforming to the time domain. The rectangular (default) window that results from performing finite Fourier transforms causes the sinc(t) function to be convolved with the true impulse response. Finally, the numerical results were normalized by the magnitude of the image reflection that occurs in the axial geometry.

The highest order of spherical Hankel and Bessel functions needed for convergence was dependent upon ka [Norton, *et al.* (1993)]. The higher the ka the more terms required for convergence. For numerical implementation the following rule was used: The highest order used was equal to two times the value of ka (minimum of ten). The maximum azimuthal index used equaled the value of ka (minimum of five). The numbers used at low ka were more than sufficient for convergence, but in order to insure convergence at high ka these relationships were used. No problems were observed throughout the numerical computations and convergence was obtained for all values of ka . All computations were carried out on a SGI Altix super computer.

IV. LABORATORY EXPERIMENT—THICK ELASTIC DISK

The measurements of acoustic diffraction by the disks were part of a larger effort to understand the scattering by benthic shells. In a laboratory experiment conducted in fall 2002, the scattering by a wide range of metal disks as well as benthic shells was measured. The importance of the diffraction by the outer edge of the shells was identified and comparisons were made between the diffraction by the edge of each shell and the diffraction from the double edge of a disk with outer dimensions comparable to those of the shells (Stanton and Chu, 2004). In this paper, the results from the measurements involving a subset of the disks are reported and analyzed. These disks are constructed of aluminum with a range of diameters (6–20 cm) at a constant thickness of 1.9 mm. Another set of disks in which the thickness was varied at a constant diameter is analyzed in Chu *et al.* (2007) where multiple diffraction between the edges is studied. Details of the experimental setup and measurement procedure are given in Stanton and Chu (2004) and in the references cited in that paper, and are briefly summarized below.

The experiments were conducted in a tank of fresh water using two closely spaced transducers, one as transmitter and the other as a receiver. A broadband chirp (linear frequency modulated) signal spanning the frequency range 40–95 kHz was transmitted. The individual targets were 3.0 m from the transducers and were rotated over the range 0°–360° in 1° increments. The echoes were processed both in the time and frequency domains. In the time domain, the echoes were temporally compressed through a cross-correlation technique resembling matched-filter processing so that echoes from features of the targets could be resolved. The range resolu-

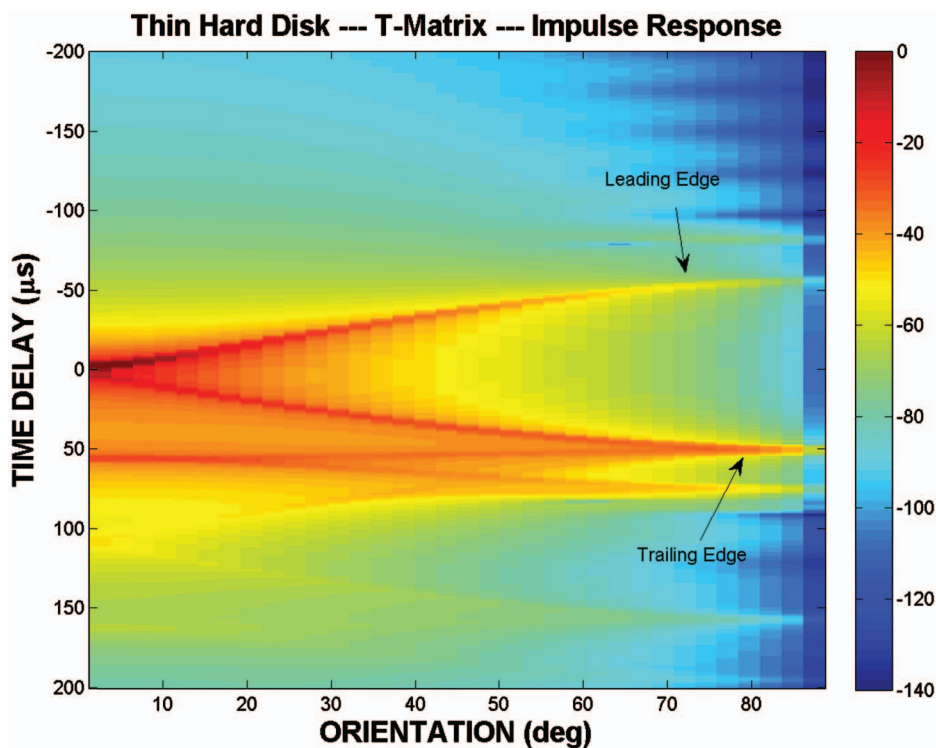


FIG. 4. Impulse response in backscatter direction for 8-cm-diameter thin disk as calculated with the T-matrix method. Normal incidence is 0° and edge-on incidence is 90°. The calculations were over the range 2.5–87.5°. The color scale is in decibels relative to the maximum value of the entire plot. The time delay of 0 μ s corresponds to the center of the disk.

tion achieved through this signal processing approach is about 2 cm, which corresponds to the inverse bandwidth of the system.

V. RESULTS

A. Thin, impenetrable disk

1. General observations of numerical predictions

The impulse response for a 8-cm-diameter thin disk (impenetrable knife-edge) was calculated with the T-matrix method (Fig. 4). The calculations were performed in the backscatter direction and over orientations of the disk spanning the range 2.5–87.5°. (i.e., near normal incidence to near edge-on incidence). The predictions show the diffraction echoes from the edges following the expected sinusoidal pattern throughout the range of orientations. The strength of the backscattered echo from the leading edge is shown to decrease as the orientation angle approaches 90°, which is consistent with the fact that diffraction by a knife edge at edge-on incidence is zero. There is also a strong echo with an arrival time that remains nearly constant at about 55 μ s throughout the entire range of orientations. This time corresponds to a wave traveling along one face of the disk and across the center (i.e., a travel distance equal to the diameter of the disk).

Partial wave target strengths (PWTS) of the leading edge echoes were calculated with the T-matrix method for the 60-kHz component of the impulse response for two thin, impenetrable disks of different diameters (8 and 20 cm). These parameters were chosen for comparison with the laboratory data collected. Here, the PWTS is the target strength that is calculated from echoes from portions of a target, as defined in [Chu and Stanton \(1998\)](#). The leading edge echo is shown to decrease for angles away from normal incidence, and dropping precipitously near edge-on angles (Fig. 5). Also, the echoes are uniformly stronger for the larger disk.

2. Comparison with deformed knife-edge model

The deformed edge formulation is evaluated for the thin impenetrable disk, using the exact solution from [Morse and Ingard \(1968\)](#) for the infinite knife-edge [Eq. (19)], and compared with the T-matrix calculations (Fig. 5). The deformed edge predictions for the leading edge of the disk possess the same general trends as with the T-matrix method, as they show a decrease with echo level for angles away from normal incidence, as well as show an increase in level for the larger disk.

The discrepancies between the two sets of predictions are relatively small for angles of orientation greater than about 20° away from normal incidence. The deformed edge predictions show a more pronounced series of oscillations in the plot of PWTS versus angle than those from the T-matrix predictions.

Also, the calculations from both methods show about a 4-dB difference between the cases of the 8-cm-diameter disk and 20-cm-diameter disk for the higher angles. This is consistent with the limiting form of the deformed edge prediction, where the diffracted field varies as $a^{1/2}$ when there are

many Fresnel zones occupying the edge (i.e., conditions under which the method of stationary phase can be used) as discussed in Sec. II B 2 c.

B. Thick, elastic disk

1. General observations of laboratory data

The temporally compressed echoes as measured in the laboratory contain resolved echoes from both the leading and trailing double edges of the disks (Fig. 6). Since the range resolution of the system is 2 cm, then the individual edges within each double edge pair are not resolvable. In addition to the edge-diffracted echoes, there are strong echoes from the flat surfaces of the disks at normal incidence to the disk face, as well as echoes associated with circumnavigations around the disk (double edge to flat surface to double-edge path) arriving after the trailing edge echoes. There is overlap between the trailing double-edge echoes and other echoes for much of the range of orientation, preventing quantitative interpretation of the trailing edge echoes. The leading double-edge diffracted echoes are easily resolved from the other types of echoes over a wide range of angles, and will there-

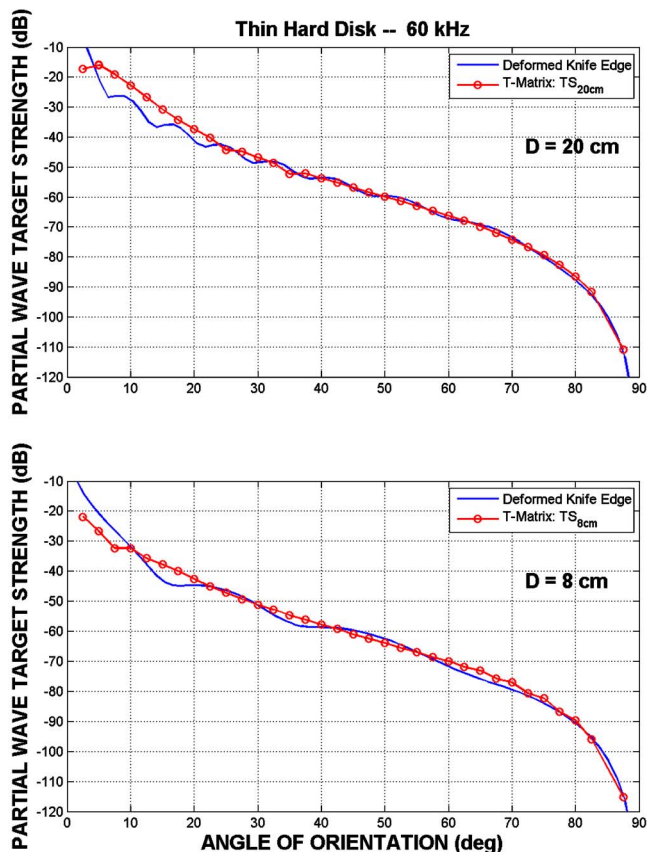


FIG. 5. (Color online) Comparisons between T-matrix and deformed-edge calculations for thin, impenetrable disks of diameters 20 cm (upper) and 8 cm (lower). The partial wave target strength (PWTS) of the diffraction by leading edge only is calculated in each case. As in Fig. 3, Eq. (19) was used for the deformed edge calculations, based on an exact solution to the infinite knife-edge. The leading edge echo was numerically separated from the trailing edge echo in the impulse response time series in the T-matrix calculations, although there was difficulty resolving the two echoes (hence resulting in some contamination) for angles below about 20°. All calculations involved 60 kHz for later comparison with the laboratory data.

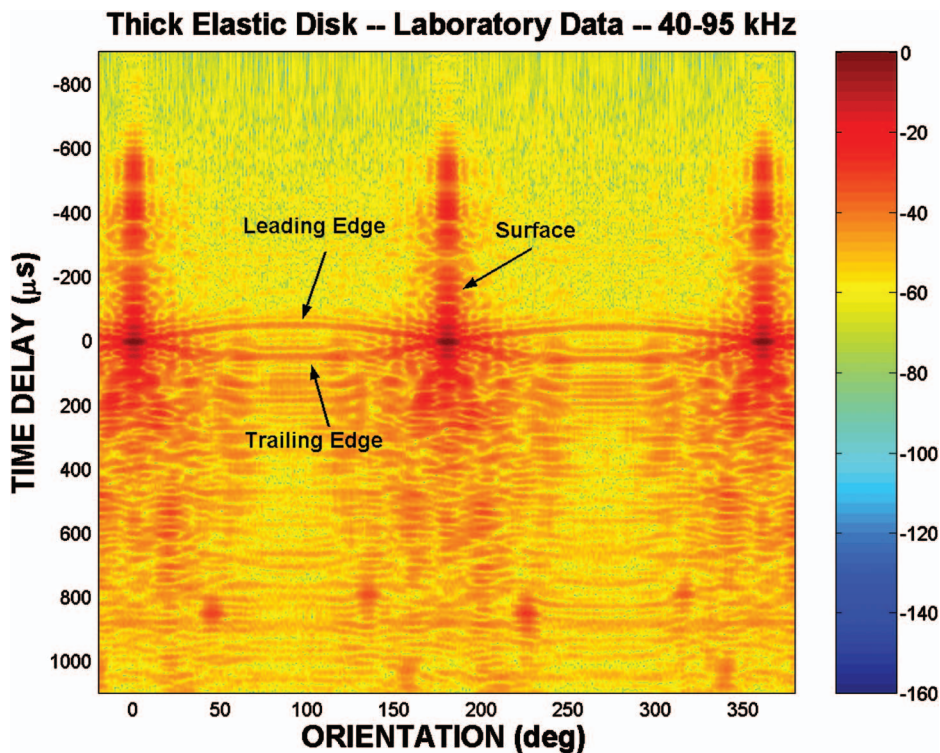


FIG. 6. Temporally compressed echo measured in backscatter direction versus orientation for 8-cm-diameter aluminum disk submerged in water. The disk is 1.9 mm thick. Normal incidence echoes (at $\theta=0^\circ$, 180° , and 360°), leading and trailing double-edge echoes, and circumnavigated echoes are resolved. The circumnavigated waves occur at approximately $110 \mu\text{s}$ after the trailing edge echoes. Other echoes arrive near the circumnavigated echoes and are out of the scope of this analysis. The color scale is in decibels relative to the maximum value of the entire plot. Apparent echoes at normal incidence arriving at negative time delays are actually processing sidelobes from the large zero-time-delay echoes. The abbreviated terminology “leading edge” and “trailing edge” correspond to the more rigorous description “leading double edge” and “trailing double edge.” From Stanton and Chu (2004).

fore be the focus of the analysis. Partial wave target strengths of the leading double-edge echoes were calculated for various spectral components of the signal. The leading double-edge echo was observed to be generally stronger for orientations closer to normal incidence (Fig. 7). Also, the leading double-edge echo tended to increase with increasing diameter of the disk (Fig. 8).

2. Comparisons with models

The measured diffracted echoes from the leading double edge of the elastic disk are generally much larger than those predicted for the thin, impenetrable disk (knife-edge) using either modeling approaches, especially at the larger angles (Figs. 5 and 7). For example, for angles of 40° away from normal incidence, the observed levels for the 8- and 20-cm-diameter-thick disks are several decibels above the corresponding knife-edge predictions. For angles approaching 90° away from normal (i.e., edge-on incidence), the observed (thick-disk) levels are tens of decibels higher, as the predicted knife-edge echoes approach $-\infty\text{dB}$. Given the latter discrepancy, the thickness of the disk needs to be taken into account. Of course, the most rigorous treatment would also require accounting for the elastic properties of the disk.

The scattering by the double edge of an elastic disk is complex. In addition to diffraction by the edges (including higher-order diffractions between the edges), there is the potential for conversion of the signal into elastic waves at the boundary. Given that the disks are relatively thin (i.e., kw is comparable to or smaller than unity, where w is the thickness), then the conversion effects on the face of the edge might be small. Also, since the acoustic impedance of aluminum in water is much greater than unity, then there will be little penetration of the incident signal into the disk. There-

fore, as a first approximation, the scattering by the double edge will be modeled as the pure diffraction by the two edges.

The data were compared with predictions using a deformed double-edge diffraction formulation by incorporating $P_{\text{diff}}^{(\infty, \text{PS})}$ of Chu *et al.* (2007) into Eq. (7) of this paper, and evaluating for the disk geometry. In this case involving a point source, the diffracted signal is given by a modified form of Eq. (15) by replacing $P_{\text{diff}}^{(\infty, \text{PW})}$ with $P_{\text{diff}}^{(\infty, \text{PS})}$ and multiplying the right-hand side by $\sqrt{2}$. The Chu *et al.* formulation involves an impenetrable, infinitely long, straight double edge and is based on the work of Pierce (1974), where the diffraction by a truncated wedge is described. Pierce derived formulas for this case through second-order diffraction. Chu *et al.* improved the accuracy of the second-order diffraction predictions as well as included all higher orders of diffraction. The expression for the term I in Eq. (16) was used to account for the leading double edge only.

The model predictions were based principally on measured dimensions of the disks, although there were parameters empirically determined from the data from one disk to describe multiple diffraction for all of the disks (Chu *et al.* 2007). As mentioned earlier, the transmitting and receiving transducers are not collocated, which results in a deviation from the true backscatter direction of an amount of 6.3° . The predictions took into account this deviation, by evaluating $P_{\text{diff}}^{(\infty, \text{PS})}$ for bistatic scattering. The predictions incorporating bistatic scattering differed by less than 1 dB relative to true backscatter.

There is generally agreement between the predictions and both the trend and much of the structure of the data over most conditions. For example, there is general agreement concerning the trend of decreasing PWTS versus orientation

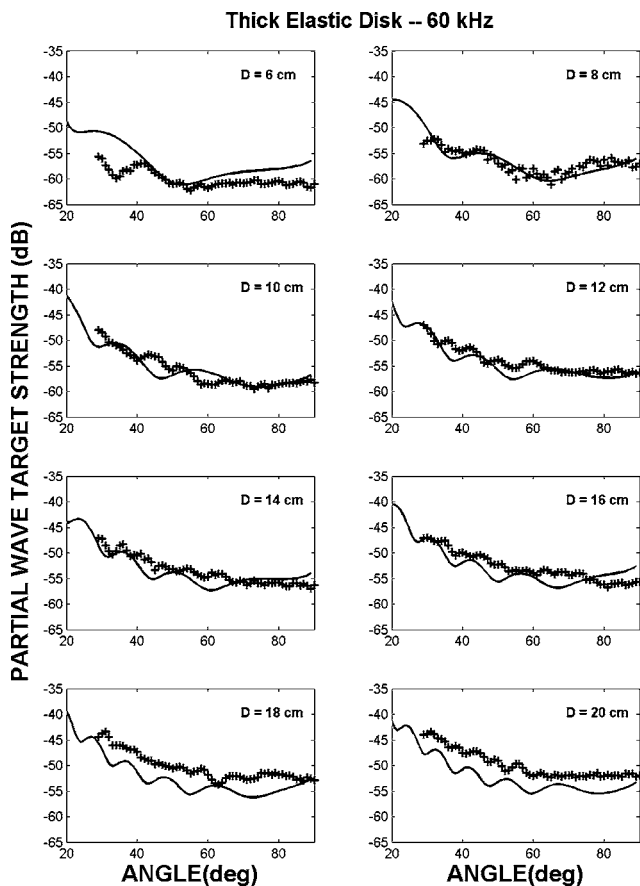


FIG. 7. Partial wave target strength (PWTS) of leading double-edge diffracted 60-kHz echo versus orientation angle for aluminum disks of various diameters. Predictions are given by the solid lines and laboratory data are given by the “+.” The diameters of the disks range from 6 to 20 cm, each with a thickness of 1.9 mm. The 0° angle corresponds to normal incidence to the flat surface of the disks, while 90° corresponds to edge-on incidence. The angle θ is illustrated in Fig. 2. The predictions are based on a formulation from *Chu et al. (2007)* that describes the diffraction by an impenetrable infinitely long, straight double edge. That formula is incorporated into the deformed edge line integral in this paper.

angle of disk (Fig. 7). In that same comparison only some of the observed structure in the data is predicted by the model. For example, all of the structure is predicted for the 8-cm-diameter disk. However, for the 10-cm-diameter disk, the structure in the range of angles 40°–60° is not predicted. For the disks of higher diameter, both predictions and obser-

variations show a more rapid set of oscillations in the pattern of PWTS versus angle, although the respective patterns do not necessarily coincide with each other. There is also very good agreement between predictions and the data concerning the increasing trend of PWTS versus diameter of disk at edge-on incidence (angle of 90° in Fig. 8). At this angle, both the data and predictions increase by about 5 dB as the disk diameter is increased from 6 to 20 cm, which is consistent with the deformed edge prediction that the diffracted field varies as $a^{1/2}$ (Sec. II B 2 c). The predicted levels tend to depart from observations for angles approaching normal incidence to the disk face and larger diameters (angles of 40° and 60° in Fig. 8). There was also reasonable agreement between predictions and data in the study of diffraction as a function of thickness of disk (not shown). Those results are presented in *Chu et al. (2007)* as part of a study of multiple diffraction.

VI. DISCUSSION

The general agreement between the deformed edge predictions and the numerical calculations and laboratory data over much of the range of conditions indicates that the approximate approach of the line-integral method to estimate effects due to deformations of the edge has merit. The most significant approximation in the predictions involved the approximation of basing the formulation, in part, on a solution involving infinitely long edges. Phase shifts using the line-integral approach were assigned to the local field or diffraction per unit length to account for deviations of the edges from a straight line. With this approximation, it is anticipated that the approach is generally only valid for slowly varying deformations and for geometries involving near-normal incidence to the edge. However, the disks involve a wide range of angles of the edge relative to the incident acoustic signal, giving rise to conditions of possibly significant error. For each orientation, although there was a wide range of angles (corresponding to different points along the perimeter of the disk), there was always a section of the disk in which the incident angles were near normal to the tangent of the edge—that is, the section of perimeter closest to the transceiver. This section contains the first Fresnel zone of the acoustic signal which, at these high frequencies, will dominate the echo. Thus, although the predictions for the portions of the edge outside of the first Fresnel zone may contain

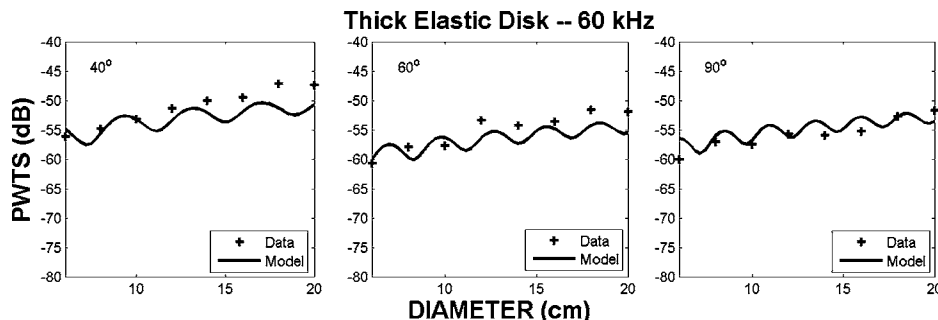


FIG. 8. PWTS of leading double-edge echo versus diameter of aluminum disk at 60 kHz and for three orientation angles. Model predictions are given by the solid lines and laboratory data are given by the “+.” The thickness of the disks is 1.9 mm. The angle θ is illustrated in Fig. 2. The predictions use the same model (impenetrable deformed double edge) as in Fig. 7. The data for the 90° angle follow the trend of varying by $a^{1/2}$, as predicted by the deformed edge model, over the entire range of diameters.

significant error, the errors tend to be canceled out through the alternating signs of the adjacent Fresnel zones. Given the occurrence of this phenomenon, the potentially significant errors in the approach are not realized with this particular scattering geometry.

Certainly, the purest test of the line-integral formula involved the thin, impenetrable disk. Here, an exact formula was used for the infinite knife-edge and the comparison was made with the T-matrix calculations which are exact. Thus, the principal source of error was in the fundamental approximation of the line-integral formula where the local diffraction per unit length was based on the infinite edge formulation. The comparisons showed that the line-integral approach was to within about 2 dB of the T-matrix calculations for the angles of incidence greater than 20° from normal incidence. There were greater discrepancies for smaller angles. These latter differences can be attributed, at least in part, to the fact that it was increasingly more difficult at the small angles to numerically resolve the echoes from the leading and trailing edges in the T-matrix calculations.

A much less pure, but interesting, test involved the comparison of the formulation with the laboratory data involving thick, elastic disks. Here, in addition to the fundamental limitation of the formulation discussed above, there were sources of error involving the differences between diffraction by an elastic body and an impenetrable one, and the fact that the infinite double-edge formula used in the integral was approximate. With these additional sources of error, the discrepancies were generally up to about 5 dB. Quantification of the error associated with elastic effects are out of the scope of this present study, although the error associated with the double-edge formula is discussed in *Chu et al. (2007)*.

VII. CONCLUSIONS

A general approximate formula was derived to predict diffraction from deformed edges of finite length using a line-integral approach. The formula is general, as it is written for arbitrary deformations and is based on an arbitrary infinite edge formula. There was reasonable agreement between predictions using this formula and exact numerical predictions for a thin, impenetrable disk and laboratory data involving thick, elastic disks. For the conditions under which the sources of error in the analysis were at a minimum, i.e., angles greater than about 20° from normal incidence of the thin, impenetrable disks, predictions using the formulation were to within about 2 dB of the exact solution. Deviations were expectedly greater (about 5 dB) with the thick, elastic disk as the infinite double-edge formula was approximate and elastic effects were not taken into account. Although there is potential for significant error for any finite-length edge geometry due to the fact that the formula is based, in part, on one associated with infinitely long edges, significant error was not realized for the scattering geometries in this paper due to the dominance of the first Fresnel zone and cancellation of the errors associated with higher-order

Fresnel zones. Given the wide range of conditions under which this line-integral approach appears to be valid, there is great utility in the approach.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous reviewers for their thoughtful advice that led to significant improvements of this paper. The authors are also grateful to Shirley Barkley, Jayne Doucette, and Craig Johnson, all from the Woods Hole Oceanographic Institution (WHOI), Woods Hole, MA, for preparation of the manuscript to this paper, drawing certain figures, and for construction of the disks, respectively. This research was supported by the U.S. Office of Naval Research (Grant No. N00014-02-0095), WHOI, and by a grant of computer time at the U.S. Department of Defense High Performance Computing Shared Resource Center (Naval Research Laboratory, Washington, DC).

- Bowman, J. J., Senior, T. B. A., and Uslenghi, P. L. E. (1987). *Electromagnetic and Acoustic Scattering by Simple Shapes* (Hemisphere, New York).
- Chu, D., and Stanton, T. K. (1998). "Application of pulse compression techniques to broadband acoustic scattering by live individual zooplankton," *J. Acoust. Soc. Am.* **104**, 39–55.
- Chu, D., Stanton, T. K., and Pierce, A. D. (2007). "Higher-order acoustic diffraction by edges of finite thickness," *J. Acoust. Soc. Am.* **122**, 3177.
- Hefner, B. T. (2000). "Acoustic backscattering enhancements for circular elastic plates and acrylic targets, the application of acoustic holography to the study of scattering from planar elastic objects, and other research on the radiation of sound," Ph.D. thesis, Washington State University.
- Hefner, B. T., and Marston, P. L. (2001). "Backscattering enhancements associated with the excitation of symmetric Lamb waves on a circular plate: Direct and holographic observations," *ARLO* **2**, 55–60.
- Hefner, B. T., and Marston, P. L. (2002). "Backscattering enhancements associated with antisymmetric Lamb waves confined to the edge of a circular plate: Direct and holographic observations," *ARLO* **3**, 101–106.
- Keiffer, R. S., Novarini, J. C., and Norton, G. V. (1994). "The impulse response of an aperture: Numerical calculations within the framework of the wedge assemblage method," *J. Acoust. Soc. Am.* **95**, 3–12.
- Kristensson, G., and Waterman, P. C. (1982). "The *T* matrix for acoustic and electromagnetic scattering by circular disks," *J. Acoust. Soc. Am.* **72**, 1612–1625.
- Medwin, H. (1981). "Shadowing by finite noise barriers," *J. Acoust. Soc. Am.* **69**, 1060–1064.
- Medwin, H., Childs, E., Jordon, E. A., and Spaulding, R. A. Jr. (1984). "Sound scatter and shadowing at a seamount: Hybrid physical solutions in two and three dimensions," *J. Acoust. Soc. Am.* **75**, 1478–1490.
- Menounou, P., and You, J. H. (2004). "Experimental study of the diffracted sound field around jagged edge noise barriers," *J. Acoust. Soc. Am.* **116**, 2843–2854.
- Menounou, P., Busch-Vishniac, I. J., and Blackstock, D. T. (2000). "Directive line source model: A new model for sound diffraction by half planes and wedges," *J. Acoust. Soc. Am.* **107**, 2973–2986.
- Morse, P. M., and Ingard, K. U. (1968). *Theoretical Acoustics* (Princeton University Press, Princeton, NJ).
- Norton, G. V., Novarini, J. C., and Keiffer, R. S. (1993). "An evaluation of the Kirchhoff approximation in predicting the axial impulse response of hard and soft disks," *J. Acoust. Soc. Am.* **93**, 3049–3056.
- Pierce, A. D. (1974). "Diffraction of sound around corners and over wide barriers," *J. Acoust. Soc. Am.* **55**, 941–955.
- Skudrzyk, E. (1971). *The Foundations of Acoustics* (Springer, New York).
- Stanton, T. K. (1989). "Sound scattering by cylinders of finite length III. Deformed cylinders," *J. Acoust. Soc. Am.* **86**, 691–705.
- Stanton, T. K., and Chu, D. (2004). "On the acoustic diffraction by the edges of benthic shells," *J. Acoust. Soc. Am.* **116**, 239–244.
- Svensson, U. P., Fred, R. I., and Vanderkooy, J. (1999). "An analytic secondary source model of edge diffraction impulse responses," *J. Acoust. Soc. Am.* **106**, 2331–2344.

Higher-order acoustic diffraction by edges of finite thickness

Dezhang Chu^{a)} and Timothy K. Stanton

Department of Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543-1053

Allan D. Pierce

Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts 02215

(Received 6 September 2005; revised 9 April 2007; accepted 9 August 2007)

A cw solution of acoustic diffraction by a three-sided semi-infinite barrier or a double edge, where the width of the midplanar segment is finite and cannot be ignored, involving all orders of diffraction is presented. The solution is an extension of the asymptotic formulas for the double-edge second-order diffraction via amplitude and phase matching given by Pierce [A. D. Pierce, *J. Acoust. Soc. Am.* **55**, 943–955 (1974)]. The model accounts for all orders of diffraction and is valid for all kw , where k is the acoustic wave number and w is the width of the midplanar segment and reduces to the solution of diffraction by a single knife edge as $w \rightarrow 0$. The theory is incorporated into the deformed edge solution [Stanton *et al.*, *J. Acoust. Soc. Am.* **122**, 3167 (2007)] to model the diffraction by a disk of finite thickness, and is compared with laboratory experiments of backscattering by elastic disks of various thicknesses and by a hard strip. It is shown that the model describes the edge diffraction reasonably well in predicting the diffraction as a function of scattering angle, edge thickness, and frequency. © 2007 Acoustical Society of America.
[DOI: 10.1121/1.2783001]

PACS number(s): 43.20.El, 43.30.Ft [AJMD]

Pages: 3177–3194

I. INTRODUCTION

To study the scattering by a seafloor covered by shells, laboratory measurements of free-field backscattering by sea shells and a variety of machined circular disks with different sizes, thicknesses, and material properties were made (Stanton and Chu, 2004). Qualitative similarities were observed between the scattering by the edges of the disks and the edges of the shells (especially the sand dollar) (Stanton and Chu, 2004). A major challenge remains for quantitatively describing acoustic diffraction by edges that occur in nature, which are deformed, of finite length, and are composed of a complex material such as one with elastic properties.

One special case of deformed edges of finite length concerning the problem of diffraction by circular disks has been studied intensively by many investigators. Sleator (1969) presented an exact solution based on the modal series solution of an oblate spheroid by letting the aspect ratio approach zero. The T matrix method, a formally exact analytical/numerical approach, has been used to study the acoustic and electromagnetic scattering by circular disks (Kristensson and Waterman, 1982). A number of approximate approaches have also been developed. Keiffer *et al.* (1994), by using the Huygens wavelets approach, presented a discrete wedge assemblage (WA) model that was based on the exact and closed form solution of scattering by an infinitely long straight wedge in the time domain (Biot and Tolstoy, 1957). This method was first proposed by Medwin *et al.* (1982) by using the truncated time series from the solution to an infinitely long straight wedge to describe the diffraction by straight

finite-length wedges. Svensson *et al.* (1999) extended the WA model to a more generalized integral form that can include the second-order diffraction by using the secondary edge sources. It was shown that a second-order diffracted component that circumnavigates the disk could be accurately included in the calculations and was calculated numerically. Recently, Li *et al.* (1998) presented a hybrid-iterative method to compute the scattering by a conducting circular disk. The model combined analytical and numerical approaches and employed a number of techniques including the vector wave eigenfunction expansion, the least-squares method, and the mode matching. A brief review of other approximate methods was also given in that paper including the physical optics (Rahmat-Samii, 1988), the physical theory of diffraction (Ufimtsev, 1962), the geometric theory of diffraction (Keller, 1962), and the method of moments (Duan *et al.*, 1991).

All of the above-mentioned methods were applied to the problem of calculating the diffraction by circular disks with single knife edges [left drawing in Fig. 1(a)], i.e., disks of zero thickness. However, realistic cases may involve edges of finite thickness—that is, a double edge or truncated wedge geometry [middle drawing in Fig. 1(a) and the three-dimensional geometry shown in Fig. 1(b)]. For example, in a recent paper, the scattering by elastic circular disks was studied for the purpose of understanding the underlying scattering mechanisms of benthic shells (Stanton and Chu, 2004). Although the disks were quite thin relative to their diameter, the value of kw , where k is the acoustic wave number and w is the thickness, was comparable to unity. In this scattering region, the thickness cannot be ignored and higher-order diffraction between the two closely spaced parallel edges along the perimeter of the disk must be taken into account [right drawing in Fig. 1(a)].

^{a)}Current address: NOAA-NMFS, Northwest Fisheries Science Center, Seattle, WA 98112. Electronic mail: dchu@whoi.edu

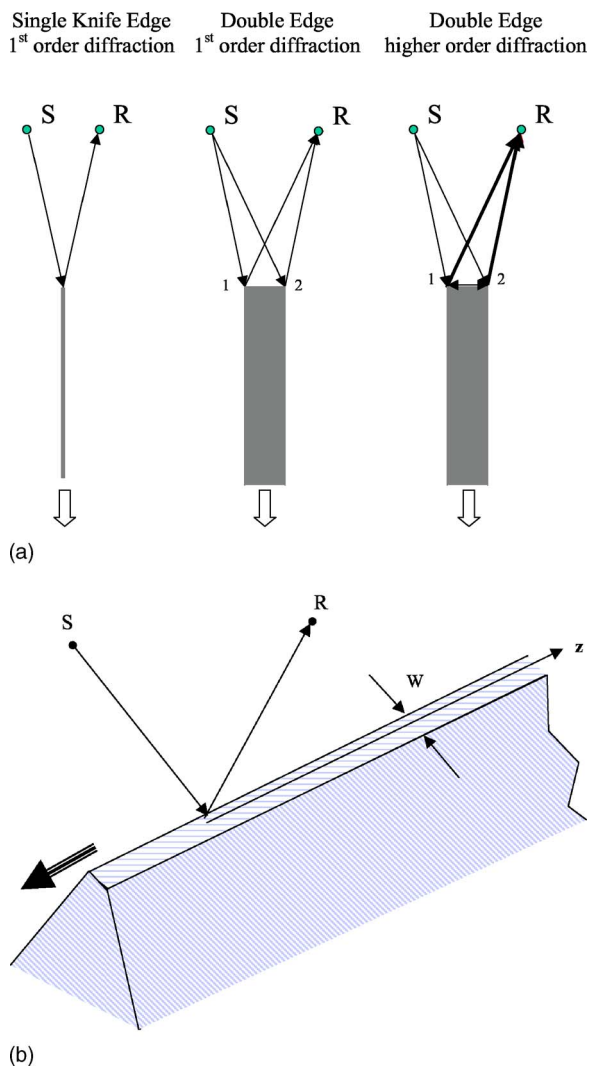


FIG. 1. (Color online) (a) 2D cross-sectional view of diffraction by infinitely long straight truncated wedges of two thicknesses. “S” and “R” indicate the locations of the point source and receiver, respectively. (b) Three-dimensional (3D) view of diffraction by an infinitely long straight truncated wedge.

To demonstrate the necessity of including higher-order diffraction, we compare the experimental data [partial wave target strength (PWTS) of the leading edge, the portion of the disk closer to the transducers, i.e., “A” on the left graph in Fig. 2] with the theoretical predictions (Fig. 2). The two diffraction models are the knife-edge single-diffraction model (solid) corresponding to the case illustrated in Fig. 1(a) (left) and the double-edge, first-order diffraction model (dashed) corresponding to the case also illustrated in Fig. 1(a) (middle). Both models are based on the asymptotic solution for an infinitely long straight wedge or a knife edge (Pierce, 1974) and are incorporated into the deformed wedge formulation (Stanton *et al.*, 2007). Clearly, the single-diffraction model of a knife edge severely underestimates the edge diffraction from scattering angles between 30° and 90° . It is well known that for an infinitely long screen of half space, or a knife edge, the diffraction approaches zero as the scattering angle (backscatter) approaches the edge-on incidence ($\theta=90^\circ$ in Fig. 2). Another natural edge-diffraction model includes a simple summation of the first-order diffrac-

tion by two wedges of right angle as shown as a dashed line in Fig. 2. This double-edge single-diffraction model provides a much better fit to the data, especially for the overall pattern, but it overestimates the overall level of diffraction by a few decibels. To better describe the diffraction by an edge of finite thickness, a more sophisticated model is desirable.

Predicting diffraction by disks of finite thickness faces two major challenges—one is to account for the deformation (circular curve) of the double edge and the other is to account for the multiple diffraction between the two closely spaced edges. The deformation can be accounted for with an approximate line integral approach and has been studied in a separate paper (Stanton *et al.*, 2007). In this paper, higher-order diffraction between two infinitely long edges is studied and formulations are developed. The approach is based on a formulation published by Pierce (1974) where the second-order diffraction is included for cases when $kw \gg 1$. In this paper, Pierce’s approach is extended to include all orders of diffraction, in which a heuristic formula is used to analytically connect the strength of the virtual sources associated with the thickness between the thin (knife edge) and thick (finite thickness) disk solutions so that the solution can be used to estimate diffraction for all thicknesses. The solution, once incorporated into the deformed edge line integral and using empirically determined coefficients, is compared with the laboratory data involving machined elastic disks of various thicknesses as a function of orientation and disk thickness. Furthermore, the model predictions are also compared with the previously published experimental data involving the diffraction by a straight strip of finite width (Medwin *et al.*, 1982). In deriving the high-order diffraction model, since we include only the first arrival (edge diffraction) from the pulse-compressed signal for each ping and the reflection coefficient for an infinite aluminum plane is close to unity, the influence of elasticity is ignored and a proposed higher-order diffraction model is based on the rigid wedge solution.

The paper is organized as follows: Section II defines the problem. Section III briefly reviews the diffraction by an infinitely long straight single wedge and examines the diffracted field across the reflection and shadow boundaries for the two extreme situations $r \rightarrow 0$ and $r \rightarrow \infty$. In Sec. IV, the higher-order diffraction model will be developed. Model-data comparison and the subsequent discussions will be included in Sec. V. Finally, summaries and conclusions are provided in Sec. VI.

II. DEFINITION OF THE DIFFRACTION PROBLEM

The total field associated with the presence of a diffracting object is the sum of the incident (p_{inc}), reflected (p_{ref}), and total diffracted ($p_{dif,tot}$) fields and can be expressed as

$$p_{tot} = p_{inc} + p_{ref} + p_{dif,tot}. \quad (1)$$

The incident and reflected fields can be determined easily based on the scattering geometry using the method of images (Biot and Tolstoy, 1957). For a single edge or wedge, the total diffracted field is solely due to the first-order (single) diffraction by the edge [Fig. 1(a), left]. However, for the diffraction by two parallel infinitely long edges (double

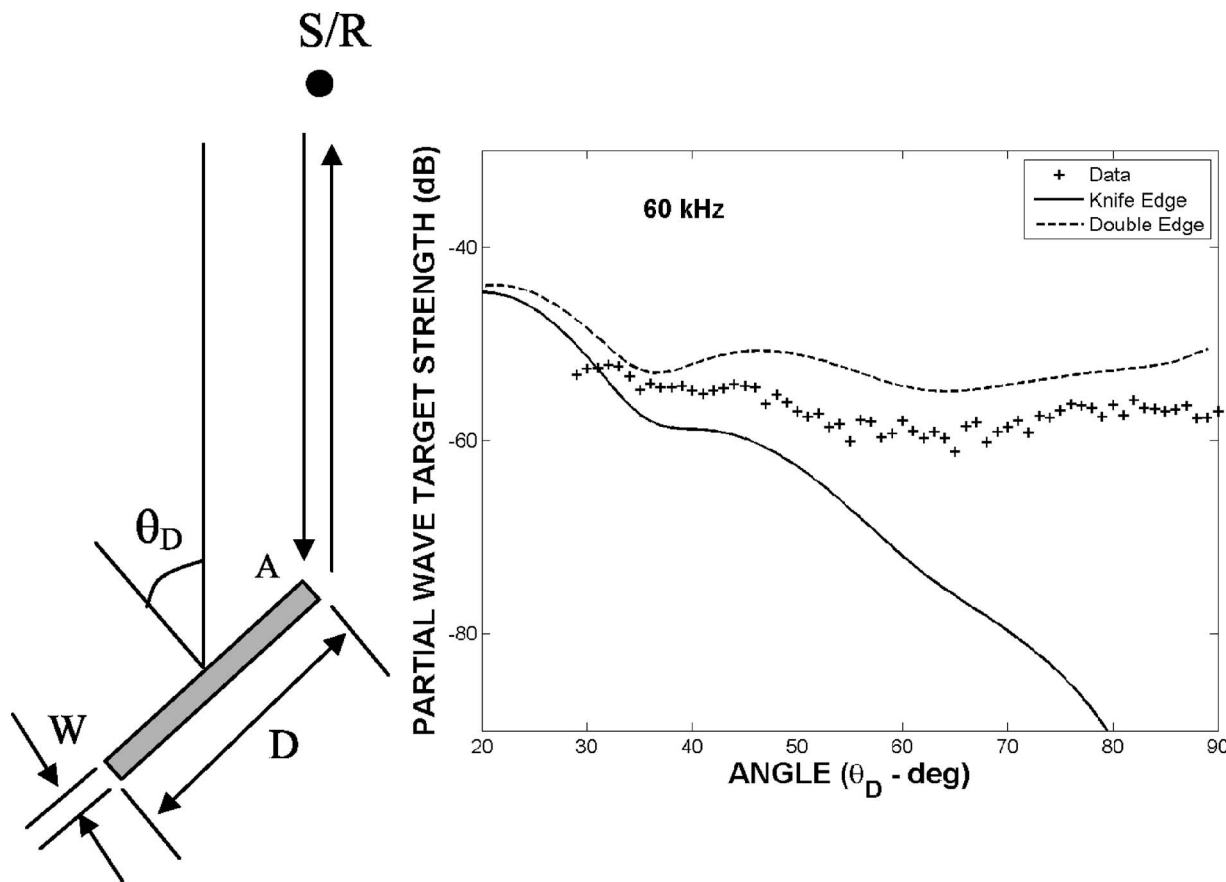


FIG. 2. Comparison between laboratory data and knife-edge diffraction model. The “+” symbol indicates measured diffraction by the leading-edge (A) of an aluminum disk with a diameter of 8 cm (D) and thickness of 1.9 mm (W). The solid curve is the model prediction based on an infinitely long knife edge from Pierce (1974) and with the curvature of the edge accounted for using the method described in Stanton *et al.* (2007). The dashed curve is the sum of the first-order diffracted waves from the two right-angle wedges at A.

edge) or more generally, a truncated wedge [Fig. 1(b)], the total diffracted field is composed of the first-order diffraction from the two edges [Fig. 1(a), middle] and the higher-order diffraction between the edges [Fig. 1(a), right]

$$P_{\text{dif,tot}} = \sum_{n=1}^{\infty} P_{\text{dif},n},$$

where $p_{\text{dif},n}$ represents the n th order diffraction from the double edge. The second-order diffraction $p_{\text{dif},2}$ that corresponds to the ray path $S \rightarrow 1 \rightarrow 2 \rightarrow R$ or $S \rightarrow 2 \rightarrow 1 \rightarrow R$ [Fig. 1(a), right] is also referred to as “double diffraction” by a number of authors (Karp and Keller, 1960; Keller, 1962; Pierce, 1974; Medwin *et al.*, 1982; Svensson *et al.*, 1999). The reason why we use the terminology “second-order diffraction” instead of double diffraction is that double diffraction, when taken literally, could also imply first-order diffraction by edges 1 and 2 of the double edge [Fig. 1(a), middle] rather than second-order diffraction. In the following sections, we will derive an analytical expression for $p_{\text{dif,tot}}$ that includes all orders of diffraction from a double edge, or a truncated wedge.

III. FIRST-ORDER DIFFRACTION

A. Background and Pierce’s solution for first-order diffraction

Diffraction of electromagnetic and acoustic waves by an infinitely long straight wedge (not truncated) has long been studied and can be traced back to Sommerfeld (1954), who, in 1896, first presented the exact solution in the frequency domain due to a plane incident wave for a knife edge (wedge angle of $\theta_w = 2\pi$, where θ_w is defined in Fig. 3) and suggested that the solution for a wedge with an arbitrary wedge angle can be derived based on the same principle. The diffraction by wedges for a cylindrical incident wave was presented by Carslaw (1899, 1920). Later on, MacDonald provided an exact solution, also in the frequency domain, for a rigid wedge due to a cylindrical source (MacDonald 1902) and a point source (MacDonald 1915) in the form of a contour integral. Biot and Tolstoy (1957), using the method of normal coordinates, presented an exact impulse solution in closed form for a rigid wedge due to a point source involving only elementary functions. Applications and developments based on Biot and Tolstoy’s (BT) solution have been reported by many investigators (Medwin *et al.*, 1982; Kinney *et al.*, 1983; Daneshvar and Clay, 1987; Chu, 1989; Clay *et al.*, 1993; Li *et al.*, 1994; Keiffer *et al.*, 1994; Medwin and Clay,

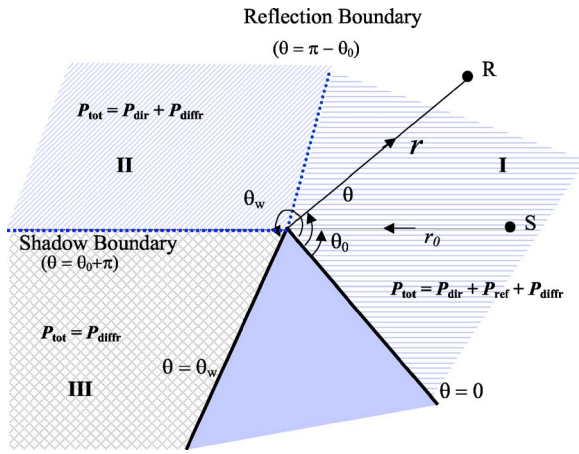


FIG. 3. (Color online) Geometry illustrating different diffraction regions associated with diffraction by an infinitely long straight wedge. 2D cross-sectional view.

1998; Feuillade *et al.*, 2004). At almost the same time when Biot and Tolstoy published their point source solution, in a much less referenced work by Friedlander (1958), an exact two-dimensional (2D) solution due to a line source in the time domain involving only the elementary functions was

also presented. A variety of solutions using other methods and focusing on other types of applications have also been published (Keller *et al.*, 1956; Keller, 1957, 1962; Ufimtsev, 1962; Morse and Ingard, 1968; Pierce, 1974, 1981; Tolstoy and Clay, 1987; Tolstoy, 1989a,b; Davis and Scharstein, 1997; Menounou *et al.*, 2000).

We choose to begin with a solution that originated in Pierce (1974) and was given in a complete form in Pierce (1981). Because of the limited bandwidth of the acoustic system used in our experiment, this frequency domain cw solution is appropriate for our application. In addition, the spectral representation is a convenient analytical form that can be manipulated easily and extended to include all orders of diffraction. The integral form of a spectral representation of the exact solution of the diffracted wave from an infinitely rigid wedge (Fig. 3) due to a point source is (Pierce, 1981, pp. 479–489)

$$p_{\text{dif}} = -\frac{S \sin \nu \pi}{2\theta_w} \sum_{\pm} \int_{-\infty}^{\infty} \frac{e^{ikR(\pi-is)}}{R(\pi-is)} F_s(s, \theta \pm \theta_0) ds, \quad (2a)$$

where S is the source strength, $\nu = \pi/\theta_w$ is the wedge parameter, and

$$F_s(s, \vartheta) = \frac{\cos \nu \pi - \cos \nu \vartheta \cos \nu s}{(\cosh \nu s - 1)^2 + 2(\cosh \nu s - 1)(1 - \cos \nu \vartheta \cos \nu \pi) + (\cos \nu \pi - \cos \nu \vartheta)^2}, \quad (2b)$$

$$\begin{aligned} R(\pi - is) &= (r^2 + r_0^2 - 2rr_0 \cos(\pi - is) + (z - z_0)^2)^{1/2} \\ &= (r^2 + r_0^2 + 2rr_0 \cosh s + (z - z_0)^2)^{1/2}. \end{aligned} \quad (2c)$$

Without loss of generality, we assume $\nu = \pi/\theta_w < 1$ and $0 \leq \theta_0 \leq \pi$ in our analysis throughout the paper.

Although Eqs. (2a)–(2c) were originally derived using a contour integral, they can also be derived from the exact impulse solution given by Biot and Tolstoy (1957), which is based on the method of normal coordinates (see the Appendix). Equations (2a)–(2c) and (A10) are exactly the same. The exact cw solution for wave propagation in a 2D waveguide based on the normal coordinates was provided by Tolstoy and Clay (1987). The integral variable s represents the imaginary part of a complex angle $\xi = \pi - is$ as described in Pierce (1981, p. 489). When ξ is a real number, it represents the angle between the receiver and the source images (Pierce, 1981; Biot and Tolstoy, 1957). For $kR(s)$ much greater than unity, Eq. (2a) reduces to a simpler but approximate form. Following the procedures outlined in Pierce (1981), the approximate solution can be expressed as

$$p_{\text{dif}} = S \frac{e^{ikL}}{L} \{D_+ + D_-\} = S \frac{e^{ikL}}{L} D_{\Sigma}, \quad kR \gg 1, \quad (3)$$

where $D_{\Sigma} \equiv D_+ + D_-$, will be referred to as “diffraction amplitude” throughout the paper and the distance

$$L = [(r + r_0)^2 + (z - z_0)^2]^{1/2} \quad (4)$$

represents the shortest distance connecting the point source, the apex of the wedge, and the point receiver. The diffraction function D_{\pm} is

$$D_{\pm} = \frac{e^{i\pi/4}}{\sqrt{2}} \frac{\sin \nu \pi}{[1 - \cos \nu \pi \cos \nu(\theta \pm \theta_0)]^{1/2}} A_D[\Gamma M_{\nu}(\theta \pm \theta_0)], \quad (5)$$

where

$$\Gamma = \sqrt{\frac{2rr_0}{\lambda L}}, \quad (6a)$$

$$M_{\nu}(\vartheta) = \frac{\cos \nu \pi - \cos \nu \vartheta}{\nu(1 - \cos \nu \pi \cos \nu \vartheta)^{1/2}}, \quad (6b)$$

and $A_D(X)$ is the diffraction integral (Pierce, 1974, 1981),

$$\begin{aligned} A_D(X) &= \frac{1}{\pi\sqrt{2}} \int_{-\infty}^{\infty} \frac{e^{-u^2} du}{\sqrt{\pi/2X - e^{-i\pi/4}u}} \\ &= \text{sign}(X)[f(|X|) - ig(|X|)], \end{aligned} \quad (6c)$$

where the functions $f(X)$ and $g(X)$ are auxiliary Fresnel functions defined in Abramowitz and Stegun (1971, p. 111). For small X , we have

$$f(X) = \frac{1}{2} - \frac{\pi}{4}X^2 + \dots, \quad (7a)$$

$$g(X) = \frac{1}{2} - X + \frac{\pi}{4}X^2 - \dots. \quad (7b)$$

For large X , we have

$$f(X) = \frac{1}{\pi X} - \frac{3}{\pi^3 X^5} + \dots, \quad (8a)$$

$$g(X) = \frac{1}{\pi^2 X^3} - \frac{15}{\pi^4 X^7} + \dots. \quad (8b)$$

For $|X| > 2$, the errors introduced by only keeping the leading term in Eqs. (8a) and (8b) are less than 1%.

If $\Gamma \gg 1$ and the receiver is not too close to either the shadow boundary ($\theta = \pi + \theta_0$ for $\theta_0 \leq \theta_w - \pi$ or $\theta = \theta_0 - \pi$ for $\theta_0 \geq \pi$) or the reflection boundary ($\theta = \pi - \theta_0$ for $\theta_0 \leq \pi$ or $\theta = 2\theta_w - \pi + \theta_0$ for $\theta_0 \geq \pi$) (Fig. 3), an asymptotic solution can be obtained by keeping only the first term of $f(X)$ and ignoring $g(X)$,

$$D_\Sigma = \left(\frac{1}{\pi\Gamma} \right) \frac{e^{i\pi/4}}{\sqrt{2}} \nu \sin \nu\pi \left[\frac{1}{\cos \nu\pi - \cos \nu(\theta + \theta_0)} + \frac{1}{\cos \nu\pi - \cos \nu(\theta - \theta_0)} \right]. \quad (9)$$

The above-presented representation can also be obtained by applying the method of stationary phase directly to Eq. (2a)–(2c). The diffraction factor given by Eq. (9) is the same as that given by Keller (1962) using the geometric theory of diffraction (GTD), and by Morse and Ingard (1968) by directly solving the wave equation.

B. Special limiting cases of first-order diffraction

In Pierce's original work (1974), the second-order diffraction involves only a special geometry in which both the receiver and the edge apex on the far side are in the shadow zone of the source, hence direct insonification and reflection can be completely ignored. To study the higher-order diffraction by edges with a more general geometry, we need to investigate the characteristics of the single diffraction by infinitely long straight wedge for some special cases. When the receiver is on or very close to either the shadow boundary or reflection boundary, i.e., either $\cos \nu(\theta + \theta_0) = \cos \nu\pi$ or $\cos \nu(\theta - \theta_0) = \cos \nu\pi$, then $M_\nu = 0$. Using Eqs. (7a) and (7b), the diffracted wave reduces to

$$p_{\text{dif}} \rightarrow S \frac{e^{ikL}}{2L}. \quad (10)$$

In other words, the diffracted field is approaching one-half of the reflected wave. This is consistent with the results in the frequency domain given by Sandness *et al.* (1983) and in the time domain given by Svensson and Calamia (2006) and Calamia and Svensson (2007). A special case is that when the wedge angle θ_w also approaches π (half space), both the numerator and the denominator of M_ν given by Eq. (6b) tend to zero. However, by applying the L'Hospital's

rule, it is straightforward to show that $M_\nu \rightarrow 0$ as $\theta_w \rightarrow \pi$, and consequently, $A_D \rightarrow 1/2$ is finite but $D_\pm \rightarrow 0$ by Eq. (5). This is exactly what is expected since for $\theta_w \rightarrow \pi$, the diffraction should approach zero.

Next, we examine the continuity of the total field when the observation point crosses the shadow or reflection boundary. Without loss of generality, we assume the source angle $\theta_0 \leq \pi$ and allow the receiver angle θ to vary between 0 and θ_w as shown in Fig. 3. There are three regions.

Region I:

$$p_{\text{tot}} = p_{\text{inc}} + p_{\text{ref}} + p_{\text{dif}}, \quad 0 \leq \theta \leq \pi - \theta_0, \quad (11a)$$

$$M_\nu(\theta - \theta_0) \leq 0, \quad M_\nu(\theta + \theta_0) \leq 0. \quad (11b)$$

Region II:

$$p_{\text{tot}} = p_{\text{inc}} + p_{\text{dif}}, \quad \pi - \theta_0 \leq \theta \leq \pi + \theta_0, \quad (12a)$$

$$M_\nu(\theta - \theta_0) \leq 0, \quad M_\nu(\theta + \theta_0) \geq 0. \quad (12b)$$

Region III (shadow region):

$$p_{\text{tot}} = p_{\text{dif}}, \quad \pi + \theta_0 \leq \theta \leq \theta_w, \quad (13a)$$

$$M_\nu(\theta - \theta_0) \geq 0, \quad M_\nu(\theta + \theta_0) \geq 0. \quad (13b)$$

The total field under two limiting conditions will be analyzed: (1) Both source and receiver are in the far field; and (2) either source or receiver is on the apex of the wedge.

Far field.

(1) Reflection boundary: $\theta \rightarrow \pi - \theta_0$.

(a) $\theta_- \rightarrow \pi - \theta_0$ (Region I):

$$M_\nu(\theta + \theta_0) \rightarrow 0_- \rightarrow D_+ = -\frac{1}{2},$$

$$p_{\text{dif}} = S \frac{e^{ikL}}{L} \left\{ -\frac{1}{2} - \left(\frac{e^{i\pi/4}}{\sqrt{2}} \right) \times \frac{\sin \nu\pi}{\sqrt{1 - \cos \nu\pi \cos \nu(2\theta_0 - \pi)}} \times A_D [\Gamma M_\nu(2\theta_0 - \pi)] \right\}, \quad (14a)$$

$$\begin{aligned} p_{\text{tot}} &= p_{\text{inc}} + p_{\text{ref}} + p_{\text{dif}} \\ &= p_{\text{inc}} + S \frac{e^{ikL}}{L} + S \frac{e^{ikL}}{L} \left\{ -\frac{1}{2} - \left(\frac{e^{i\pi/4}}{\sqrt{2}} \right) \times \frac{\sin \nu\pi}{\sqrt{1 - \cos \nu\pi \cos \nu(2\theta_0 - \pi)}} \times A_D [\Gamma M_\nu(2\theta_0 - \pi)] \right\} \\ &= p_{\text{inc}} + S \frac{e^{ikL}}{2L} - S \frac{e^{ikL}}{L} \left(\frac{e^{i\pi/4}}{\sqrt{2}} \right) \times \frac{\sin \nu\pi}{\sqrt{1 - \cos \nu\pi \cos \nu(2\theta_0 - \pi)}} \times A_D [\Gamma M_\nu(2\theta_0 - \pi)]. \end{aligned} \quad (14b)$$

(b) $\theta_+ \rightarrow \pi - \theta_0$ (Region II):

$$M_\nu(\theta + \theta_0) \rightarrow 0_+ \rightarrow D_+ = \frac{1}{2},$$

$$p_{\text{dif}} = S \frac{e^{ikL}}{L} \left\{ \frac{1}{2} - \left(\frac{e^{i\pi/4}}{\sqrt{2}} \right) \times \frac{\sin \nu\pi}{\sqrt{1 - \cos \nu\pi \cos \nu(2\theta_0 - \pi)}} \times A_D[\Gamma M_\nu(2\theta_0 - \pi)] \right\}, \quad (14c)$$

$$p_{\text{tot}} = p_{\text{inc}} + p_{\text{dif}} = p_{\text{inc}} + S \frac{e^{ikL}}{L} \left\{ \frac{1}{2} - \left(\frac{e^{i\pi/4}}{\sqrt{2}} \right) \times \frac{\sin \nu\pi}{\sqrt{1 - \cos \nu\pi \cos \nu(2\theta_0 - \pi)}} \times A_D[\Gamma M_\nu(2\theta_0 - \pi)] \right\} = p_{\text{inc}} + S \frac{e^{ikL}}{2L} - S \frac{e^{ikL}}{L} \left(\frac{e^{i\pi/4}}{\sqrt{2}} \right) \frac{\sin \nu\pi}{\sqrt{1 - \cos \nu\pi \cos \nu(2\theta_0 - \pi)}} \times A_D[\Gamma M_\nu(2\theta_0 - \pi)]. \quad (14d)$$

Equations (14b) and (14d) are exactly the same, confirming the continuity of the total field across the reflection boundary. The subscripts of θ_\pm and 0_\pm stand for approaching θ and 0 from above (+) and below (-), respectively. It can be seen that when the receiver approaches the reflection boundary from below [Eq. (14a)], part of the diffracted field approaches one-half of the reflected field but changes sign to compensate for the disappearance of the reflected field when the receiver enters Region II [Eq. (14c)].

(2) Shadow boundary: $\theta \rightarrow \pi + \theta_0$.

(a) $\theta_- \rightarrow \pi + \theta_0$ (Region II):

$$M_\nu(\theta - \theta_0) \rightarrow 0_- \rightarrow D_- = -\frac{1}{2},$$

$$p_{\text{dif}} = S \frac{e^{ikL}}{L} \left\{ -\frac{1}{2} - \left(\frac{e^{i\pi/4}}{\sqrt{2}} \right) \times \frac{\sin \nu\pi}{\sqrt{1 - \cos \nu\pi \cos \nu(2\theta_0 + \pi)}} \times A_D[\Gamma M_\nu(2\theta_0 + \pi)] \right\}, \quad (15a)$$

$$p_{\text{tot}} = p_{\text{inc}} + p_{\text{dif}} = S \frac{e^{ikL}}{L} + S \frac{e^{ikL}}{L} \left\{ -\frac{1}{2} - \left(\frac{e^{i\pi/4}}{\sqrt{2}} \right) \times \frac{\sin \nu\pi}{\sqrt{1 - \cos \nu\pi \cos \nu(2\theta_0 + \pi)}} \right\}$$

$$\left. \begin{aligned} & \times A_D[\Gamma M_\nu(2\theta_0 + \pi)] \Big\} \\ & = S \frac{e^{ikL}}{2L} - S \frac{e^{ikL}}{L} \left(\frac{e^{i\pi/4}}{\sqrt{2}} \right) \\ & \quad \times \frac{\sin \nu\pi}{\sqrt{1 - \cos \nu\pi \cos \nu(2\theta_0 + \pi)}} \\ & \quad \times A_D[\Gamma M_\nu(2\theta_0 + \pi)]. \end{aligned} \right\} \quad (15b)$$

(b) $\theta_+ \rightarrow \pi + \theta_0$ (Region III):

$$M_\nu(\theta - \theta_0) \rightarrow 0_+ \rightarrow D_- = \frac{1}{2},$$

$$p_{\text{dif}} = S \frac{e^{ikL}}{L} \left\{ \frac{1}{2} - \left(\frac{e^{i\pi/4}}{\sqrt{2}} \right) \times \frac{\sin \nu\pi}{\sqrt{1 - \cos \nu\pi \cos \nu(2\theta_0 + \pi)}} \times A_D[\Gamma M_\nu(2\theta_0 + \pi)] \right\}, \quad (16a)$$

$$p_{\text{tot}} = p_{\text{dif}} = S \frac{e^{ikL}}{2L} - S \frac{e^{ikL}}{L} \left(\frac{e^{i\pi/4}}{\sqrt{2}} \right) \times \frac{\sin \nu\pi}{\sqrt{1 - \cos \nu\pi \cos \nu(2\theta_0 + \pi)}} A_D[\Gamma M_\nu(2\theta_0 + \pi)]. \quad (16b)$$

Again, we see that the total field is continuous across the shadow boundary.

Receiver on the apex (near field).

The results given by Eq. (10) and those studied previously are valid only for receiver being away from the apex of a wedge. To find the exact diffracted field at the apex, we need to reevaluate the integral given in Eq. (2a)–(2c) directly instead of using the asymptotic solution given by Eqs. (3)–(5) and (6a)–(6c). By setting $r=0$ in Eq. (2a)–(2c), we obtain

$$\begin{aligned} p_{\text{dif}} &= -S \frac{e^{ikR_0}}{R_0} \frac{\sin \nu\pi}{2\theta_w} \int_{-\infty}^{\infty} \sum_{\pm} F_s(s, \theta \pm \theta_0) ds \\ &= -S \frac{e^{ikR_0}}{R_0} \sum_{q=1}^4 \frac{\sin \nu x_q}{2\theta_w} \int_{-\infty}^{\infty} \frac{ds}{\cosh \nu s - \cos \nu x} \\ &= -S \frac{e^{ikR_0}}{R_0} \frac{1}{2\pi} \sum_{q=1}^4 \sin \nu x_q \int_0^{\infty} \frac{du}{\cosh u - \cos \nu x} \\ &= -S \frac{e^{ikR_0}}{R_0} \frac{1}{2\pi} \sum_{q=1}^4 \sin \nu x_q I_q, \end{aligned} \quad (17a)$$

where $R_0 = [r_0^2 + (z - z_0)^2]^{1/2}$. The terms x_1, x_2, x_3 , and x_4 correspond to $\pi + \theta + \theta_0, \pi + \theta - \theta_0, \pi - \theta + \theta_0$, and $\pi - \theta - \theta_0$, respectively. From the first to the second line in Eq. (17a), we have also used the relation (Pierce, 1981, p.489)

TABLE I. Values of $[x_q]$ and diffraction in three diffraction regions shown in Fig. 3 computed using Eqs. (11)–(13), (19), and (20). In obtaining the results, we have assumed $\nu = \pi/\theta_w < 1$ and $0 \leq \theta_0 \leq \pi$ in our computations. Note that since $r=0$, $p_{\text{inc}} = p_{\text{ref}} = S e^{ikR_0}/R_0$, where $R_0 = [r_0^2 + (z - z_0)^2]^{1/2}$.

	Region I	Region II	Region III
Restrictions	$\theta \geq 0$ $\theta + \theta_0 \leq \pi$	$\theta + \theta_0 \geq \pi$ $\theta - \theta_0 \leq \pi$ $\pi + \theta_0 \leq \theta_w$	$\theta \leq \theta_w$
$[x_1]$	$\pi + \theta + \theta_0$	$\pi + \theta + \theta_0$	$\pi + \theta + \theta_0$
$[x_2]$	$\pi + \theta - \theta_0$	$\pi + \theta - \theta_0$	$\pi + \theta - \theta_0$
$[x_3]$	$\pi - \theta + \theta_0$	$\pi - \theta + \theta_0$	$\pi - \theta + \theta_0 + 2\theta_w$
$[x_4]$	$\pi - \theta - \theta_0$	$\pi - \theta - \theta_0 + 2\theta_w$	$\pi - \theta - \theta_0 + 2\theta_w$
$\sum_{q=1}^4 [x_q]$	4π	$4\pi + 2\theta_w$	$4\pi + 4\theta_w$
p_{dif}	$-S \frac{e^{ikR_0}}{R_0} \left(2 - \frac{2\pi}{\theta_w}\right)$	$-S \frac{e^{ikR_0}}{R_0} \left(1 - \frac{2\pi}{\theta_w}\right)$	$S \frac{e^{ikR_0}}{R_0} \left(\frac{2\pi}{\theta_w}\right)$
p_{tot}	$p_{\text{inc}} + p_{\text{ref}} + p_{\text{dif}}$ $= S \frac{e^{ikR_0}}{R_0} \left(\frac{2\pi}{\theta_w}\right)$	$p_{\text{inc}} + p_{\text{dif}}$ $= S \frac{e^{ikR_0}}{R_0} \left(\frac{2\pi}{\theta_w}\right)$	p_{dif} $= S \frac{e^{ikR_0}}{R_0} \left(\frac{2\pi}{\theta_w}\right)$

$$\begin{aligned} \sin \nu \pi \sum_{\pm} F_s(s, \theta \pm \theta_0) &= \sum_{q=1}^4 \frac{\sin \nu x_q}{\cosh \nu s - \cos \nu x} \\ &= \sum_{q=1}^4 \frac{\sin \nu x_q}{\cosh \nu s + \cos(\pi - \nu x_q)}. \end{aligned} \quad (17b)$$

The analytical expression for I_q is tabulated in [Gradshteyn and Ryzhik \(1980\)](#) (3.514.1),

$$I_q = \frac{1}{\sin \nu x_q} (\pi - \nu [x_q]), \quad (18)$$

where the quantity in the square brackets may be expressed as

$$\begin{aligned} [x_q] &= 2\theta_w + x_q, \quad x_q < 0, \\ [x_q] &= x_q, \quad 0 \leq x_q \leq 2\theta_w, \\ [x_q] &= x_q - 2\theta_w, \quad x_q > 2\theta_w. \end{aligned} \quad (19)$$

The above-presented results are obtained by ensuring that the angle $[x_q]$ falls in the range of $0 \leq [x_q] \leq 2\theta_w$, where x_q could be any combination of $\pi \pm \theta \pm \theta_0$. Substituting Eq. (18) into Eq. (17a) leads to

$$p_{\text{dif}} = -S \frac{e^{ikR_0}}{R_0} \sum_{q=1}^4 \frac{1}{2} (1 - [x_q]/\theta_w). \quad (20)$$

The values of $[x_q]$ in the three regions (Fig. 3) are listed in Table I. The total field at the apex is amplified by a factor of $2\pi/\theta_w$, which is consistent with that discussed by [Pierce \(1981, p. 480\)](#).

Receiver position moves from the apex to the far field.

For a receiver position that varies continuously between the apex and the far field, to our knowledge, there are no closed-form analytical solutions for a cw with satisfactory accuracy available for a general case of an arbitrary wedge angle. We therefore need to evaluate the integral, Eq.

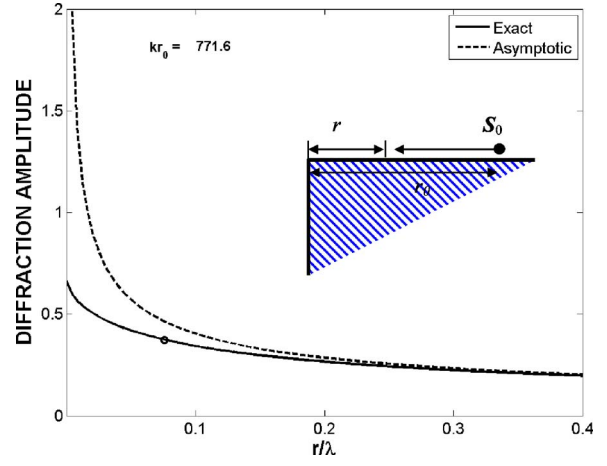


FIG. 4. (Color online) Comparison between the exact solution [Eq. (2a)–(2c)] normalized by $S e^{ikL}/L$ and the diffraction amplitude defined in Eq. (5) and based on an asymptotic expansion of D_Σ , i.e., Eq. (9), as a function of range to the apex. The circle on the solid curve appearing at $r/\lambda \approx 0.08$ corresponds to the position about 1.9 mm away from the apex, or the thickness of the disk used in generating Fig. 2.

(2a)–(2c), numerically. For the case when the source is located on the surface of the wedge but in the far field region with the receiver location varying along the wedge surface from the wedge apex to the source, Fig. 4 illustrates the comparison between the exact and asymptotic solutions for a right-angle wedge ($\theta_w = 3\pi/2$). The solid curve is based on the exact integral solution, Eq. (2a)–(2c), computed numerically and normalized by $S e^{ikL}/L$ [see Eq. (3)], while the dashed curve is the diffraction amplitude, D_Σ , defined in Eq. (5) but using the asymptotic solution, Eq. (9). At $r/\lambda = 0$, the value from the exact solution is $2\pi/\theta_w = 2/3$. The circle on the solid curve appearing at $r/\lambda \approx 0.08$ corresponds to the position of about 1.9 mm away from the apex, or the thickness of the disk used in generating Fig. 2.

IV. HIGHER-ORDER DIFFRACTION

The diffraction by an infinitesimally thin edge, or a knife edge, can be characterized by first-order diffraction or single diffraction. However, real edges have a finite thickness and there may be higher-order diffraction associated with the boundaries of the “edge” (or, more precisely, truncated wedge) that have appreciable magnitude. In the early 1950s, [Jones \(1953\)](#) proposed an exact solution of the diffraction of electromagnetic waves by a thick semi-infinite plate for both parallel and perpendicular polarizations. The solutions involve contour integrals that cannot be easily evaluated. Keller and his colleagues presented approximate solutions for higher-order diffraction by an aperture on a hard screen, a problem that can be related to the scattering by a hard strip by applying Babinet’s principle, using the method of the GTD when the width of the aperture is not too small ([Keller, 1957, 1962](#); [Karp and Keller, 1960](#)). [Medwin et al. \(1982\)](#) used a discrete Huygens wavelet approach to describe the second-order diffraction by double edges. Tolstoy presented an exact solution to a truncated rigid wedge or a double edge ([1989a,b](#)), but his solution is derived based on a line source. [Pierce \(1974\)](#) used a ray-based “virtual source” approach and

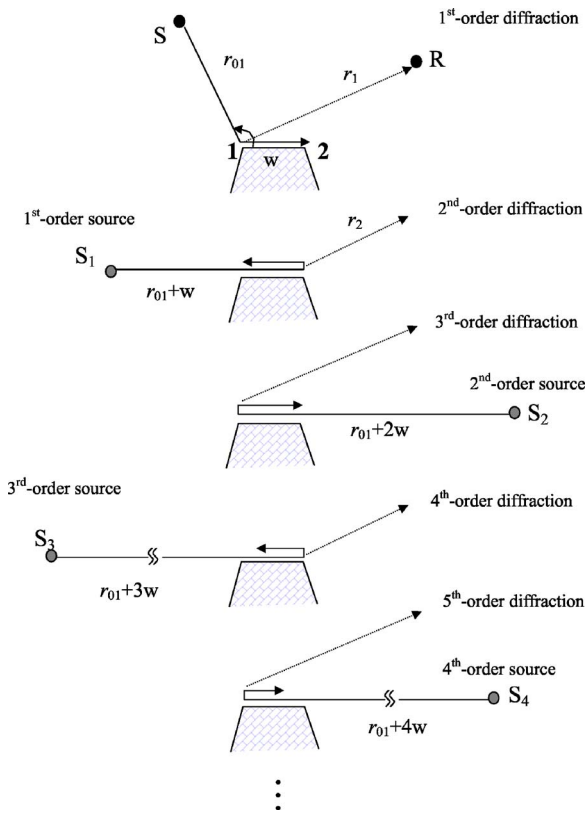


FIG. 5. (Color online) Diagram of higher-order diffraction.

presented an approximate solution of the second-order diffraction in a relatively simple and convenient analytical form.

Our approach is to follow Pierce's method and extend his second-order solution to include all orders, by requiring that the solution converges to the exact solution of an infinitely long knife edge as the thickness of the double edge approaches zero. The diagram of higher-order diffraction is depicted in Fig. 5. S_j with $j=1, 2, \dots, \infty$ are the virtual sources located respectively at the hypothetical extension of the corresponding side of a wedge. The relationship between the consecutive virtual sources is assumed to be

$$S_j = \alpha S_{j-1} D_{\Sigma_p}, \quad (21)$$

where D_{Σ_p} is the first-order diffraction from the edge p , where $p=1$ or 2 corresponding to the edge number (Fig. 6), and α is a coefficient to be determined ($|\alpha| < 1$), which relates diffraction of order j to that of order $j-1$. Note that S_0 is the source strength of the real source at the source position S (Fig. 5). As discussed in Pierce (1974), once the width of the truncated wedge increases, the diffraction of order greater than two can be ignored and the "strength" of the virtual source approaches a value of one-half of that when the wedge is not extended because the total "effective" incident field is the sum of the virtual source and its image. It should be pointed out that since this method is based on the approximate representation, Eq. (3), that requires $kR \gg 1$, where R is defined in Eq. (2c), it implies that the condition $kR \gg 1$ can still be satisfied even if kr approaches zero provided that $kr_0 \gg 1$.

Second-order diffraction.

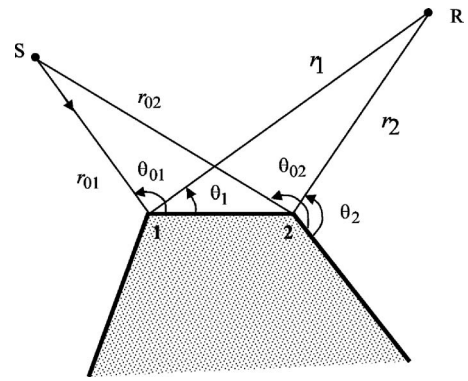


FIG. 6. Geometry illustrating model angles and ranges for the diffraction by an infinitely long straight truncated wedge or a double-edge. 2D cross-sectional view.

There are two ray paths for the second-order diffraction that can be considered separately.

Path 1: $S \rightarrow 1 \rightarrow 2 \rightarrow R$

(a) $S \rightarrow 1 \rightarrow 2$ (first-order diffraction by edge "1" evaluated at edge 2 by assuming edge 2 is not present, i.e. both sides of edge 1 extend to infinity)

$$p_{\text{dif}}^{S12} = S_0 \frac{e^{ik(r_{01}+w)}}{r_{01}+w} D_{\Sigma_1}(w, 0; r_{01}, \theta_{01}). \quad (22a)$$

(b) $S \rightarrow 1 \rightarrow 2 \rightarrow R$ (second-order diffraction by edge 2 evaluated at the receiver by assuming both sides of edge 2 extend to infinity)

$$\begin{aligned} p_{\text{dif}}^{S12R} &= p_{\text{dif}}^{S12} \left(\frac{r_{01}+w}{r_{01}+w+r_2} \right) e^{ikr_2} D_{\Sigma_2}(r_2, \theta_2; r_{01}+w, \theta_{w_2}) \\ &= \alpha S_0 \frac{e^{ikL}}{L} D_{\Sigma_1}(w, 0; r_{01}, \theta_{01}) D_{\Sigma_2}(r_2, \theta_2; r_{01}+w, \theta_{w_2}), \end{aligned} \quad (22b)$$

where $L = r_{01} + w + r_2$.

Path 2: $S \rightarrow 2 \rightarrow 1 \rightarrow R$

(a) $S \rightarrow 2 \rightarrow 1$ (first-order diffraction by edge 2 evaluated at edge 1 by assuming both sides of edge 2 extend to infinity)

$$p_{\text{dif}}^{S21} = S_0 \frac{e^{ik(r_{02}+w)}}{r_{02}+w} D_{\Sigma_2}(w, \theta_w; r_{02}, \theta_{02}). \quad (23a)$$

(b) $S \rightarrow 2 \rightarrow 1 \rightarrow R$ (second-order diffraction by edge 1 evaluated at the receiver by assuming both sides of edge 1 extend to infinity)

$$\begin{aligned} p_{\text{dif}}^{S21R} &= p_{\text{dif}}^{S21} \left(\frac{r_{02}+w}{r_{02}+w+r_1} \right) e^{ikr_1} D_{\Sigma_1}(r_1, \theta_1; r_{02}+w, 0) \\ &= \alpha S_0 \frac{e^{ikL}}{L} D_{\Sigma_2}(w, \theta_{w_2}; r_{02}, \theta_{02}) D_{\Sigma_1}(r_1, \theta_1; r_{02}+w, 0), \end{aligned} \quad (23b)$$

where $L = r_{02} + w + r_1$.

If we set $\alpha = 1/2$, and using the approximate solution of the diffraction function given in Eq. (9), the resultant solution for the second-order diffraction by a double edge is ex-

actly the same as that given in [Pierce \(1974\)](#). Incorporating the same concept of virtual sources, we can extend the second-order solution to higher orders.

All orders of diffraction—General formulas.

Similar to the above-presented approach given for second-order diffraction, general formulas can be obtained for all orders of diffraction,

(1) Even Orders: $2(n+1)^{\text{th}}$, $n=0, 1, \dots$

$$\begin{aligned}
 & \text{Path 1: } S \rightarrow 1 \rightarrow \overbrace{2 \leftrightarrow 1}^{n-1} \rightarrow 2 \rightarrow R \\
 p_{\text{dif}}^{S1(2n1)2R} &= \alpha S_0 \frac{e^{ikL_n}}{L_n} D_{\Sigma_1}(w, 0; r_{01}, \theta_{01}) D_{\Sigma_2}(r_2, \theta_2; r_{01} \\
 & + (2n+1)w, \theta_{w2}) \left\{ \prod_{m=0}^{n-1} \alpha^2 D_{\Sigma_2}(w, \theta_{w2}; r_{01} + (2m \right. \\
 & \left. + 1)w, \theta_{w2}) D_{\Sigma_1}(w, 0; r_{01} + 2(m+1)w, 0) \right\}, \quad (24)
 \end{aligned}$$

where $L_n = r_{01} + (2n+1)w + r_2$. In Eq. (24), the convention $\prod_{m=0}^{-1}(\dots) \equiv 1$ is assumed. Note that when $n=0$, Eq. (24) reduces to the second-order diffraction p_{dif}^{S12R} , Eq. (22b).

$$\begin{aligned}
 & \text{Path 2: } S \rightarrow 2 \rightarrow \overbrace{1 \leftrightarrow 2}^n \rightarrow 1 \rightarrow R. \\
 p_{\text{dif}}^{S2(1n2)1R} &= \alpha S_0 \frac{e^{ikL_n}}{L_n} D_{\Sigma_2}(w, \theta_{w2}; r_{02}, \theta_{02}) D_{\Sigma_1}(r_1, \theta_1; r_{02} \\
 & + (2n+1)w, 0) \left\{ \prod_{m=0}^{n-1} \alpha^2 D_{\Sigma_1}(w, 0; r_{02} + (2m \right. \\
 & \left. + 1)w, 0) D_{\Sigma_2}(w, \theta_{w2}; r_{02} + 2(m+1)w, \theta_{w2}) \right\}, \quad (25)
 \end{aligned}$$

where $L_n = r_{02} + (2n+1)w + r_1$. Also note that when $n=0$, it reduces to the second-order diffraction p_{dif}^{S21R} , Eq. (23b).

(2) Odd Orders: $(2n+1)^{\text{th}}$, $n=1, 2, \dots$

$$\begin{aligned}
 & \text{Path 1: } S \rightarrow 1 \rightarrow \overbrace{2 \leftrightarrow 1}^{n-1} \rightarrow R. \\
 p_{\text{dif}}^{S1(2n1)R} &= \alpha^2 S_0 \frac{e^{ikL_n}}{L_n} D_{\Sigma_1}(w, 0; r_{01}, \theta_{01}) D_{\Sigma_2}(w, \theta_{w2}; r_{01} \\
 & + w, \theta_{w2}) D_{\Sigma_1}(r_1, \theta_1; r_{01} + 2nw, 0) \\
 & \times \left\{ \prod_{m=1}^{n-1} \alpha^2 D_{\Sigma_1}(w, 0; r_{01} \right. \\
 & \left. + 2mw, 0) D_{\Sigma_2}(w, \theta_{w2}; r_{01} + (2m+1)w, \theta_{w2}) \right\}, \quad (26)
 \end{aligned}$$

where $L_n = r_{01} + 2nw + r_1$ and the convention $\prod_{m=1}^0(\dots) \equiv 1$ is assumed.

$$\begin{aligned}
 & \text{Path 2: } S \rightarrow 2 \rightarrow \overbrace{1 \leftrightarrow 2}^{n-1} \rightarrow R. \\
 p_{\text{dif}}^{S2(1n2)R} &= \alpha^2 S_0 \frac{e^{ikL_n}}{L_n} D_{\Sigma_2}(w, \theta_{w2}; r_{02}, \theta_{02}) D_{\Sigma_1}(w, 0; r_{02} \\
 & + w, 0) D_{\Sigma_2}(r_2, \theta_2; r_{02} + 2nw, \theta_{w2}) \\
 & \times \left\{ \prod_{m=1}^{n-1} \alpha^2 D_{\Sigma_2}(w, \theta_{w2}; r_{02} \right. \\
 & \left. + 2mw, \theta_{w2}) D_{\Sigma_1}(w, 0; r_{02} + (2m+1)w, 0) \right\}, \quad (27)
 \end{aligned}$$

where $L_n = r_{02} + 2nw + r_2$.

Having obtained the diffraction for all even and odd orders, the total diffracted field will be the summation of all orders of diffraction including the first-order diffractions from both edges [see Fig. 1(a)—middle drawing]:

$$p_{\text{dif,tot}} = \underbrace{p_{\text{dif}}^{S1R} + p_{\text{dif}}^{S2R}}_{\text{first-order}} + \underbrace{\sum_{n=0}^{\infty} (p_{\text{dif}}^{S1(2n1)2R} + p_{\text{dif}}^{S2(1n2)1R}) + \sum_{n=1}^{\infty} (p_{\text{dif}}^{S1(2n1)R} + p_{\text{dif}}^{S2(1n2)R})}_{\text{higher-order}}. \quad (28)$$

Note that even if $w \rightarrow 0$, the condition $kL_n \gg 1$ can still be satisfied as long as $k(r_{0i} + r_j) \gg 1$, where $i, j=1, 2$, and we can still use the approximate solution Eq. (3) to factor out the diffraction amplitude D_{Σ} from the pressure field.

Determination of coefficient α .

To determine the coefficient α included in Eq. (21), we need to apply the requirement that the total field p_{tot} that includes the incident, reflected, and diffracted wave components is equal to $(2\pi/\theta_w)e^{ikr_0}/r_0$ ($r_{01} = r_{02} = r_0$ as $w \rightarrow 0$) at the apex in the case of a knife edge or a double edge (Table I and also [Pierce, 1981](#)). The total diffracted fields in the different

regions (Fig. 3), even as the observation point approaches to the apex ($w \rightarrow 0$), should be consistent with those described by Eqs. (11)–(13). Each term in Eq. (28) can be expressed as a product of a source term, a propagation term, and a diffraction function, $S_0(e^{ikL}/L)D_{\Sigma}$, with $L=r+r_0$. Therefore, each summation term, in the limiting case of $w \rightarrow 0$, reduces to

$$\begin{aligned} \sum_{n=0}^{\infty} p_{\text{dif}}^{S_1(2n1)2R} &\xrightarrow{w \rightarrow 0} \alpha S_0 D_{\Sigma_1}(0,0;r_{01},\theta_{01}) D_{\Sigma_2}(r_2,\theta_2;r_{01},\theta_{w2}) \\ &\times \sum_{n=0}^{\infty} \alpha^{2n} [D_{\Sigma_2}(0,\theta_{w2};r_{01},\theta_{w2}) D_{\Sigma_1}(0,0;r_{01},0)]^n \\ &= \frac{\alpha D_{\Sigma_1}(0,0;r_{01},\theta_{01}) D_{\Sigma_2}(r_2,\theta_2;r_{01},\theta_{w2})}{1 - \alpha^2 D_{\Sigma_2}(0,\theta_{w2};r_{01},\theta_{w2}) D_{\Sigma_1}(0,0;r_{01},0)}, \end{aligned} \quad (29a)$$

where we have assumed $|\alpha^2 D_{\Sigma_2}(0,\theta_{w2};r_{01},\theta_{w2})$

$\times D_{\Sigma_1}(0,0;r_{01},0)| < 1$ to assure the convergence of the series summation in Eq. (29a). Similarly,

$$\begin{aligned} \sum_{n=0}^{\infty} p_{\text{dif}}^{S_2(1n2)1R} &\xrightarrow{w \rightarrow 0} \alpha S_0 D_{\Sigma_2}(0,0;r_{02},\theta_{02}) D_{\Sigma_1}(r_1,\theta_1;r_{02},0) \\ &\times \sum_{n=0}^{\infty} \alpha^{2n} [D_{\Sigma_1}(0,0;r_{02},0) D_{\Sigma_2}(0,\theta_{w2};r_{02},\theta_{w2})]^n \\ &= \frac{\alpha D_{\Sigma_2}(0,\theta_{w2};r_{02},\theta_{02}) D_{\Sigma_1}(r_1,\theta_1;r_{02},0)}{1 - \alpha^2 D_{\Sigma_1}(0,0;r_{02},0) D_{\Sigma_2}(0,\theta_{w2};r_{02},\theta_{w2})}, \end{aligned} \quad (29b)$$

where, as in Eq. (29a), we have assumed $|\alpha^2 D_{\Sigma_1}(0,0;r_{02},0) D_{\Sigma_2}(0,\theta_{w2};r_{02},\theta_{w2})| < 1$ to assure the convergence of the series summation in Eq. (29b). The two summations for the odd orders are

$$\begin{aligned} \sum_{n=1}^{\infty} p_{\text{dif}}^{S_1(2n1)R} &\xrightarrow{w \rightarrow 0} \alpha^2 S_0 D_{\Sigma_1}(0,0;r_{01},\theta_{01}) D_{\Sigma_2}(0,\theta_{w2};r_{01},\theta_{w2}) D_{\Sigma_1}(r_1,\theta_1;r_{01},0) \\ &\times \sum_{n=0}^{\infty} \alpha^{2n} [D_{\Sigma_1}(0,0;r_{01},0) D_{\Sigma_2}(0,\theta_{w2};r_{01},\theta_{w2})]^n \\ &= \frac{\alpha^2 D_{\Sigma_1}(0,0;r_{01},\theta_{01}) D_{\Sigma_2}(0,\theta_{w2};r_{01},\theta_{w2}) D_{\Sigma_1}(r_1,\theta_1;r_{01},0)}{1 - \alpha^2 D_{\Sigma_1}(0,0;r_{01},0) D_{\Sigma_2}(0,\theta_{w2};r_{01},\theta_{w2})}, \end{aligned} \quad (29c)$$

and

$$\begin{aligned} \sum_{n=0}^{\infty} p_{\text{dif}}^{S_2(1n2)R} &\xrightarrow{w \rightarrow 0} \alpha^2 S_0 D_{\Sigma_2}(0,\theta_{w2};r_{02},\theta_{02}) D_{\Sigma_1}(0,0; \\ &\times r_{02},0) D_{\Sigma_2}(r_2,\theta_2;r_{02},\theta_{w2}) \\ &\times \sum_{n=0}^{\infty} \alpha^{2n} [D_{\Sigma_2}(0,\theta_{w2};r_{02},\theta_{w2}) D_{\Sigma_1}(0,0;r_{02},0)]^n \\ &= \frac{\alpha^2 D_{\Sigma_2}(0,\theta_{w2};r_{02},\theta_{02}) D_{\Sigma_1}(0,0;r_{02},0) D_{\Sigma_2}(r_2,\theta_2;r_{02},\theta_{w2})}{1 - \alpha^2 D_{\Sigma_2}(0,\theta_{w2};r_{02},\theta_{w2}) D_{\Sigma_1}(0,0;r_{02},0)}, \end{aligned} \quad (29d)$$

respectively. Since as $w \rightarrow 0$, $r_{01}=r_{02}=r_0$, $r_1=r_2=r$, and the double edge reduces to a single knife edge with a wedge angle:

$$\theta_{sw} = \theta_{w1} + \theta_{w2} - \pi. \quad (30)$$

Note that for $w \rightarrow 0$, L_n 's in Eqs. (24)–(27) are the same. Using Eq. (3), the total diffraction amplitude, $D_{\Sigma_{\theta_{sw}}}$, can be obtained by normalizing Eq. (28) with $S_0 e^{ikL}/L$ and can be written as

$$\begin{aligned} D_{\Sigma_{\theta_{sw}}}(r,\theta;r_0,\theta_0) &= D_{\Sigma_{\theta_{w1}}}(r,\theta+\pi-\theta_{w2};r_0,\theta_0+\pi-\theta_{w2}) \\ &+ D_{\Sigma_{\theta_{w2}}}(r,\theta;r_0,\theta_0) + D_{\Sigma_m}(\alpha), \end{aligned} \quad (31)$$

where the last term, $D_{\Sigma_m}(\alpha)$, on the right-hand side of Eq.

(31) represents the summation of four multiple-order diffractions given by Eqs. (29a)–(29d). In all of the above-presented equations, Eqs. (21), (22a), (22b), (23a), (23b), (24)–(28), (29a)–(29d), (30), and (31), the angles are measured counterclockwise from the right side of the corresponding wedge (Fig. 6). In addition, we have assumed that $r \gg w$ and $r_0 \gg w$, which lead to $\theta_{01} = \theta_0 + \pi - \theta_{w2}$ and $\theta_1 = \theta + \pi - \theta_{w2}$. Solving Eq. (31) for an angle dependent $\alpha(\theta)$, we obtain a coefficient that characterizes the virtual source, $\alpha_0(\theta)$, for $w=0$. On the other hand, we know that when $w \rightarrow \infty$, $\alpha_{\infty}(\theta) \rightarrow 1/2$ as discussed previously in Sec. III B (see also Pierce, 1974). To obtain α for an arbitrary w , we use an interpolation function

$$\alpha_w(\theta) = [\alpha_0(\theta) - \alpha_\infty(\theta)]e^{-\beta w/\lambda} + \alpha_\infty(\theta), \quad (32)$$

where λ is the wavelength and the coefficient β controls the rate of $\alpha_w(\theta)$ approaching $\alpha_\infty(\theta)$. This function, chosen heuristically, is used because it smoothly connects the limiting values of α over the full range $0 \leq w < \infty$. The coefficient β can be determined through comparison of the prediction with either data (which will be the case in this paper) or with an exact solution.

Having obtained $\alpha_w(\theta)$, we can compute all diffraction terms defined in Eqs. (24)–(27), and hence the total diffraction given in Eq. (28). It should be pointed out that Eq. (28) is based on the scattering geometry shown in Fig. 6. If, however, one of the edges is in the shadow zone of the source, such as edge 2 is in the shadow zone or $\theta_{01} > \pi$, the corresponding first-order diffraction term, p_{dif}^{S2R} , as well as the corresponding terms for the higher-order diffraction, $p_{\text{dif}}^{S2(1n2)1R}$ and $p_{\text{dif}}^{S2(1n2)R}$, should be removed. Furthermore, if the receiver is also in the shadow zone relative to one of the edges, such as $\theta_2 < \theta_{w2} - \pi$, which is the geometry studied in Pierce (1974), the remaining first-order diffraction term, p_{dif}^{S1R} , and an additional higher-order diffraction term, $p_{\text{dif}}^{S1(2n1)R}$, should also be removed. As a result, the total diffraction contains

only one term of the higher-order diffraction, $p_{\text{dif}}^{S1(2n1)2R}$.

V. RESULTS AND DISCUSSION

Numerical example: Determination of $\alpha(\theta)$ for a semi-infinite plate of finite thickness.

For a plate, we have $\theta_{w1} = \theta_{w2} = \theta_w = 3\pi/2$, which results in $D_{\Sigma_1} = D_{\Sigma_2} = D_\Sigma$ and $\theta_{sw} = 2\pi$ by Eq. (30). Furthermore, if we assume a backscattering geometry and let $w \rightarrow 0$, we have $\theta_2 = \theta_{02} = \theta$, $\theta_1 = \theta_{01} = \theta - \pi/2$ (Fig. 6), and $r_{01} = r_{02} = r_0 = r_1 = r_2 = r$. Using the symmetry of the wedge, we have

$$\begin{aligned} D_{\Sigma_2}(0, \theta_{w2}; r_{01}, \theta_{w2}) &= D_{\Sigma_1}(0, 0; r_{01}, 0) = D_{\Sigma_1}(0, 0; r_{02}, 0) \\ &= D_{\Sigma_2}(0, \theta_{w2}; r_{02}, \theta_{w2}) \\ &= D_\Sigma(0, 0; r_0, 0) = D_\Sigma(0, 0; r, 0). \end{aligned} \quad (33)$$

Thus, the denominators in Eqs. (29a)–(29d) can be expressed as

$$\text{Denominator} = 1 - \alpha^2 D_\Sigma^2(0, 0; r, 0). \quad (34)$$

By combining Eq. (29a) with Eq. (29d), and Eq. (29b) with Eq. (29c), the summation of the four terms involving higher-order diffraction in Eq. (31) can be rewritten as

$$\begin{aligned} D_{\Sigma_m}(\alpha) &= \frac{\alpha D_\Sigma(r, \theta; r, \theta_w) [D_\Sigma(0, 0; r, \theta - \pi/2) + \alpha D_\Sigma(0, \theta_w; r, \theta) D_\Sigma(0, 0; r, 0)]}{1 - \alpha^2 D_\Sigma^2(0, 0; r, 0)} \\ &+ \frac{\alpha D_\Sigma(r, \theta - \pi/2; r, 0) [D_\Sigma(0, \theta_w; r, \theta) + \alpha D_\Sigma(0, 0; r, \theta - \pi/2) D_\Sigma(0, 0; r, 0)]}{1 - \alpha^2 D_\Sigma^2(0, 0; r, 0)} \\ &= \frac{\alpha D_\Sigma(0, \theta_w; r, \theta) [D_\Sigma(r, \theta; r, \theta_w) + D_\Sigma(r, \theta - \pi/2; r, 0)]}{1 - \alpha D_\Sigma(0, 0; r, 0)}. \end{aligned} \quad (35)$$

To obtain the last result of Eq. (35) we limited the scattering angle by $\pi < \theta < 3\pi/2$, a condition consistent with the experiment configuration discussed in the next section. As a result, $D_\Sigma(0, 0; r, \theta - \pi/2)$, $D_\Sigma(0, \theta_w; r, \theta)$, and $D_\Sigma(0, 0; r, 0)$ are all in Region I (Fig. 3), and thus have the same value (Table I):

$$\begin{aligned} D_\Sigma(0, \theta_w; r, \theta) &= D_\Sigma(0, 0; r, \theta - \pi/2) \\ &= D_\Sigma(0, 0; r, 0) = \frac{2\pi}{\theta_w} - 2 = -2/3. \end{aligned} \quad (36)$$

Substituting Eq. (36) into Eq. (35) leads to

$$D_{\Sigma_m}(\alpha) = -\left(\frac{2\alpha}{3}\right) \frac{[D_\Sigma(r, \theta; r, \theta_w) + D_\Sigma(r, \theta - \pi/2; r, 0)]}{1 + 2\alpha/3}. \quad (37)$$

If we now express θ_D in terms of θ_w and θ as illustrated in Fig. 2,

$$\theta_D = \theta_w - \theta, \quad (38)$$

we are able to apply the simulation results directly to the diffraction by the leading edge of a circular disk. For $\pi < \theta < 3\pi/2$, we have $0 < \theta_D < \pi/2$. Inserting Eq. (37) into Eq. (31) and replacing θ with θ_D result in

$$\begin{aligned} D_{\Sigma_{2\pi}}(r, 3\pi/2 - \theta_D; r, 3\pi/2 - \theta_D) &= D_{\Sigma_{3\pi/2}}(r, \pi - \theta_D; r, \pi - \theta_D) \\ &+ D_{\Sigma_{3\pi/2}}(r, 3\pi/2 - \theta_D; r, 3\pi/2 - \theta_D) - \left(\frac{2\alpha}{3}\right) \\ &\times \frac{D_{\Sigma_{3\pi/2}}(r, 3\pi/2 - \theta_D; r, \theta_w) + D_{\Sigma_{3\pi/2}}(r, \pi - \theta_D; r, 0)}{1 + 2\alpha/3}. \end{aligned} \quad (39)$$

In Eq. (39), we use subscripts 2π and $3\pi/2$ to differentiate the wedge angles between a knife edge ($\theta_w = 2\pi$) and a right-angle wedge ($\theta_w = 3\pi/2$) explicitly. If we let

$$D_{\Sigma_{\text{knife}}} = D_{\Sigma_{2\pi}}(r, 3\pi/2 - \theta_D; r, 3\pi/2 - \theta_D), \quad (40a)$$

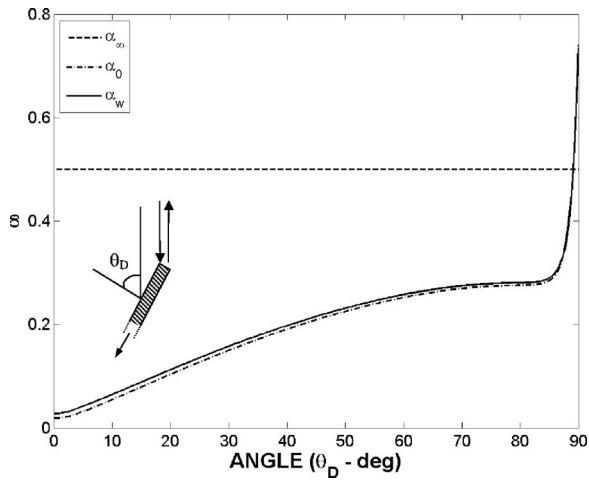


FIG. 7. Coefficient α vs angle for several cases. The dashed curve corresponds to the limit of the thickness $w \rightarrow \infty$, the dot-dashed curve is for the limit $w \rightarrow 0$ determined by Eq. (41), and the solid curve corresponds to interpolated α computed from the interpolation function Eq. (32) with $w = 1.9$ mm and the frequency of 60 kHz and a fit to laboratory data, as described later in this section.

$$D_{\Sigma_{1st}} = D_{\Sigma_{3\pi/2}}(r, \pi - \theta_D; r, \pi - \theta_D) + D_{\Sigma_{3\pi/2}}(r, 3\pi/2 - \theta_D; r, 3\pi/2 - \theta_D), \quad (40b)$$

and

$$D_{\Sigma_{high}} = D_{\Sigma_{3\pi/2}}(r, 3\pi/2 - \theta_D; r, \theta_w) + D_{\Sigma_{3\pi/2}}(r, \pi - \theta_D; r, 0) \quad (40c)$$

represent the terms associated with the diffraction by the single knife edge, the first-order diffraction by the double edge, and the higher-order diffraction by the double edge, the coefficient $\alpha_0(\theta_D)$ can be obtained by

$$\alpha_0(\theta_D) = -\frac{3}{2} \left(1 + \frac{D_{\Sigma_{high}}}{D_{\Sigma_{knife}} - D_{\Sigma_{1st}}} \right)^{-1}. \quad (41)$$

The computed coefficient $\alpha_0(\theta_D)$ (dot-dashed line) is illustrated in Fig. 7. From Fig. 7, we see that $\alpha_0(\theta_D)$ is less than $\frac{1}{2}$ when θ_D is well away from edge-on incidence, or $\theta_D \ll \pi/2$, indicating a small contribution from the higher-order diffraction to the total diffraction field. As θ_D increases but is not too close to edge-on incidence, $\alpha_0(\theta_D)$ increases with a moderate rate (slope). Once the scattering angle θ_D approaches the edge-on angular position, $\theta_D > 85^\circ$, the coefficient $\alpha_0(\theta_D)$ increases with a much higher rate and approaches the value of $\frac{3}{4}$ at the edge-on angular position ($\theta_D = \pi/2$), indicating a much increased contribution from the higher-order diffraction.

One of the interesting problems in edge diffraction we want to address is how thin a plate needs to be in order for a plate to behave as a knife edge, or a screen, for which only the first-order diffraction from a knife edge ($\theta_w = 2\pi$) needs to be considered. Jones (1953) suggested that for the case of a plane incident wave, once the thickness of the plate is one-tenth of the acoustic wavelength, the plate can be treated as a knife edge. However, Jones did not provide any angle dependence or a quantitative basis for the statement. An example of such a problem is illustrated in Fig. 8, where the

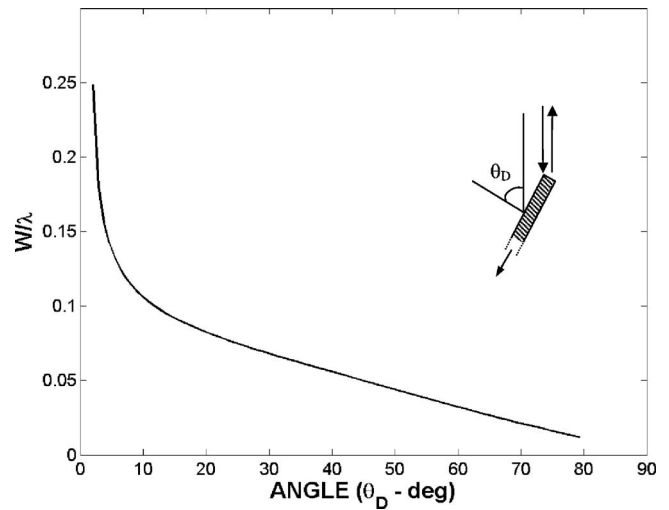


FIG. 8. The ratio of width to wavelength of the truncated wedge, which is required to make the difference in edge diffraction between a knife edge and a double edge of finite thickness less than 5%, vs the scattering angle for a backscattering geometry.

ratio of the plate width to the wavelength (w/λ) that is required to provide a difference in diffraction level between a knife edge and a double edge of less than 5%, or 0.5 dB, is plotted as a function of θ_D . It can be seen that at a scattering angle of $\theta_D < 10^\circ$, w/λ is greater than 0.1, indicating that when the incident wave approaches the direction parallel to the normal of the plate flat surface (broadside incidence), the higher-order diffraction resulting from the interaction between the two edges can be ignored. As the scattering angle increases, the ratio or the projected width of the plate ($D \cos \theta_D$, where D is the diameter of the disk) decreases, indicating that more and more contribution from the higher-order diffraction has to be considered. For $\theta_D < 2^\circ$, the difference between the knife-edge solution and Eq. (28) is less than 5%. Further calculation indicates that for the limiting case as $\theta_D \rightarrow 0$, the maximum difference between the two models is less than 1.5%.

When the scattering angle is greater than 80° , the ratio approaches zero, corresponding to a sudden change in slope of the α curve shown in Fig. 7. The physical explanation for this phenomenon is not clear at this point, it may be related to our heuristic approach that forces the total diffraction based on the ray representation, Eq. (28), to converge to the first-order diffraction by a knife edge once the width of a double edge approaches zero.

Figure 9 demonstrates the relative importance of the contributions from higher-order diffraction in a different way at two incident angles, 10° and 89° , respectively. It can be seen that for an incident angle close to the broadside of the disk ($\theta_D = 10^\circ$), the total field including only the first-order diffraction can provide satisfactory result with a relative error of about 10% or less. In contrast, as the angle of incidence moves away from the broadside ($\theta_D = 89^\circ$), the relative error increases to more than 400% when including only the first-order diffraction, to about 70% when the second-order diffraction is included, and to less than 10% when the third-order diffractions are included. For the diffraction order

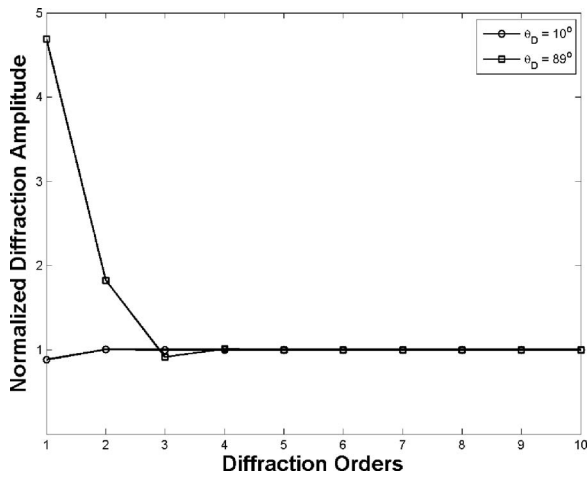


FIG. 9. Diffraction amplitude normalized by the converged amplitude ($n \rightarrow \infty$) as a function of diffraction orders at two different incident angles for a monostatic scattering geometry.

greater than 4, the diffraction is basically indistinguishable from the converged value for $n \rightarrow \infty$. These results are consistent with those shown in Figs. 7 and 8.

Comparison with experimental data.

(1) Diffraction by metallic disks. Laboratory experiments involving a wide range of metal circular disks were conducted in Fall 2002. The purpose of the experiments was to understand the acoustic scattering by the seafloor covered with benthic shells. Metal disks were chosen as part of these free-field scattering measurements as they possess similar characteristics to the shells (Stanton and Chu, 2004). The experiments were conducted in an elongated rectangular tank with the scattering targets located about 3 m away from the transmitter-receiver pair in a near-backscattering configuration. Details of the experimental setup can be found in Stanton and Chu (2004). The PWTS, representing the edge diffraction from leading edge, was extracted from the backscattered time series by using a pulse compression technique (Chu and Stanton, 1998). Although the elasticity of the metal disk could also play a role in the diffraction process, the exclusion of the arrivals later than the first arrival for each ping can greatly reduce, if not eliminate, the influence resulting from the possible elastic surface and plate waves. In addition, since the reflection coefficient for an infinite aluminum plane is close to unity, the diffraction model based on the rigid wedge approximation is assumed to be valid in the following data/model comparison.

As mentioned in Sec. I, there are two aspects in constructing a complete diffraction model by a circular disk from the solution by an infinitely long knife edge. One is to take into account the finite thickness by including the higher-order diffraction, which is the subject of this paper, and the other is to account for the deformation of the edge. The deformation factor can be computed by a line integral over the leading edge of the disk and can be expressed as (Stanton et al., 2007):

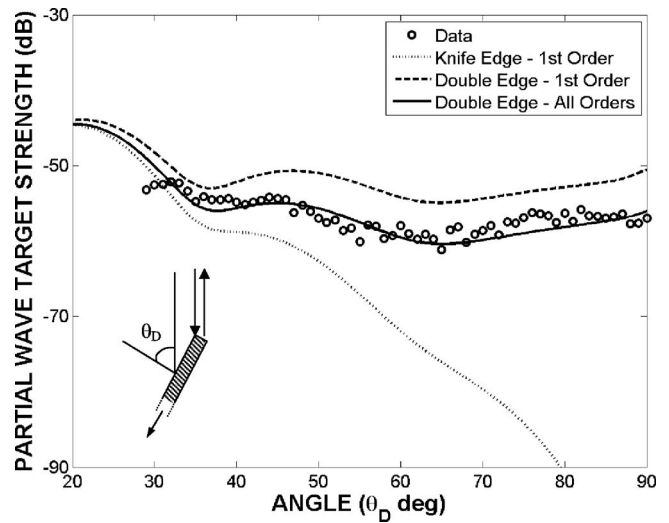


FIG. 10. Comparison of laboratory data (60 kHz) involving an aluminum disk with various diffraction models.

$$f_D = a \left[J_0(2ka \sin \theta_D) - i \frac{2}{\pi} s_{0,0}(2ka \sin \theta_D) \right], \quad (42)$$

where a is the radius of the disk, $J_0(x)$ is the Bessel function of order zero, and $s_{0,0}$ is the Lommel function of order zero (Abramowitz and Stegun, 1974). The complete solution for the diffraction by a circular disk of finite thickness is the product of $p_{dif,tot}$ from Eq. (28) and the deformation factor f_D .

A comparison of the model and the data is given in Fig. 10. The diffraction data as a function of backscattering angle shown in Fig. 2 are replotted in Fig. 10 (open circles), superimposed with several diffraction models. The disk is made of aluminum with a diameter of 8 cm and a thickness of about 1.9 mm (0.075 in.), which corresponds to 0.08λ at 60 kHz (open circle in Fig. 4). One of the diffraction models assumes a knife-edge geometry, or a semi-infinite screen (dotted). As described in Sec. I, the diffraction by a knife-edge decreases rapidly as $\theta_D \rightarrow \pi/2$, and approaches zero once the incidence is edge-on. The contribution from the two first-order diffractions from edges 1 and 2 defined in Eq. (28) is also plotted in Fig. 10 (dashed). The solid curve representing the total diffraction, including all orders of diffraction, is computed from Eq. (28) with source coefficient $\alpha_w(\theta_D)$ computed from Eq. (32) and illustrated in Fig. 7 (solid). The interpolation coefficient, β in Eq. (32) is assumed to have a linear relation with the ratio parameter, w/λ ,

$$\beta = \beta_1 w/\lambda + \beta_0. \quad (43)$$

Equation (43) results in a quadratic dependence of w/λ for the exponent of the first term of the interpolation function in Eq. (32). The two constants $\beta_0 = -2.7$ and $\beta_1 = 34.9$ are determined based on the visual fit of the computed PWTS to the data in Fig. 10. From Fig. 7, we see that the interpolated $\alpha_w(\theta_D)$ used to provide the “best” fit to the data is much closer to $\alpha_0(\theta_D)$ than to the asymptotic $\alpha_\infty(\theta_D)$, which is $\frac{1}{2}$.

Figure 11 shows the comparison of the model with the diffraction data from a number of aluminum disks with the same diameter but different thickness and at different back-

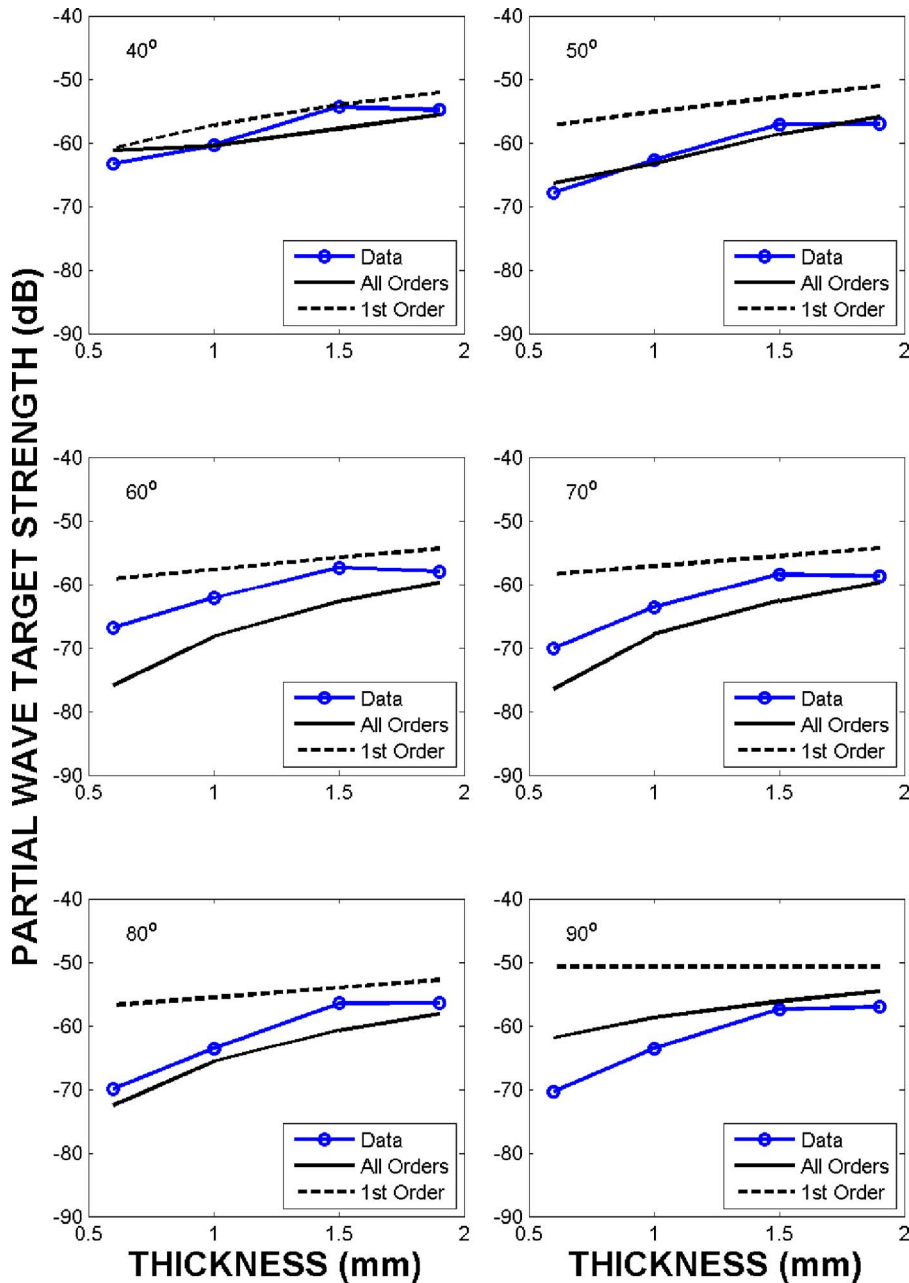


FIG. 11. (Color online) Comparison of the measured partial wave target strength of an aluminum disk (diameter of 8 cm) of various thickness ($w = 0.6, 1.0, 1.5,$ and 1.9 mm) with the model of all orders of diffractions [Eq. (28)] at different angles. For the scattering angle at 90° (bottom right), the actual angle is 89.6° to exclude the “edge-on” specular term.

scattering angles. The same value of β_0 and β_1 that were determined in Fig. 10 were used to compute the theoretical higher-order diffraction in each of the plots in Fig. 11. In general, the PWTS at all angles decreases as the disk becomes thinner. For all plots, the slopes of the model are very close to those of the data. The predicted absolute values of PWTS are in general agreement with some of the data but no more than 10 dB difference for the rest of the data such as at 60° . The comparison indicates that the proposed higher-order diffraction model has significant merit in describing the edge diffraction by circular disks of finite thickness. The mismatch in this comparison may result from a number of factors including approximations made in deriving the deformation factor, the assumptions made in deriving factor α , and the approximation of neglecting the elasticity of the disks, as well as the uncertainty in measuring the scattering geometry. For the purpose of comparison, the model predictions includ-

ing only the first-order diffraction are also presented, which overestimate the total diffraction in its overall level. Furthermore, the slope predicted by the double-edge first-order diffraction model approaches zero and deviates more and more from the data as $\theta_D \rightarrow 90^\circ$. This is consistent with the results of Figs. 9 and 10.

(2) Diffraction by a hard strip. Medwin *et al.* (1982) used the concept of Huygens impulsive wavelets to model the second-order diffraction by a hard strip. Their solution was based on the exact impulse solution of the diffraction by an infinitely long straight wedge (Biot and Tolstoy, 1957), or BT solution, and was transformed to the frequency domain. Their model predictions agreed with the experimental data very well. Here we will present an alternative solution to the problem to demonstrate the applicability of our approach.

For a hard strip, the n th order diffraction results from the contributions corresponding to 2^n different diffraction ray

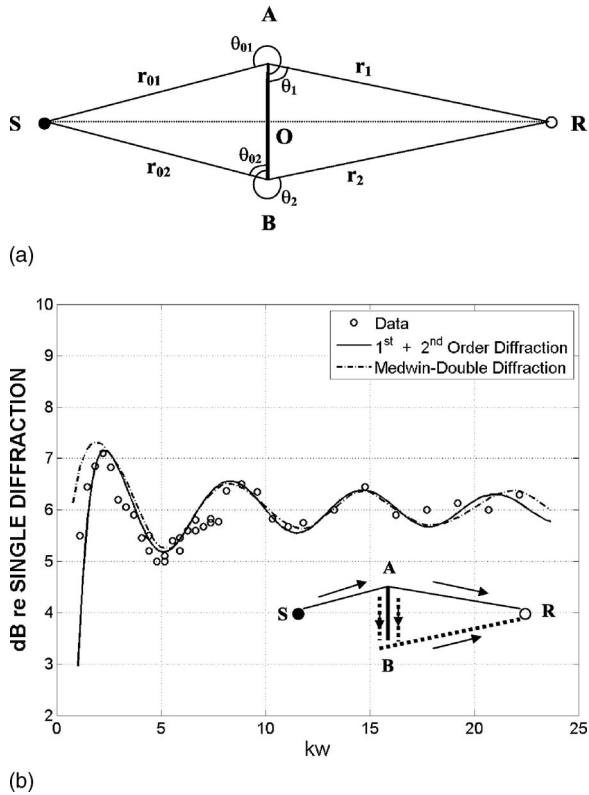


FIG. 12. (a) Geometry and notations of the parameters for diffraction by a hard strip. The view is along the infinite length of the strip and the distance between points “A” and “B” is the width w . (b) Comparison of data (Medwin *et al.*, 1982) with the second-order diffraction model [Eq. (47)] (solid) and Medwin’s “double” or second-order diffraction (dashed). The data are the ratio of the total field to that of one-half of the total first-order diffraction (ray paths SAR and SBR). The model parameters are $\alpha_0=0.2$, $\alpha_\infty=0.5$, $\beta_0=-2.7$, and $\beta_1=34.9$ [Eq. (32)]. The distances from the strip to the source and the receiver are 14.5 cm [OS, see (a)] and 23.2 cm [OR, see (a)], respectively. The width of the strip (w) is 4 cm. The frequencies are from about 1.5 to 30 kHz. Note that the horizontal axis in Medwin *et al.* is frequency in log-scale and we use a dimensionless variable kw on a linear scale, where w is the width of the strip.

paths. This is because for each new higher-order diffraction there will be twice as many ray paths propagating along both sides of the strip. For this particular geometry, it is extremely hard, if not impossible, to find a general expression of the total diffraction in closed form by summing all orders of the diffraction similar to Eq. (28) with the explicit expressions given in Eqs. (24)–(27). As a result, we cannot use the same method to determine the magnitude of the virtual source, α_0 , as for the double-edge problem associated with a disk of finite thickness. However, since we know that if the value of kw , which can be determined from the width of the strip (w) and the acoustic frequency used in their experiment, is large compared to unity, the diffraction with order higher than 2 is negligible. To compare directly with the laboratory experimental data for a hard strip published by Medwin *et al.* (1982), instead of deriving the solution for the total field as we did for the disk experiments, we now use the solutions given by Eq. (3), Eqs. (22a)–(22b), and Eqs. (23a)–(23b), to express as the ratio of the “total diffracted field” (first and second) to the first-order diffraction from edge “ k ,” where k can be either 1 or 2 corresponding to edge “A” and “B,” respectively [Fig. 12(a)],

$$R_N = 2 + \frac{\alpha}{\xi_k D_{\Sigma_k}(r_k, \theta_k; r_{0k}, \theta_{0k})} \times \{ \xi_{12} D_{\Sigma_1}(w, 0; r_{01}, \theta_{01}) D_{\Sigma_2}(r_2, \theta_2; r_{01} + w, \theta_w) + \xi_{12} D_{\Sigma_1}(w, \theta_w; r_{01}, \theta_{01}) D_{\Sigma_2}(r_2, \theta_2; r_{01} + w, 0) + \xi_{21} D_{\Sigma_2}(w, 0; r_{02}, \theta_{02}) D_{\Sigma_1}(r_1, \theta_1; r_{02} + w, \theta_w) + \xi_{21} D_{\Sigma_2}(w, \theta_w; r_{02}, \theta_{02}) D_{\Sigma_1}(r_1, \theta_1; r_{02} + w, 0) \}, \quad (44)$$

where the factors ξ_k , with $k=1$ or 2, is the propagation term associated with edge A or B, and ξ_{ij} , where $i, j=1, 2$, are the propagation terms corresponding to the different ray paths between the first- and second-order diffractions,

$$\xi_k = \frac{e^{ik(r_{0k}+r_k)}}{r_{0k} + r_k}, \quad \xi_{ij} = \frac{e^{ik(r_{0i}+w+r_j)}}{r_{0i} + w + r_j}. \quad (45)$$

For a symmetrical scattering geometry shown in Fig. 12(a), we have $r_{01}=r_{02} \equiv r_0$, $r_1=r_2 \equiv r$, $\theta_{w1}=\theta_{w2} \equiv \theta_w$, $\theta_{01}=\theta_w-\theta_{02} \equiv \theta_0$, and $\theta_2=\theta_w-\theta_1 \equiv \theta$. Since $\theta_w=2\pi$ and $\nu = \pi/\theta_w=1/2$ for a strip, it is easy to show that

$$\begin{aligned} D_{\Sigma_1}(r_1, \theta_1; r_{01}, \theta_{01}) &= D_{\Sigma_2}(r_2, \theta_2; r_{02}, \theta_{02}) \\ &\equiv D_{\Sigma}(r, \theta; r_0, \theta_w - \theta_0) \\ &= -D_{\Sigma}(r, \theta; r_0, \theta_0) \\ &= D_{\Sigma}(r, \theta_w - \theta; r_0, \theta_0). \end{aligned} \quad (46)$$

Using these relations, Eq. (44) can be simplified to

$$R_N = 2 - 4\alpha e^{ikw} \left(\frac{r + r_0}{r_0 + r + w} \right) \times \frac{D_{\Sigma}(w, 0; r_0, \theta_0) D_{\Sigma}(r, \theta; r_0 + w, 0)}{D_{\Sigma}(r, \theta; r_0, \theta_0)}. \quad (47)$$

Comparison of the computed theoretical predictions with the laboratory experimental data published by Medwin *et al.* (1982) is shown in Fig. 12(b). Since the experiment was performed in the air, the strip could be regarded approximately as a “hard” material. The model parameters needed to compute $\alpha_w(\theta)$ in Eq. (32) are $\alpha_0=0.2$ and $\alpha_\infty=0.5$. As stated previously, the determination of α_0 for this problem is different from the backscattering case [Eq. (39)]. As an approximation, we choose 0.2, an average value of α_0 shown in Fig. 7, as the estimated α_0 value. The same values of β_0 and β_1 that were used in Fig. 10 with the disk data were also used in obtaining Fig. 12(b). The width of the strip, w , is 4 cm. The frequencies vary from about 15 to 30 kHz. The distances from the strip to the source and the receiver are 14.5 cm [OS, see Fig. 12(a)] and 23.2 cm [OR, see Fig. 12(a)], respectively. The agreement between the data and the theory is very good over a wide range of kw . To illustrate the difference between Medwin’s second-order diffraction model (Medwin *et al.*, 1982) and the solution presented here, we superimposed the theoretical predictions from their diffraction model as a comparison. It is seen that at lower frequencies, or small kw , the present model provides better fit and the two models are converged as kw increases. This is expected since at lower frequencies, higher-order diffraction is important (see

Fig. 9). As kw increases, $\alpha_w(\theta)$ approaches $\alpha_\infty=0.5$, and the current diffraction model reduces to the Pierce's second-order diffraction model. Since the BT solution, from which Medwin's diffraction model was derived, is exactly the same as the Pierce's solution (see the Appendix), from which the current higher-order diffraction is derived, Medwin's second-order diffraction model is essentially the same as our higher-order diffraction model for large kw . The deviation between the two theoretical models at higher frequencies ($kw > 18$) is likely due to the difficulties in numerical computations of the double diffraction.

From the three examples (Figs. 10–12), we can conclude that although the two empirical parameters, β_0 and β_1 , are determined originally from fitting particular backscattering data set, the fact that the reasonable agreement between data and theory by using the same β_0 and β_1 for all three different cases, i.e., versus incident angle (Fig. 10), width (Fig. 11), and frequency (Fig. 12), suggests that the assumed relations given by Eqs. (32) and (43) work reasonably well at least for the diffraction by metal disks of finite thickness and by strip of finite width. Particularly, since the latter two examples involve varying w/λ , the reasonable agreement between the model and the data suggests that the approximation of the exponent term of the interpolation function, Eqs. (32) and (43), by a second-order polynomial of w/λ in obtaining α may be justified. In addition, the three examples also suggest that the two parameters work well not only for backscattering application (Figs. 10 and 11) but also for bistatic scattering application (Fig. 12). This is actually expected since in theory, the general solution of Eq. (28), as well as Eq. (32), is valid for all scattering geometry.

VI. SUMMARY AND CONCLUSION

We have developed an approximate analytical solution of bistatic scattering by a double edge of finite thickness that includes all orders of diffraction from the edges. The solution is an extension of a previously developed second-order diffraction model (Pierce, 1974). An amplitude function of the virtual source (α), a function of the scattering geometry, the thickness of the double edge, and the acoustic frequency, is introduced heuristically to make the ray-based solution converge to the exact solution of an infinitely long knife edge as the thickness of the double edge approaches zero. A heuristic interpolation function is used to obtain the α for a double edge of an arbitrary thickness with two empirically determined parameters.

The solution, when combined with the approximate solution to account for the deformation of a general edge (Stanton *et al.*, 2007), is used to describe the edge diffraction by a disk of finite thickness. It is found that the contribution of the higher-order diffraction is small at or near broadside incidence, increases with the incident angle at a moderate rate if the incidence is not too close to edge-on, and increases drastically as the incidence approaches the edge-on direction ($\theta_D \rightarrow \pi/2$). The agreement between the theory and the data is generally good. An approximate and closed form solution of the diffraction by a hard strip that takes into account the first- and second-order diffractions is presented and is com-

pared favorably with the laboratory experimental data over a wide range of kw values, where k is the wave number and w is the width of the strip.

In addition, the equivalence between the exact impulse solution of an infinitely long straight wedge in the time domain (Biot and Tolstoy, 1957) and the integral solution in the frequency domain given by various authors (Sommerfeld, 1954; Pierce, 1981) has been shown. Furthermore, the behavior of the diffracted field due to the presence of an infinitely long straight wedge under different conditions has been studied systematically and analytically. It is found that the diffracted field increases drastically as it approaches either the reflection or shadow boundaries and has an amplitude value of one-half of the incident and/or reflected waves. At or near the apex of the wedge, the diffracted field is constant within the same diffraction regions but discontinuous across the reflection and shadow boundaries by a magnitude value of unity (Table I). It is also found that the edge diffraction has the same order of magnitude as the incident and/or the reflected fields.

Despite the promising comparison between the data and our higher-order diffraction model, it is recognized that in deriving the approximate solution for the diffraction by a double edge we have used several assumptions: (1) The relationship between the virtual sources of adjacent order specified in Eq. (21); (2) an empirical interpolation function defined by Eq. (32); and (3) a linear relation between the interpolation coefficient β and w/λ [Eq. (43)], where w is the distance between the apexes of the two edges [Figs. 1(b) and 5] or the width of the strip [Fig. 12(a)] and λ is the wavelength. These heuristic relations (assumptions) allow us to understand the more complicated double-edge diffraction problem involving all orders of diffraction in a relatively simple way and gain more physical insight into the mechanisms of the higher-order diffraction by a double edge.

In conclusion, we have derived an approximate ray-based solution of diffraction by double edges in a general form that can take into account all orders of diffraction. The solution can, in principle, be extended to include more than two edges.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their helpful comments and suggestions, which helped to improve the quality of the paper significantly. This work was supported by the US Office of Naval Research and by the Woods Hole Oceanographic Institution.

APPENDIX: EQUIVALENCY BETWEEN THE EXACT TRANSIENT IMPULSE SOLUTION AND THE CW REPRESENTATION OF THE DIFFRACTED WAVE DUE TO AN INFINITELY LONG RIGID WEDGE

Pierce (1981, pp. 489–490) briefly provided the procedures of how to convert the contour-integral-based spectral representation to a transient solution, but did not provide the closed-form impulse solution. Here we use a different method to prove the equivalency between the exact spectral representation [Eq. (2a)–(2c)] and the exact impulse solution given by Biot and Tolstoy (1957).

The exact impulse solution of the diffracted wave due to a point source for a rigid wedge, given by [Biot and Tolstoy \(1957\)](#), is

$$p_{\text{diffr}}(t) = \frac{\rho c}{4\pi\theta_w r r_0} \frac{e^{-\nu\eta}}{\sinh \eta} \sum_{\pm} \frac{\sin[\nu(\pi \pm \theta \pm \theta_0)]}{1 - 2e^{-\nu\eta} \cos[\nu(\pi \pm \theta \pm \theta_0)] + e^{-2\nu\eta}} u(t - \tau_0), \quad (\text{A1})$$

where $u(x)$ is a unit step function and $\eta(t)$ is a function of time t and the other geometric parameters

$$\eta(t) = \cosh^{-1} \frac{c^2 t^2 - (r^2 + r_0^2 + (z - z_0)^2)}{2rr_0}, \quad (\text{A2})$$

where c is the sound speed. τ_0 is the shortest time required for the wave traveling from the source to the apex, and then to the receiver, or the time when $\eta(t)=0$,

$$\tau_0 = \frac{((r + r_0)^2 + (z - z_0)^2)^{1/2}}{c} = \frac{L}{c}, \quad (\text{A3})$$

where L is given in Eq. (4). The source term ρ is the volume flux and is related to the source strength S in Eq. (2a)–(2c) by $S = -\rho/4\pi$, i.e., volume flux per unit solid angle. The summation in (A1) represents four combinations: $++$, $+-$, $-+$, and $--$. It can be shown that straightforward manipulations of Eq. (A1) lead to

$$p_{\text{diffr}}(t) = \frac{\rho c}{4\pi\theta_w r r_0} \frac{\sin \nu\pi}{\sinh \eta} \sum_{\pm} \frac{\cos \nu\pi - \cosh \nu\eta \cos[\nu(\theta \pm \theta_0)]}{\cos \nu\eta - \cos[\nu(\theta \pm \theta_0)] - \sin^2 \nu\eta \sin^2[\nu(\theta \pm \theta_0)]} u(t - \tau_0). \quad (\text{A4})$$

The spectral representation can be obtained via the Fourier transform of Eq. (A1) as

$$p_{\text{dif}}(f) = \int_{-\infty}^{\infty} p_{\text{diffr}}(t) e^{i\omega t} dt = \int_{\tau_0}^{\infty} p_{\text{diffr}}(t) e^{i\omega t} dt. \quad (\text{A5})$$

Using Eq. (A2), we can change the integration variable t to η ,

$$dt = \frac{r r_0}{c} \sinh \eta \frac{d\eta}{R(\eta)}, \quad (\text{A6})$$

where

$$R(\eta) = ct = [r^2 + r_0^2 + 2rr_0 \cos \eta + (z - z_0)^2]^{1/2}. \quad (\text{A7})$$

Substituting Eqs. (A6) and (A7) into Eq. (A5), we obtain

$$p_{\text{dif}}(f) = \frac{\rho \sin \nu\pi}{4\pi\theta_w} \int_0^{\infty} \frac{e^{ikR(\eta)}}{R(\eta)} \sum_{\pm} \frac{\cos \nu\pi - \cosh \nu\eta \cos[\nu(\theta \pm \theta_0)]}{\cos \nu\eta - \cos[\nu(\theta \pm \theta_0)] - \sin^2 \nu\eta \sin^2[\nu(\theta \pm \theta_0)]} d\eta. \quad (\text{A8})$$

Using the trigonometric identity $\sin^2 x = 1 - \cos^2 x$, the denominator of Eq. (A8) becomes

$$\begin{aligned} \text{Denominator} &= (\cosh \nu\eta - 1)^2 + 2(\cosh \nu\eta - 1)(1 \\ &\quad - \cos \nu\vartheta \cos \nu\pi) + (\cos \nu\pi - \cos \nu\vartheta)^2, \end{aligned} \quad (\text{A9})$$

where $\vartheta = \theta \pm \theta_0$. By using the relation $S = -\rho/4\pi$, we obtain

$$p_{\text{dif}}(f) = -\frac{S \sin \nu\pi}{2\theta_w} \int_{-\infty}^{\infty} \frac{e^{ikR(\eta)}}{R(\eta)} D_{\eta} d\eta, \quad (\text{A10})$$

where

$$D_{\eta} = \frac{\cos \nu\pi - \cosh \nu\eta \cos[\nu\vartheta]}{(\cosh \nu\eta - 1)^2 + 2(\cosh \nu\eta - 1)(1 - \cos \nu\vartheta \cos \nu\pi) + (\cos \nu\pi - \cos \nu\vartheta)^2}. \quad (\text{A11})$$

A factor of 2 included in Eq. (A10) results from extending the integral from 0 to $-\infty$ by recognizing that the integrand is an even function of η . Equation (A10), along with Eqs. (A7) and (A11), is exactly the same as Eq. (2a)–(2c).

Abramowitz, M., and Stegun, I. A., eds. (1974). *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables* (Dover, New York).

Biot, M. A., and Tolstoy, I. (1957). "Formulation of wave propagation in infinite media by normal coordinates with an application to diffraction," *J. Acoust. Soc. Am.* **29**, 381–391.

Calamia, P. T. and Svensson, U. P. (2007). "Fast Time-Domain Edge-Diffraction Calculations for Interactive Acoustic Simulations," *EURASIP Journal on Advances in Signal Processing*, Vol. **2007**, pp. 10.

Carslaw, H. S. (1899). "Some multiform solutions of the partial differential equations of physics and mathematics and their applications," *Proc. London Math. Soc.* **30**, 121–163.

Carslaw, H. S. (1920). "Diffraction of waves by a wedge of any angle," *Proc. London Math. Soc.* **18**, 1291–1306.

Chu, D. (1989). "Impulse response of density contrast wedge using normal coordinates," *J. Acoust. Soc. Am.* **86**, 1883–1896.

Chu, D., and Stanton, T. K. (1998). "Application of pulse compression techniques to broadband acoustic scattering by live individual zooplankton," *J. Acoust. Soc. Am.* **104**, 39–55.

Clay, C. S., Chu, D., and Li, S. (1993). "Specular reflections of transient pressures from finite width plane facets," *J. Acoust. Soc. Am.* **94**, 2279–2286.

Daneshvar, M. R., and Clay, C. S. (1987). "Imaging of rough surfaces for impulsive and continuously radiating sources," *J. Acoust. Soc. Am.* **82**, 360–369.

Davis, A. M. J., and Scharstein, R. W. (1997). "The complete extension of the Biot–Tolstoy solution to the density contrast wedge with sample calculations," *J. Acoust. Soc. Am.* **101**, 1821–1835.

Duan, D.-W., Rahmat-Samii, Y., and Mahon, J. P. (1991). "Scattering from a circular disk: A comparative study of PTD and GTD techniques," *Proc. IEEE* **79**, 1472–1480.

Feuillade, C., Chu, D., and Clay, C. S. (2004). "Space-time variations of the acoustic field scattered from a penetrable isovelocity wedge," *J. Acoust. Soc. Am.* **116**, 777–789.

Friedlander, F. G. (1958). *Sound Pulses* (Cambridge University Press, New York), pp. 108–144.

Gradshteyn, I. S., and Ryzhik, I. M. (1980). *Table of Integrals, Series, and Products* (Academic, Orlando).

Jones, D. S. (1953). "Diffraction by a thick semi-infinite plate," *Proc. R. Soc. London* **217**, 153–175.

Karp, S. N., and Keller, J. B. (1960). "Multiple diffraction by an aperture in a hard screen," *Opt. Acta* **8**, 61–72.

Keiffer, R. S., Novarini, J. C., and Norton, G. V. (1994). "The impulse response of an aperture: Numerical calculations within the framework of the wedge assemblage method," *J. Acoust. Soc. Am.* **95**, 3–12.

Keller, J. B. (1957). "Diffraction by an aperture," *J. Appl. Phys.* **28**, 426–444.

Keller, J. B. (1962). "Geometric theory of diffraction," *J. Appl. Phys.* **52**, 116–130.

Keller, J. B., Lewis, R. M., and Seckler, B. D. (1956). "Asymptotic solution of some diffraction problems," *Commun. Pure Appl. Math.* **9**, 207–265.

Kinney, V. A., Clay, C. S., and Sandness, G. A. (1983). "Scattering from a corrugated surface: Comparison between experiment, Helmholtz-

Kirchhoff theory, and the facet ensemble method," *J. Acoust. Soc. Am.* **73**, 183–194.

Kristensson, G., and Waterman, P. C. (1982). "The T matrix for acoustic and electromagnetic scattering by circular disks," *J. Acoust. Soc. Am.* **72**, 1612–1625.

Li, L. W., Kooi, P. S., Qiu, Y. L., Yeo, T. S., and Leong, M. S. (1998). "Analysis of electromagnetic scattering of conducting circular disk using a hybrid method," *Progress in Electromagnetics Research* **20**, 101–123. doi: 10.2528/PIER97111200.

Li, S., Chu, D., and Clay, C. S. (1994). "Time domain reflections and diffractions from facet-wedge constructions: Acoustic experiments including double diffractions," *J. Acoust. Soc. Am.* **96**, 3715–2720.

MacDonald, H. M. (1902). *Electric Waves* (Cambridge University Press, London), pp. 187–198.

MacDonald, H. M. (1915). "A class of diffraction problems," *Proc. London Math. Soc.* **14**, 410–427.

Medwin, H., Childs, E., and Jebsen, G. M. (1982). "Impulse studies of double diffraction: Adiscrete Huygens interpretation," *J. Acoust. Soc. Am.* **72**, 1005–1013.

Medwin, H., and Clay, C. S. (1998). *Fundamentals of Acoustic Oceanography* (Academic, Boston), Chaps. 11 and 12.

Menounou, P., Busch-Vishniac, I. J., and Blackstock, D. T. (2000). "Directive line source model: A new model for sound diffraction by half planes and wedges," *J. Acoust. Soc. Am.* **107**, 2973–2986.

Morse, P. M., and Ingard, K. U. (1968). *Theoretical Acoustics* (McGraw-Hill, New York), pp. 400–466.

Pierce, A. D. (1974). "Diffraction of sound around corners and over wide barriers," *J. Acoust. Soc. Am.* **55**, 941–955.

Pierce, A. D. (1981). *Acoustics: An Introduction to its Physical Principles and Applications* (McGraw-Hill, New York), pp. 479–501.

Rahmat-Samii, Y. (1988). "Reflector antennas," in *Antennas Handbook*, edited by Y. T. Lo and S. W. Lee (Rahmat-Samii, a Nostrand-Reinhold Company, New York), Chap. 15.

Sleator, F. B. (1969). The Disc, Ch. 14 in *Electromagnetic and Acoustic Scattering by Simple Shapes*, edited by J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi (Elsevier, New York).

Sommerfeld, A. (1954). *Optics: Lectures on Theoretical Physics* (Academic, New York), pp. 204–205.

Stanton, T. K., and Chu, D. (2004). "On the acoustic diffraction by the edges of benthic shells," *J. Acoust. Soc. Am.* **116**, 239–244.

Stanton, T. K., Chu, D., and Norton, G. V. (2007). "Acoustic diffraction by deformed edges of finite length: Theory and experiment," *J. Acoust. Soc. Am.* **122**, 3167.

Svensson, U. P., Fred, R. I., and Vanderkooy, J. (1999). "An analytic secondary source model of edge diffraction impulse responses," *J. Acoust. Soc. Am.* **106**, 2331–2344.

Svensson, U.P. and Calamia, P.T. (2006). "Edge-diffraction impulse responses near specular- and shadow-zone boundaries," *Acta. Acust. Acust.* **92**, 501–512.

Tolstoy, I. (1989a). "Diffraction by a hard truncated wedge and a strip," *IEEE J. Ocean. Eng.* **14**, 4–16.

Tolstoy, I. (1989b). "Exact, explicit solutions for diffraction by hard sound barriers and seamounts," *J. Acoust. Soc. Am.* **85**, 661–669.

Tolstoy, I., and Clay, C. S. (1987). *Ocean Acoustics: Theory and experiment in underwater sound*. (American Institute of Physics, New York), Chap. 3.

Ufimtsev, P. Y. (1962). "Method of edge waves in the physical theory of diffraction," *Izd-Vo Sovyetskoye Radio*, translated by US Air Force Foreign Tech. Div., 1154 pp.

On appropriate equivalent monopole sets for rigid body scattering problems

Yves J. R. Gounot^{a)}

Acoustics and Vibration Laboratory, Department of Mechanical Engineering/COPPE/UFRJ, Universidade Federal do Rio de Janeiro, C.P. 68503, Rio de Janeiro, 21941-972, Brazil

Ricardo E. Musafir

Acoustics and Vibration Laboratory, Department of Mechanical Engineering/COPPE/UFRJ and Water Resources and Environmental Engineering Department/EP/UFRJ, Universidade Federal do Rio de Janeiro, C.P. 68503, Rio de Janeiro, 21941-972, Brazil

(Received 21 December 2006; revised 10 August 2007; accepted 23 September 2007)

The present paper investigates properties of the solutions obtained with the equivalent sources method in scattering problems with the aim of identifying suitable monopole arrangements. Simple geometry scatterers—parallelepipeds with different aspect ratios—are considered and easy-to-implement source supports are tested: linear, circular, elliptical, and a “double linear” one. It was found that the supports providing best solutions differ according to the body geometry and the incidence angle of the impinging wave. Moreover, in the situations in which the other supports fail, it is shown that the double linear one provides satisfactory solutions with a minimum number of monopoles. Rules that furnish appropriate numbers of sources to use as well as their positioning are given for the different cases. A simple procedure based on these rules is proposed for scatterers with a more complex geometry, guaranteeing, still with a low number of monopoles, solutions with satisfactory accuracy. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2799504]

PACS number(s): 43.20.Fn, 43.20.Ei [LLT]

Pages: 3195–3205

I. INTRODUCTION

The equivalent sources method (ESM) is a simple numerical technique for solving radiation or scattering problems which basic principle lies in substituting the real body by a set of virtual sources that should reproduce the original problem boundary condition. Compared to classical methods like the boundary elements (BEMs) or finite elements (FEMs) one, commonly used for exterior and interior problems, respectively, ESM presents advantages of different natures. First, for being more direct, ESM represents a lower computational cost than BEM and FEM. Second, since it permits satisfactory reconstructions of radiated or scattered fields using only a few sources, the equivalent sources method constitutes an efficient and economical simulation technique, being quite attractive for practical applications, like active noise control. In order to find an appropriate source arrangement—that yields an acceptable value for the mean error of the simulated velocity on the boundary—one has to choose the type of sources, as well as their number and position. As for the type, most common are expansions in spherical wave functions at one or various points,^{1–3} or arrays of monopoles located at different points.^{4–14} In either case, the essential issue remains how to pick an appropriate number of sources and how to choose their position. Although in the case of expansions, the issue of the source positioning can be considered solved—see, for example, Ochmann’s *multi-point multipole method*²—choosing the po-

sition of the equivalent sources still constitutes an important difficulty when sets of monopoles are used since, in this case, no general rules exist. The most frequently used source supports are internal surfaces with the same shape as the body itself, since, as shown by Jeans and Mathews,⁴ the fact that the distance between the support and the boundary is constant yields a good numerical stability. For example, Song *et al.*⁵ and Fahline and Koopmann⁶ treat the bi-dimensional cases of an infinite cylinder using line sources located on an inner cylinder-shaped surface and Koopmann *et al.*⁷ solve simple three-dimensional (3D) problems by means of sources positioned on spherical and cubic shaped supports. However, for all these cases, a recurrent question is what should be the optimal distance between the support and the boundary, and the answers given differ. While some authors have shown that the solution accuracy increases with the separating distance⁷ or with the compactness of the source set,⁹ others suggest that the ratio between support and body characteristic lengths should be about 0.5.^{4,5,8} Koopmann *et al.*,⁷ on the other hand, recommend limiting acceptable values for this ratio, between 0.05 and 0.5. Another approach consists of positioning the sources on a grid placed in the body inner space; the mesh can be bi- or three dimensional, as the ones employed by Kropp and Svensson¹¹ and Herrin and Seybert,¹² respectively. Pavic¹³ also used a grid, but with a different approach: the grid points represent possible positions for the sources and an iterative process is used in order to select, one by one, appropriate locations. As for the issue of the number of sources, published results are even less conclusive. When the *Full Field Equations*² is used as minimization technique, the solution precision increases with

^{a)}Author to whom correspondence should be addressed. Electronic mail: ygounot@mecanica.ufrj.br

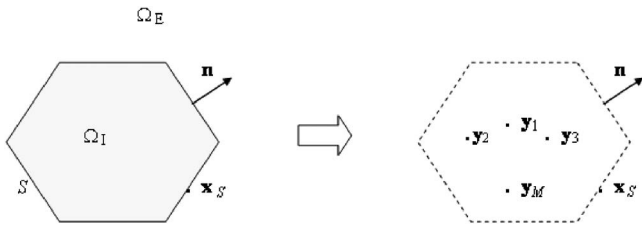


FIG. 1. Principle of the equivalent source method: representation of the “real” body and its substitution by “virtual” equivalent sources.

the number of sources, while with the *Least Square Method* (the more widespread choice), the evolution of the solution quality presents two distinct phases: as shown by Ochmann² and by Johnson *et al.*,¹⁴ after increasing with the number of monopoles, the solution generally suffers a significant degradation when this number exceeds a certain value and, moreover, both this value and the precision loss depend strongly on the support geometry, as shown by Gounot *et al.*⁹ Stepanishen and Ramakrishna¹⁰ have shown that, for each frequency, there is a band of acceptable values for the number of sources, this bandwidth becoming smaller as the frequency increases. In spite of all these results and recommendations, no general rules that could constitute working guidelines for the user are available.

Unlike the case of the radiation problem, for which the solution depends on the distribution of the vibration velocity over the body surface, which can be arbitrary, scattering from relatively simple geometry bodies will, frequently, present a high degree of symmetry. In these cases, it is reasonable to assume the existence of simple *a priori* appropriate source arrangements. The present paper aims to answer this query, by pointing out, for simple geometry scatterers, some easy-to-implement efficient source supports, and to establish rules that provide, depending on the case, the appropriate number of sources and their positioning, what should allow a straightforward and safer use of the ESM.

II. THEORETICAL BACKGROUND

The scattering problem due to the impinging of an acoustic wave p_{inc} on a rigid body surrounded by a region Ω_E characterized by a uniform mean density ρ_0 , can be described, in the frequency domain, as follows:¹⁵ the complex scattered pressure $p_{sc}(\mathbf{x})$ has to satisfy, in Ω_E , the Neumann boundary value problem described by the Helmholtz equation, the Sommerfeld radiation condition and the prescription, on the body boundary S , of the normal velocity expressed by $\bar{u}_n = -v_n^{inc}$, where v_n^{inc} is the normal velocity that would be generated by the incident wave in the absence of the body. ESM substitutes the real body by a set of M sources at points \mathbf{y}_m , located in Ω_I , the region previously filled by the body volume (see Fig. 1); the sound field due to these sources is expressed in terms of their unknown complex amplitudes (the source strengths A_m) and of a function G —or of a family of functions, if an expansion in spherical wave functions¹⁶ is used—describing their radiation. The scattered pressure and velocity at \mathbf{x} can be written, if only monopoles are used, as

$$p_{sc}(\mathbf{x}) = \sum_{m=1}^M A_m G(\mathbf{x}, \mathbf{y}_m) \quad (1)$$

$$v_n^{sc}(\mathbf{x}) = \frac{-1}{i\omega\rho_0} \sum_{m=1}^M A_m \frac{\partial G(\mathbf{x}, \mathbf{y}_m)}{\partial n}, \quad (2)$$

where $\omega = kc_0$ is the angular frequency, k is the wave number and c_0 is the speed of sound; G is the free-space Green function, $e^{ikr}/4\pi r$, where $r = |\mathbf{x} - \mathbf{y}_m|$.

In general, the equivalent source set does not satisfy exactly the boundary condition, a local boundary error, given by the difference between the velocity due to the equivalent source system and the theoretical values, being generated. The solution $\{A_m\}$ is obtained by minimizing the global velocity error on S . In the present work, the minimization technique used is the least squares method—that consists in minimizing the sum (over the N boundary nodes) of the squares of the local errors—and the solution quality will be evaluated through the normalized boundary error e_{BC} , given by

$$e_{BC} = \frac{\sum_{i=1}^N |v_n^{sc}(\mathbf{x}_i) - \bar{u}_n(\mathbf{x}_i)|^2}{\sum_{i=1}^N |\bar{u}_n(\mathbf{x}_i)|^2}. \quad (3)$$

In previous work⁹ it has been shown that the quality of the solution depends strongly, for a given source support, on the number of monopoles employed and on their relative positioning, both with respect to the other sources and to the boundary. Characteristics of the equivalent sources corresponding to good solutions have been pointed out: source phases are such that the monopoles are approximately disposed like a set of dipoles, while their amplitudes may vary significantly. For a relatively small number of monopoles, this behavior, responsible for a high degree of cancellation, permits a good reconstitution of the boundary condition, while, when the number of sources exceeds a limiting value, there is a significant increase in the mean value of the monopole magnitudes \hat{A} , which leads to the degradation of the solution quality. Consequently, selecting the range of M corresponding to good solutions is equivalent to selecting a range of acceptable values for \hat{A} . Since, in general, it is more useful to have information on the appropriate number of sources than on its mean magnitude, the former parameter (M) will be emphasized in the present investigation. As for the quality criterion, numerous trials¹⁷ relative to the case of a cubic scatterer (with side equal to the wavelength, λ) permitted to determine a limiting upper value of the acceptable e_{BC} as 0.5. This value has been chosen for corresponding to a good pressure field reconstitution in a region close to the boundary, 95% of the control points located on a circle with radius $R = 2\lambda$ presenting an error lower than 1 dB.

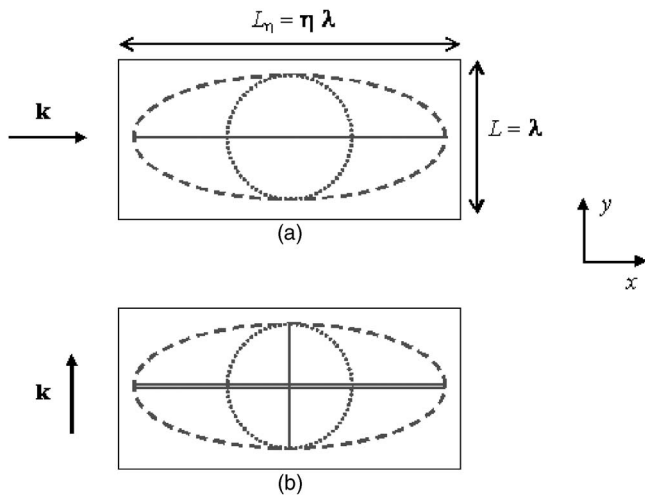


FIG. 2. Representation of the scatterer and the source supports used for the cases $\Phi_{\text{inc}}=0$ (a) and $\Phi_{\text{inc}}=\pi/2$ (b).

III. NUMERICAL EXPERIMENT AND RESULTS

A. General description

The results presented below are relative to the scattering of a plane wave by a rigid parallelepiped-shaped scatterer with dimension $(\eta \times 1 \times 1)\lambda$, its largest dimension, referred as L_η , being parallel to the x axis. Different cases are investigated, according to the scatterer aspect ratio η ($\eta=1, 2, 3$, and 4) and also to the incidence angle considered, $\Phi_{\text{inc}}=0$ and $\Phi_{\text{inc}}=\pi/2$ (see Fig. 2).

Different source supports of easy computational implementation are used: linear (L), circular (C) and elliptical (E); a fourth one, the “double linear” support (LL), will also be employed for $\Phi_{\text{inc}}=\pi/2$. All are located in the $z=0$ plane, which is the scattered field symmetry plane. Their centers are coincident with the body geometric center and their sizes are obtained by multiplying the maximum size acceptable (such that all sources are strictly inside the body) by a *reduction factor* a , $0 < a < 1$. The ratio between the axis of the elliptical support is given by η , so that proportions between the support and the body shape are maintained. Orientation for L and for the wave vector \mathbf{k} is given, in each case, by the angle made with the axis parallel to the larger side of the structure, Φ and Φ_{inc} , respectively. For the particular case $\eta=1$, the cube—for which naturally only the linear and circular supports will be considered—the investigation is focused on the effect of varying the frequency, those investigated corresponding to $kL=2\pi n$, with $n=0.5, 1, 1.5$, and 2. The source configurations are referred as L_M, C_M , and E_M , where M denotes the total number of monopoles regularly positioned on the support.

For each case investigated (i.e., for a given frequency, support type and size), 24 solutions are computed, corresponding to $M=2-25$ monopoles, the solution quality being evaluated through the corresponding e_{BC} .

For rigid compact bodies, the scattered velocity can be reasonably approximated by the first two terms in the multipole expansion of v_n , i.e., by a monopole and a dipole component with axis parallel to \mathbf{k} .¹⁵ Even though, as the compactness diminishes, higher order terms appear, while the

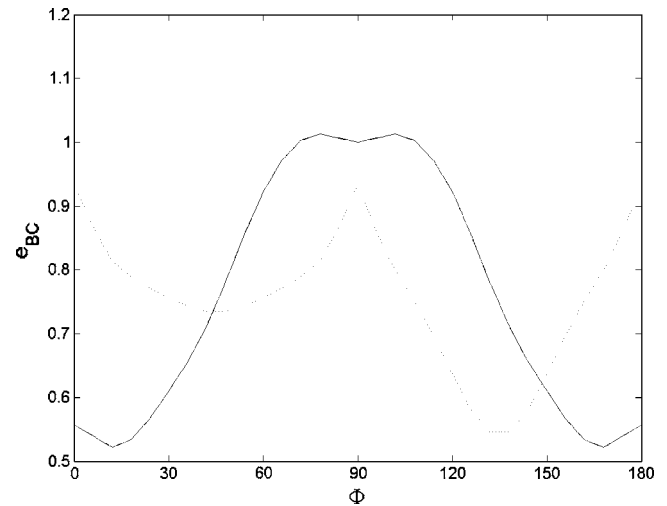


FIG. 3. Influence of the orientation of the linear support (Φ) on the boundary error for the cube and $kL=2\pi$, obtained with L_5 for two incidences of the plane wave: $\Phi_{\text{inc}}=0$ (—) and $\Phi_{\text{inc}}=135^\circ$ (----).

relative importance of the monopole term decreases, a strong dipolar component is maintained, as illustrated in Ref. 9. The better way to simulate such a dipolar field with a linear configuration of monopoles seems to be when the in-line monopoles are parallel to the wave vector \mathbf{k} . In order to illustrate the influence of the orientation of the support and to confirm this reasoning, it was considered (for $\eta=1$ and $kL=2\pi$) the configuration L_5 and, for two different angles of incidence for \mathbf{k} , were computed the solutions for Φ , the angle between the support and the x axis, varying from 0 to 180° (see Fig. 3). The results show that the best solutions (i.e., the lowest e_{BC} values) are always obtained when the axis of the linear support is parallel to \mathbf{k} , as expected. In what follows, the linear support L will always be taken parallel to the wave vector of the impinging plane wave.

B. Effect of the frequency

Since the solution quality can vary significantly with the frequency, the scatterer dimension, the number and the position of the sources, a given source arrangement providing a good solution in a situation can yield a poor quality response in another situation. The present section investigates, for the cube, which supports could guarantee a satisfactory solution independently of the frequency, in the range tested.

Figure 4 shows, for $kL=2\pi n$, with $n=0.5, 1, 1.5$, and 2, the scattered and total pressure fields on the control circle (of radius $R=2\lambda$), computed with a BEM code.¹⁸ It can be observed that, as kL increases, the scattered pressure field becomes more directional and its peaks are higher. This implies that, when using monopoles to reconstruct the field, the number of sources required should also increase with kL .

For each case considered, the number of monopoles used and the boundary error relative to the best solution over the computed 24 ones are given in Table I. In order to evaluate the efficiency of these monopole arrangements, comparison is made with the solution obtained using expansions according to a procedure proposed by Ochmann.^{1,2} This procedure consists of dividing the body in a set of cubic or

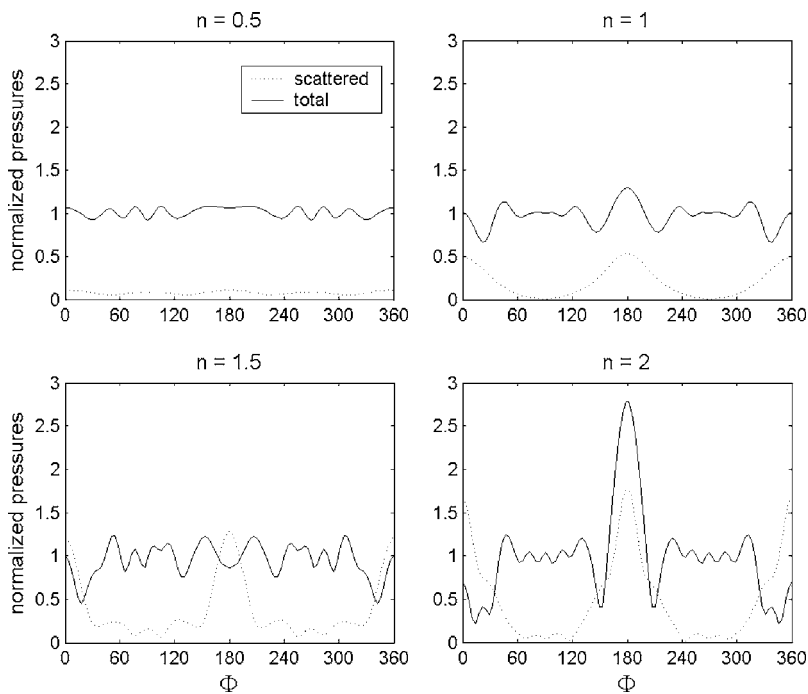


FIG. 4. Scattered and total pressure magnitude (normalized by $|p_{inc}|$) on the control circle due to the scattering by a cube for $kL=2\pi n$, with $n=0.5, 1, 1.5$ and 2 .

spherical substructures, in whose centers are placed the expansion points. In the present section, since the scatterer is a cube, no division is required, and a single expansion with order up to α , referred as X^α , is used in the geometric center of the cube. The e_{BC} values corresponding to $\alpha=1, 2, 3$, and 4 are given in the last four lines of Table I. It is worth underlining the fact that X^α contains $(\alpha+1)^2$ functions, i.e., an expansion referred to as X^1, X^2, X^3 , or X^4 corresponds to 4, 9, 16 or 25 “sources” with different orders, respectively: monopole and dipoles for $\alpha=1$, monopole, dipoles, and quadrupoles for $\alpha=2$, etc.

Results show that, for each case, the best solution obtained with the linear support is always better than the one provided by the circular one and that, when e_{BC} values are comparable, the number of monopoles required with L is always significantly smaller than with C. Results also show that the higher the frequency, the more pronounced the discrepancy between the efficiency of the two supports. In other words, for the reconstitution of such directional fields, the linear arrangement is more efficient and economical than the circular one.

Comparing the results obtained with the linear arrangements and those obtained with the expansions, it can be observed that, since the solution accuracy increases with the expansion order (α), there is always a value of α for which X^α provides a better result than the monopole sets do. However, this option can be very costly, since, as shown in these cases for $kL \sim 2\pi$, the first expansion more efficient than L_4 (only four monopoles) is X^3 , which corresponds to 16 sources with different orders and, for $kL \sim 3\pi$, the first expansion more efficient than L_7 (seven monopoles) is X^4 , which involves 25 sources. Yet, as the frequency increases (for $kL \geq 12$), it is worth noting that the solution obtained with only ten monopoles (corresponding to $e_{BC}=0.35$) is better than the one corresponding to the higher order expansion tested, X^4 . This means that, when the body dimension exceeds 2λ , the “single expansion” efficiency decreases with L, while the linear configuration still provides, with only a few monopoles, good quality solutions. In order to understand the reasons for the linear support efficiency, it is shown, in Fig. 5 (as an illustration), the source strengths (magnitude and phase) corresponding to the L_{10} solution for $n=2$. The

TABLE I. The e_{BC} and M values corresponding to the “best” solution, for each frequency, support type, and size; “Ochmann expansions” e_{BC} results. Cube case.

Freq		n=0.5			n=1			n=1.5			n=2		
C	a	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
	e_{BC}^*	0.55	0.55	0.56	0.44	0.44	0.54	0.55	0.78	0.78	0.55	0.72	0.96
	M^*	10	10	10	14	17	13	18	4	16	25	25	24
L	a	0.50	0.49	0.51	0.45	0.52	0.60	0.33	0.35	0.46	0.33	0.33	0.35
	e_{BC}^*	0.50	0.49	0.51	0.45	0.52	0.60	0.33	0.35	0.46	0.33	0.33	0.35
	M^*	6	6	5	4	3	3	7	7	5	20	11	10
Expansion	X^1	0.66			0.46			0.80			0.72		
	X^2	0.00			0.46			0.42			0.72		
	X^3	0.00			0.27			0.42			0.46		
	X^4	0.00			0.27			0.27			0.46		

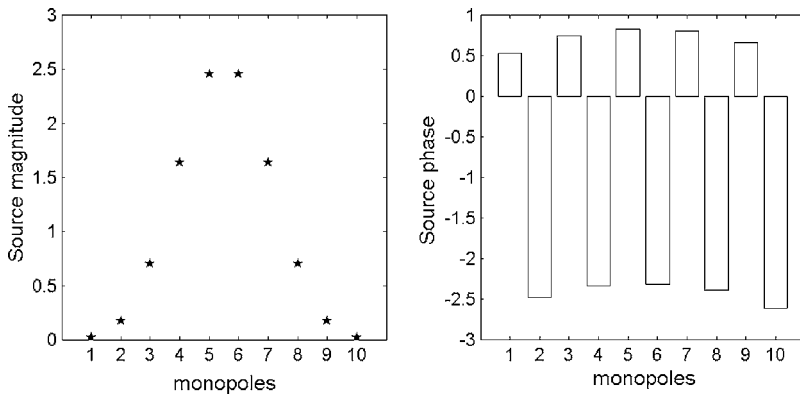


FIG. 5. The L_{10} source strengths magnitude (normalized by the mean value, \hat{A}) and phase, for $n=2$.

typical source arrangement described in a previous work⁹—the monopoles forming, roughly, a string of local “dipoles”—clearly shows up. When the whole source set is considered, a high degree of cancellation is obtained between the in-line monopoles, the emission occurring thus mainly along the support axis.

For each frequency, support size (given by the reduction factor, a) and number of monopoles considered, the distance between two adjacent sources, d_s , can be expressed, in a nondimensional form, for the linear and circular case, respectively, by

$$kd_s = \frac{2\pi na}{M-1}, \quad (4a)$$

$$kd_s = 2\pi na \sin\left(\frac{\pi}{M}\right). \quad (4b)$$

Results show that, compared with the kd_s range explored (from 0.01 to 10), the values of kd_s corresponding to acceptable solutions ($e_{BC} \leq 0.5$) are always low, with a mean value of about 0.6. Furthermore, if one considers only the best solution (out of the 24 ones) for each frequency, it is found that the corresponding curves “ $kd_s \times a$ ” can be reasonably well approximated by lines with roughly the same angular coefficient (with mean value around 1.5), both for the linear and circular supports (see Fig. 6). Insertion of this value in equations (4) yields a rule relating, for a given support, both the cube edge and the frequency to the ‘optimal’ number of monopoles, which is given, in the linear and circular cases, respectively by

$$M \approx 1 + \frac{kL}{1.5}, \quad (5a)$$

$$M \approx \frac{\pi}{1.5}kL. \quad (5b)$$

It was verified that Eqs. (5) provide, if not exactly the number of monopoles associated to the best solution, numbers that always correspond to a solution fulfilling the quality criterion adopted. As an illustration, the rule is here applied to the case of a higher frequency, corresponding to $n=3$, i.e., for $kL \sim 19$. For the linear support, the number of monopoles suggested [Eq. (5a)] is, in this case, $M=13$. The solution obtained with the L_{13} configuration shows a good precision ($e_{BC}=0.41$), while the best solution provided by expansions— X^4 , which involves twice the number of sources—scarcely corresponds to an acceptable error value ($e_{BC}=0.73$).

C. Aspect ratio effect

It is convenient to investigate which source supports permit to obtain, with the minimum number of monopoles, acceptable solutions when the scatterer presents an aspect ratio different than unity. Figure 7 shows the pressure fields (scattered and total) on the control circle (obtained with the BEM code), corresponding to two limit values of the angle of incidence $\Phi_{inc}=0$ and $\Phi_{inc}=\pi/2$, i.e., to the two types of normal incidence.

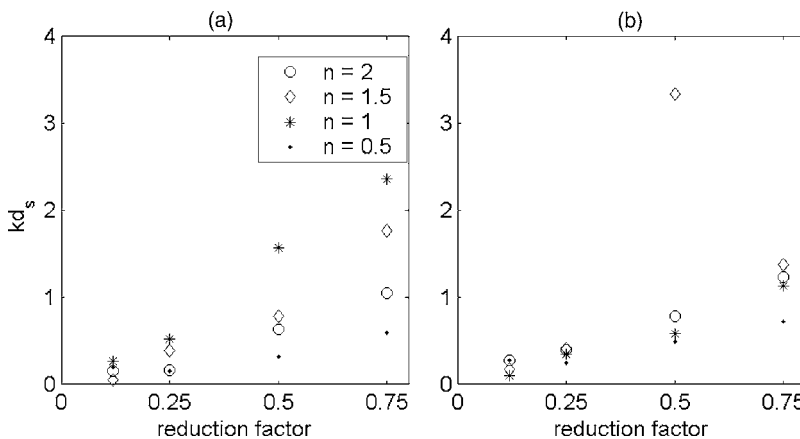


FIG. 6. The kd_s values corresponding to the best solution over the 24 tested, with the linear (a) and circular (b) supports for different reduction factors and frequencies ($kL=2\pi n$).

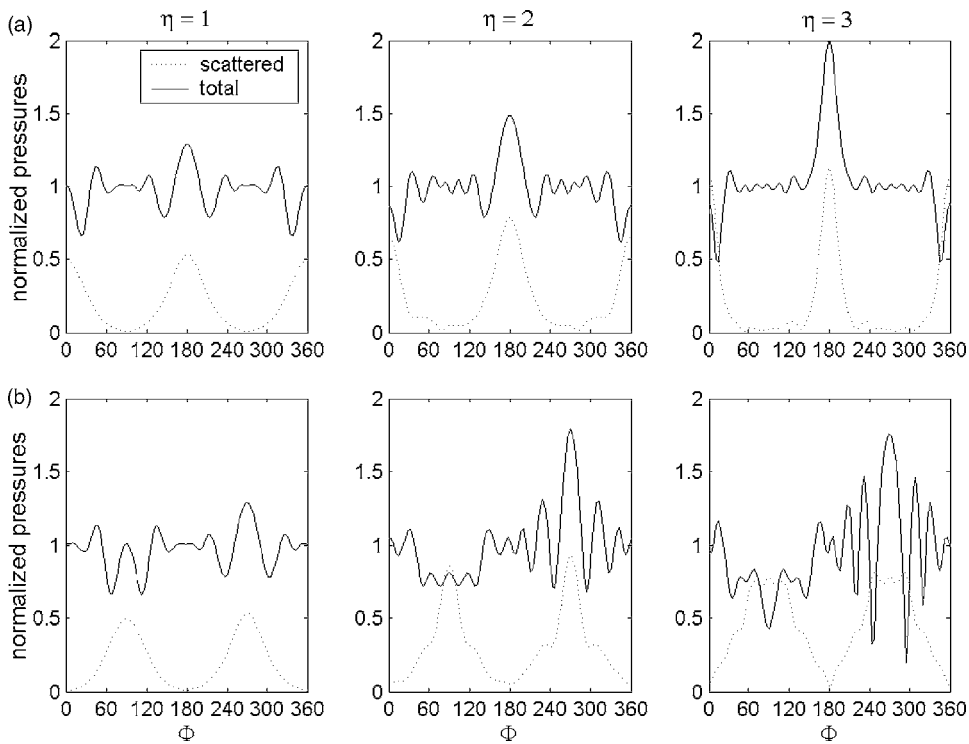


FIG. 7. Scattered and total pressure magnitude (normalized by $|p_{\text{inc}}|$) on the control circle due to the scattering by a parallelepiped ($\eta=1, 2,$ and 3) for a) $\Phi_{\text{inc}}=0$ and b) $\Phi_{\text{inc}}=\pi/2$.

For parallelepiped-shaped bodies and a normal incidence, since $\bar{u}_n = -v_n^{\text{inc}}$ is nonzero only on the two faces normal to \mathbf{k} , the scattered fields are still basically dipolar. However, as η increases, the main difference between the two incidence cases considered is that, while for $\Phi_{\text{inc}}=0$ the scattered field becomes more and more concentrated along the \mathbf{k} direction, when $\Phi_{\text{inc}}=\pi/2$, the affected region increases with η (see Fig. 7).

Figure 8 shows, for $\Phi_{\text{inc}}=0$ and for the different values of η , the evolution of the boundary error (restricted to $e_{\text{BC}} \leq 1$) when, for a given support (linear, circular and elliptical), the number of sources increases from 2 to 25. In order to also capture the influence of the “compactness” of the source sets, three support sizes, corresponding to $a=0.25, 0.5,$ and 0.75 , were investigated. Results relative to the circular support show that, while C yields good quality solu-

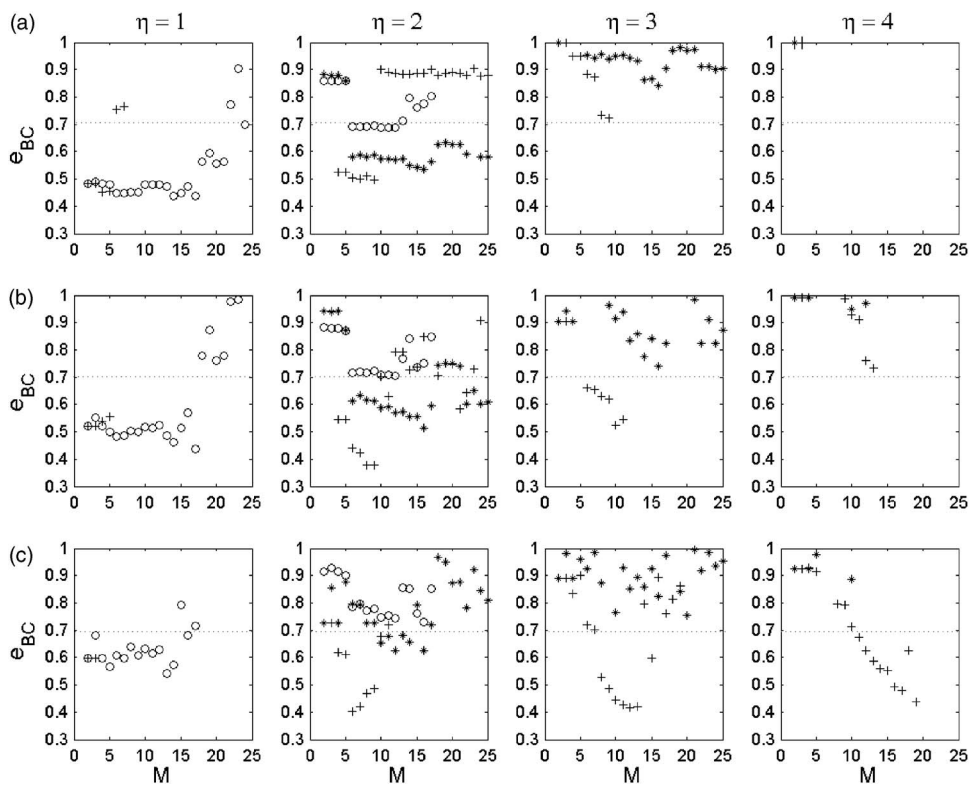


FIG. 8. Evolution of the boundary error with the number of sources for $\Phi_{\text{inc}}=0, \eta=1, 2, 3,$ and 4 , with L (+), C (○) and E (◆); $a=0.25$ (a), 0.5 (b) and 0.75 (c).

TABLE II. The e_{BC} and M values corresponding to the “best” solution, for each aspect ratio, support type, and size; “Ochmann expansions” e_{BC} results. Case \mathbf{k} parallel to L_η

		$\eta=1$			$\eta=2$			$\eta=3$			$\eta=4$		
	a	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
C	e_{BC}	0.44	0.44	0.54	0.69	0.7	0.73	1	1.01	1	1.08	1.05	1.01
	M	14	17	13	11	12	16	6	6	2	2	2	2
E	e_{BC}	0.44	0.44	0.54	0.54	0.50	0.63	0.84	0.74	0.75	1	0.95	0.89
	M	14	17	13	16	16	16	16	16	20	2	10	10
L	e_{BC}	0.45	0.52	0.6	0.5	0.38	0.4	0.72	0.53	0.42	1	0.73	0.44
	M	4	3	3	9	8	6	9	10	12	2	13	19
η -expans.	X^1		0.46			0.40			0.40			0.40	
	X^2		0.46			0.31			0.31			0.30	
	X^3		0.27			0.27			0.27			0.26	
	X^4		0.27			0.21			0.19			0.18	

tions for $\eta=1$ (the cube), it only provides solutions “just about acceptable” ($e_{BC} \sim 0.7$) for $\eta=2$ and is totally inappropriate for higher aspect ratios. For $\eta=2$, while the best solutions are reached with the linear support with a limited value for M , the elliptical one (with $a \leq 0.5$) yields good solutions for all M . When $\eta > 2$, the only appropriate support is the linear one, provided its size increases with η , what makes it able to deal with a higher number of monopoles and allows a better covering of the body boundary. Actually, the pressure field is so highly concentrated on the x axis that, with a relatively small number of monopoles, it only can be simulated efficiently when these sources are located on this axis. Table II shows, for each incidence and each support, the values of e_{BC} and M corresponding to the best solution over the 24 tested.

It was verified that the rule furnishing an appropriate number of monopoles for the linear support in the case of the cube [Eq. (5a)] is still valid for scatterers with other aspect ratio values, provided L is substituted by L_η . Table II also presents, for comparison sake, results obtained with the “expansion technique,” which, in these cases, requires η expansions X^α , i.e., $\eta(\alpha+1)^2$ sources with different orders (see Sec. III B). Since each expansion operates in a quite limited region (a cube with side λ what corresponds roughly to a spherical-shaped structure with radius $\lambda/2$), the expansion technique is very efficient, and the solution accuracy increases with the order of the expansions. Nevertheless, it can be observed that source sets with only 6, 12 and 19 monopoles provide solutions with a precision comparable to those

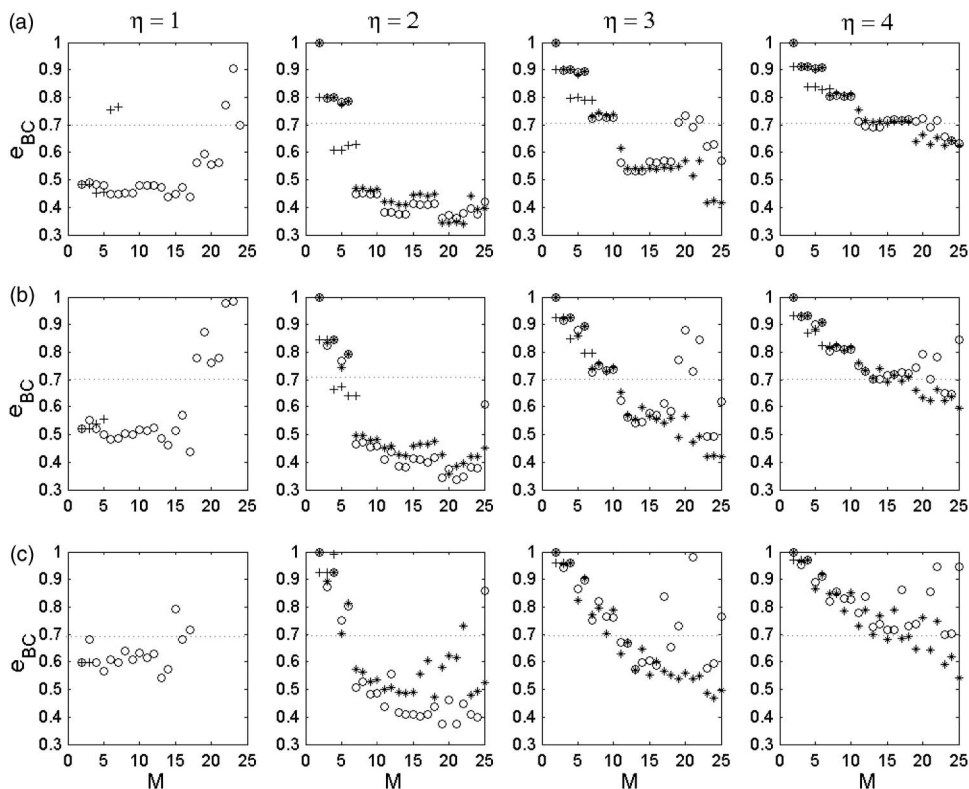


FIG. 9. Evolution of the boundary error with the number of sources for $\Phi_{inc} = \pi/2$, $\eta=1, 2, 3$, and 4, with L (+), C (O) and E (◆); $a=0.25$ (a), 0.5 (b) and 0.75 (c).

TABLE III. The e_{BC} and M values corresponding to the “best” solution, for each aspect ratio, support type, and size; “Ochmann expansions” e_{BC} results. Case \mathbf{k} normal to L_η

		$\eta=1$			$\eta=2$			$\eta=3$			$\eta=4$		
C	a	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
	e_{BC}	0.44	0.44	0.54	0.36	0.34	0.37	0.53	0.49	0.57	0.63	0.65	0.7
E	M	14	17	13	21	21	19	13	24	13	25	24	23
	e_{BC}	0.44	0.44	0.54	0.34	0.36	0.47	0.42	0.42	0.47	0.63	0.6	0.54
L	M	14	17	13	22	20	18	25	23	24	23	25	25
	e_{BC}	0.45	0.52	0.6	0.61	0.64	0.93	0.79	0.8	0.96	0.83	0.82	0.97
LL	M	4	3	3	4	6	3	6	6	3	6	7	3
	e_{BC}	0.45	0.44	0.44	0.41	0.42	0.43	0.55	0.43	0.38	0.68	0.65	0.37
η -expans.	M	6	6	6	10	6	8	10	10	10	10	8	10
	X^1		0.46			0.43			0.40			0.38	
	X^2		0.46			0.41			0.38			0.37	
	X^3		0.27			0.24			0.22			0.21	
	X^4		0.27			0.24			0.21			0.20	

obtained with 2, 3 and 4 expansions of order up to one (monopole plus dipoles) for $\eta=2, 3$ and 4, respectively.

When \mathbf{k} is normal to the largest side of the parallelepiped, i.e., for $\Phi_{inc}=\pi/2$, the results show (Fig. 9) that L becomes rapidly inappropriate. For $\eta \geq 2$, the best solutions are obtained with the circular and elliptical supports, solution quality tending to increase with M . However, it can be observed that, while a quality loss occurs with C for $M > 17$, this does not happen with E. This is due to the fact that, since the elliptical support provides a better match for the body geometry than the circular one does, the set of distances between the sources and the nodes on the boundary is more homogeneous in this case, yielding a more stable linear system to be solved. Consequently, even for high values of M , the source amplitudes generated with E are kept lower than the corresponding one with C, thus avoiding a degradation of the solution quality (not shown here). Even though a precision loss shows up as η increases (e_{BC} values lower than 0.6 being rarely obtained for $\eta=4$), E is the only one of the three supports that remains efficient for $\eta > 3$. However, since it appears that the number of monopoles required increases with η , it is reasonable to suppose that the geometry of the elliptical support may be modified and optimized in order to improve its efficiency. This comment has led the authors to investigate a fourth “easy to implement” support. Actually, as the aspect ratio increases, a natural extension of the elliptical support is the double linear one, made of two parallel lines normal to \mathbf{k} , containing, each, $M/2$ monopoles regularly distributed. This configuration, denoted by LL_M , is such that the sources are arranged in pairs, favoring the formation of dipoles parallel to \mathbf{k} . For all the cases considered, LL_M has been tested for $M=4, 6, 8$ and 10.

Results show (Table III) that the double linear support is efficient in the majority of cases, notably for the highest η value used ($\eta=4$), when all the “single” supports tested failed. While in this case e_{BC} values lower than 0.7 are reached only with the elliptical support with, at least, 20 monopoles, the same quality solutions are obtained with the double linear support with only six sources (i.e., three pairs of monopoles). For $M=8$ or 10, solutions with e_{BC} around 0.4 are generated. Comparison of the results obtained with

the double linear configuration and the expansion technique attests to the efficiency of this support. For example, even for the higher aspect ratio values, LL with only ten monopoles provides a better solution than expansions with 16 sources (with different orders); the expansion technique only yielding solutions significantly better when the order of the four expansions is up to three (i.e., each one containing monopole, dipoles and quadripoles). Figure 10 illustrates the typical arrangement that the sources obtained with the double linear support presents: LL always yields the formation of dipoles (made of the monopoles numbered as “1” and “2,” “3” and “4,” etc...). It should be noted that the representation of the equivalent sources as a set of exact dipoles is a feature of the particular problem addressed (plane wave incidence on a parallelepiped with dimensions given by integer numbers of wavelengths); in general, the multipole expansion of the exact solution contains a monopole component.^{9,15} The use of the double linear support with monopole sources permits the formation of dipoles in the proportion needed in each case.

With the aim of investigating the influence, on the solution quality, of the number of pairs of monopoles and of their relative positioning along the scatterer axis, the solutions obtained with the source configurations LL_4, LL_6, LL_8 and LL_{10} were computed for a between 0.01 and 0.99, i.e., when the size of the support varies from nearly a “point” to almost the entire size of the body, L_η . The corresponding e_{BC} curves

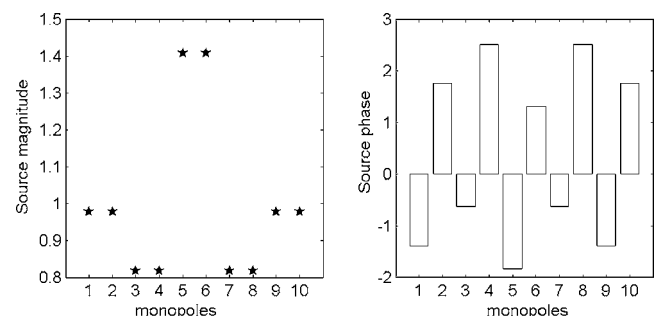


FIG. 10. The LL_{10} source strengths magnitudes (normalized by the mean value, \hat{A}) and phase, for $\eta=2$.

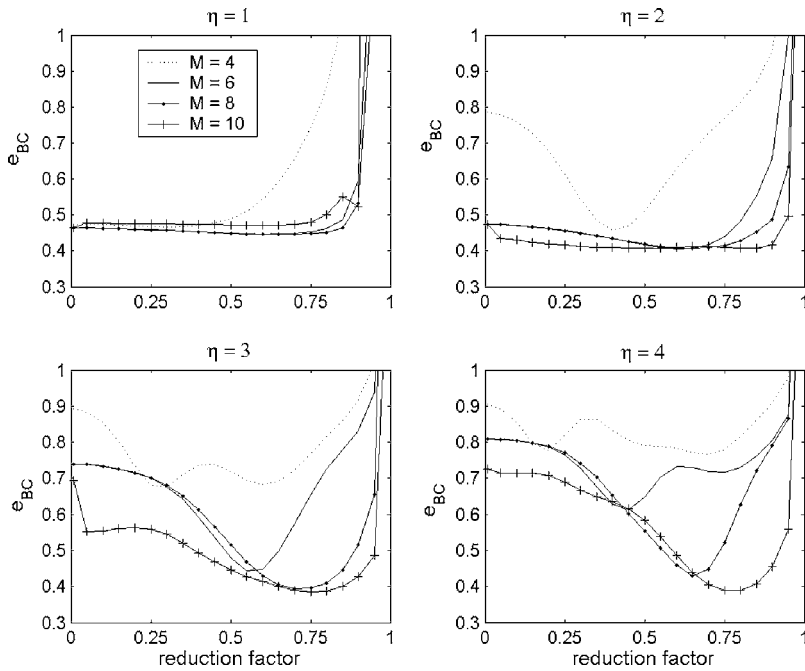


FIG. 11. Evolution of the boundary error with the reduction factor, a , for a parallelepiped with $\eta = 1, 2, 3, 4$ and $\Phi_{\text{inc}} = \pi/2$, obtained with LL₄, LL₆, LL₈, and LL₁₀.

are shown in Fig. 11. Although for each η , the lowest e_{BC} is generally obtained with the highest M , the data suggest the existence of a minimum value for M that guarantees a good solution, corresponding to

$$M \approx 2(\eta + 1). \quad (6)$$

For each curve, a specific band of values of the reduction factor a corresponding to good solutions shows up (Fig. 11). The width and positioning of these bands, centered on a specific value of the reduction factor, a^* , vary with M and η . The knowledge of a^* provides the value of d_s^* , the “optimal” distance between two monopoles in the same line. For all the cases investigated it was found that, whereas the nondimensional quantity kd_s^* varies significantly, the ratio kd_s^*/η is roughly constant, around 1.1. With Eq. (5a) this leads, for M up to 14, to

$$a^* \approx 0.17 \left(\frac{1}{2} M - 1 \right). \quad (7)$$

Equations (6) and (7) constitute a simple rule for determining, given the body aspect ratio, the optimal number of pairs of monopoles to use as well as their positioning, or more precisely, the size the support should have.

D. Case of scatterers with a more complex geometry

Obviously, the more complex the geometry of the scatterer, the knottier the velocity distribution over its boundary, and, consequently, the more complicated its reconstruction, which will require, as well, more complex source configurations. While in the case of expansions this implies increasing the order (and/or the number) of the expansions, for monopoles sets, one must augment significantly their number as well as the degree of complexity of their support geometry. In either case, an increase in the computational effort is involved.

In this section, a simple procedure for dealing with more complex scatterer geometry is proposed, which is based on approximating the structures by a finite number of elements with parallelepipedic shape. Each one of these substructures is to be substituted by an equivalent monopole-source arrangement, determined using the rules given in Secs. III B and III C. Then, the problem is solved globally, considering, as “the” source set, the union of these individual source sets. Obviously, the response quality will depend on how well the original body boundary is reconstructed, i.e., on the number of simple substructures used. Nevertheless, in order to guarantee that the collection of sources still represents an adequate equivalent source set for the whole body, special attention must be paid to the manner in which the structure breaking up is done, this determinant point being illustrated in the following example. A 3D “L-shaped” structure is considered (see Fig. 12); its largest dimensions are $(2 \times 3 \times 1)\lambda$ and its faces are referred to as F_1, \dots, F_8 ; F_7 and F_8 are the lower and upper faces, i.e., those in the $z = -\lambda/2$ and $z = \lambda/2$ planes, respectively. This structure geometry is such that the two simplest ways to break it up into two elements, e_1 and e_2 , are:

- case No. 1: two parallelepipeds with dimensions $(1 \times 2 \times 1)\lambda$ and $(2 \times 1 \times 1)\lambda$;
- case No. 2: one parallelepiped $(1 \times 3 \times 1)\lambda$ plus one cube $(1 \times 1 \times 1)\lambda$

These two cases yield two different combinations of source sets: in the first case, LL₆ and L₉ for e_1 and e_2 , respectively, and LL₈ and L₅ in the second one. Table IV shows the boundary error for the whole structure (e_{BC}) as well as the contribution (in percentage) of each one of the eight faces for the total error obtained with these two monopole sets and with the expansions. The results show that, although the expansion technique provides the desired preci-

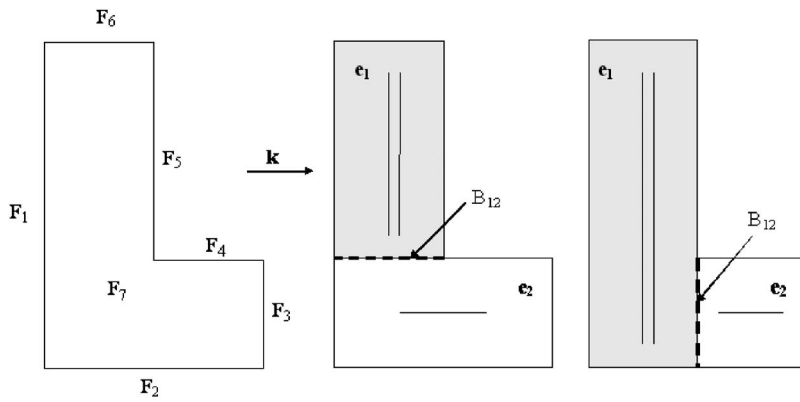


FIG. 12. Representation (from left to right) of the L -shaped structure and the two simplest breaking up cases with the corresponding appropriate supports.

sion, even if requiring, sometimes, a large number of sources (notice the abrupt increase in their number with the expansion order), the use of monopole arrangements, which is a lighter tool, also yields satisfactory solutions. For instance, the source set corresponding to the first case (LL_6 coupled with L_9) leads to a good precision ($e_{BC}=0.5$), having in mind that the number of monopoles is quite low ($M=15$), about 20 times smaller than the number of nodes (288).

Numerical experiments—achieved by varying the size of the support or/and the number of monopoles in one of the two subsource configurations—showed that other solutions with comparable e_{BC} can be obtained. These solutions, however, in spite of presenting local accuracy improvements, always involve deterioration in the velocity reconstitution in other regions. It was verified that the solution corresponding to the rules given in Secs. III B and III C, besides yielding a satisfactory e_{BC} , always presents a good homogeneity in the precision over the entire structure boundary. As for the second case, the solution obtained using LL_8 coupled with L_5 shows, relatively to the first case one, a significant loss in precision, the corresponding e_{BC} being about twice higher. This e_{BC} deterioration is essentially due to the poor velocity reconstitution on F_1 (which, in this case, concentrates 60% of the error—see Table IV) and F_5 , as evidenced in Fig. 13, which shows the normal velocity generated by each of the two source sets.

The difference in the solution quality obtained with the two source sets can be explained as follows. The appropriate source configurations (the linear and the double linear one) were determined separately for the individual substructures. When the substructures are connected, the nodes located on the now common surface, B_{12} , which correspond to inner points of the “new” structure, do not exist anymore (see Fig. 12). On the other hand, since the driving nodes for the determination of the source strengths are those in which the normal velocity is nonzero, the missing nodes (the ones on B_{12}) have a much more important contribution to the boundary condition in the second case than in the first one. As a consequence, while in the first case, the union of the two subsource sets still corresponds to an adequate set when the whole L -shaped structure is considered, this does not happen in the second one, in which the sources located close to B_{12} (in the two subsets) are sort of “confused” by the absence of an important boundary part, leading up to a poor velocity reconstruction on the lower parts of F_1 and F_5 . The way the original structure is divided into parallelepipeds is therefore a crucial point to guarantee that the individual appropriate monopole sets will still constitute, together, an appropriate source set for the complete structure. The simplest way to express the adequate procedure for the structure division is that the separating virtual surface between two substructures should always be taken as parallel to the wave vector.

TABLE IV. The e_{BC} values corresponding to the two breaking up cases of the L -shaped structure; “Ochmann expansions” e_{BC} results.

		Case No. 1		Case No. 2		“Expansion technique” 4 expansions...			
		e_1	e_2	e_1	e_2	X^1	X^2	X^3	X^4
Substructures									
Source sets		LL_6	L_9	LL_8	L_5				
a		0.32	0.5	0.48	0.25				
M (total)		15		13		16	36	64	100
e_{BC}		0.50		1.03		0.39	0.32	0.18	0.17
% contribution	F_1	24		60		29	27	25	27
	F_2	6		3		5	5	6	5
	F_3	7		3		8	6	9	9
	F_4	8		3		3	1	2	1
	F_5	20		15		21	21	17	17
	F_6	3		0		6	5	6	6
	F_7	16		8		14	17	18	17
	F_8	16		8		14	17	18	17

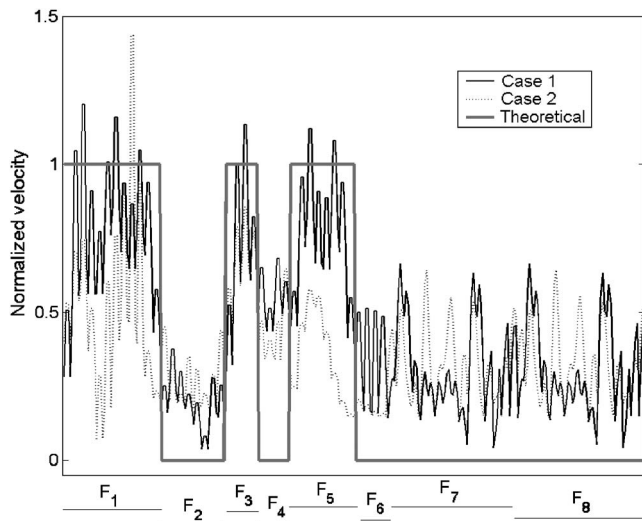


FIG. 13. Normal velocity at the “L-shaped structure” boundary nodes for the two breaking up cases and corresponding theoretical value.

IV. CONCLUSIONS

Easy-to-implement source supports which permit obtaining satisfactory solutions for the scattering problem with a small number of monopoles have been found. It was shown that, for scatterers with an aspect ratio of about unity, a linear support parallel to the incident wave vector is fairly efficient, provided an appropriate number of sources is chosen. A rule relating the nondimensional frequency kL to the suitable number of monopoles has been established. For scatterers with aspect ratio (η) higher than unity, two limiting situations were considered. When \mathbf{k} is parallel to the largest side of the parallelepiped, the best solutions are also reached with the linear support, provided its size increases with η , the same rule found for the cube case still applying. When \mathbf{k} is normal to the largest side of the parallelepiped, all the single source supports investigated (linear, circular, and elliptical) become inappropriate or else require a large number of monopoles. In this case, it has been shown that a double linear support with a relatively small number of sources, arranged in pairs, turns out to be very efficient, leading always to good solutions, which are significantly better than the best ones obtained with the single supports. In this case, a simple rule was found that furnishes, for a given aspect ratio, both a suitable size for the support and an appropriate number of monopoles. For scatterers presenting a more complex geometry, a procedure is proposed which permits composing a suitable source arrangement. It consists of substituting the structure by a proper number of parallelepipeds and then, endowing a suitable source set to each substructure. An easy but determinant way to achieve a structure breaking up that

guarantees the efficiency of the resulting source set is furnished. These simple rules constitute some helpful working guidelines that should contribute to an easier and safer use of the equivalent source method.

ACKNOWLEDGMENT

Financial support was provided by the National Research Council of Brazil, CNPq.

- ¹M. Ochmann, “The source simulation technique for acoustic radiation problems,” *Acustica* **81**, 512–527 (1995).
- ²M. Ochmann, “The full-field equations for acoustic radiation and scattering,” *J. Acoust. Soc. Am.* **105**, 2574–2584 (1999).
- ³M. G. Imhof, “Multiple multipole expansions for acoustic scattering,” *J. Acoust. Soc. Am.* **97**, 754–763 (1995).
- ⁴R. Jeans and I. C. Mathews, “The wave superposition method as a robust technique for computing acoustic fields,” *J. Acoust. Soc. Am.* **92**, 1156–1166 (1992).
- ⁵L. Song, G. H. Koopmann, and J. B. Fahline, “Numerical errors associated with the method of superposition for computing acoustic fields,” *J. Acoust. Soc. Am.* **89**, 2625–2633 (1991).
- ⁶J. B. Fahline 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**, 2808–2818 (1991).
- ⁷G. H. Koopmann, L. Song, and J. B. Fahline, “A method for computing acoustic fields based on the principle of wave superposition,” *J. Acoust. Soc. Am.* **86**, 2433–2438 (1989).
- ⁸R. D. Miller, J. R. Moyer, H. Huang, and H. Uberall, “A comparison between the boundary element method and the wave superposition approach for the analysis of the scattered fields from rigid bodies and elastic shells,” *J. Acoust. Soc. Am.* **89**, 2185–2196 (1991).
- ⁹Y. J. R. Gounot, R. E. Musafir, and J. G. Slama, “A comparative study of two variants of the equivalent sources method in scattering problems,” *Acust. Acta Acust.* **91**, 860–872 (2005).
- ¹⁰P. R. Stepanishen and S. Ramakrishna, “Acoustic radiation from cylinders with a plane of symmetry using internal multipole line source distributions. 1a,” *J. Acoust. Soc. Am.* **93**, 658–672 (1993).
- ¹¹W. Kropp and P. U. Svensson, “Application of the time domain formulation of the method of equivalent sources to radiation and scattering problems,” *Acustica* **81**, 528–543 (1995).
- ¹²D. W. Herrin, T. W. Wu, and A. F. Seybert, “The energy source simulation method,” *J. Sound Vib.* **177**, 447–477 (2003).
- ¹³G. Pavic, “A technique for the computation of sound radiation by vibrating bodies using multipole substitute sources,” *Acust. Acta Acust.* **92**, 112–126 (2006).
- ¹⁴M. E. Johnson, S. J. Elliot, K.-H. Baek, and J. Garcia-Bonito, “An equivalent source technique for calculating the sound field inside an enclosure containing scattering objects,” *J. Acoust. Soc. Am.* **104**, 1221–1231 (1998).
- ¹⁵A. D. Pierce, *Acoustics. An Introduction to Its Physical Principles and Applications*, 2nd ed. (Acoustical Society of America, Melville, 1991).
- ¹⁶P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
- ¹⁷Y. J. R. Gounot, “Características e localização das fontes equivalentes no problema de espalhamento acústico (Characteristics and positioning of equivalent sources for acoustic scattering problems),” Ph.D. dissertation, Univ. Federal do Rio de Janeiro, RJ, 2006.
- ¹⁸T. W. Wu, S. F. Wu, and A. F. Seybert, *Boundary Element Acoustics* (University of Kentucky, 2000).

The integrated extinction for broadband scattering of acoustic waves

Christian Sohl,^{a)} Mats Gustafsson,^{b)} and Gerhard Kristensson^{c)}

Department of Electrical and Information Technology, Lund University, Box 118, S-221 00, Lund, Sweden

(Received 5 June 2007; revised 30 September 2007; accepted 1 October 2007)

In this paper, physical bounds on scattering of acoustic waves over a frequency interval are discussed based on the holomorphic properties of the scattering amplitude in the forward direction. The result is given by a dispersion relation for the extinction cross section which yields an upper bound on the product of the extinction cross section and the associated bandwidth of any frequency interval. The upper bound is shown to depend only on the geometry and the material properties of the scatterer in the static or low-frequency limit. The results are exemplified by permeable and impermeable scatterers with homogeneous and isotropic material properties. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2801546]

PACS number(s): 43.20.Fn, 43.30.Ft, 43.20.Bi [RMW]

Pages: 3206–3210

I. INTRODUCTION

Linear acoustics with propagation and scattering of waves in air and water has been a subject of considerable interest for more than a century. Major contributions to the scattering theory of both acoustic and electromagnetic waves from bounded obstacles was provided by Rayleigh in a sequence of papers. From a theoretical point of view, scattering of acoustic waves shares many features with electromagnetic and elastodynamic wave interaction. For a comprehensive introduction to linear acoustics, see, e.g., Refs. 1 and 2.

The objective of this paper is to derive physical bounds on broadband scattering of acoustic waves. In more detail, the scattering problem discussed here involves how a scatterer of arbitrary shape perturbs some known incident field over a frequency interval. The analysis is based on a forward dispersion relation for the extinction cross section applied to a set of passive and linear constitutive relations. This forward dispersion relation, known as the integrated extinction, is a direct consequence of causality and energy conservation via the holomorphic properties of the scattering amplitude in the forward direction. As far as the authors know, the integrated extinction was first introduced in Ref. 3 concerning absorption and emission of electromagnetic waves by interstellar dust. The analysis in Ref. 3, however, is restricted to homogeneous and isotropic spheroids. This narrow class of scatterers was generalized in Ref. 4 to include anisotropic and heterogeneous obstacles of arbitrary shape.

The present paper is a direct application to linear acoustics of the physical limitations for scattering of electromagnetic waves introduced in Refs. 4 and 5. The broad usefulness of the integrated extinction is illustrated by its diversity of applications, see, e.g., Ref. 5 for upper bounds on the bandwidth of metamaterials associated with electromagnetic interaction. The integrated extinction has also fruitfully been applied to antennas of arbitrary shape in Ref. 6 to establish

physical bounds on directivity and bandwidth. The theory for broadband scattering of acoustic waves is motivated by the dispersion relations in Ref. 7 and the analogy with causality in the scattering theory for particles in Ref. 8.

In Sec. II, the integrated extinction is derived based on the holomorphic properties of the scattering amplitude in the forward direction. The derivation utilizes only the properties of the field in the exterior region, and the result is independent of the boundary conditions imposed on the scatterer as long as the material in the scatterer is linear. The effects of various boundary conditions are discussed in Sec. III, and there applied to the results in Sec. II. In the final section, Sec. IV, the main results of the paper are summarized and possible applications of the integrated extinction are discussed.

II. THE INTEGRATED EXTINCTION

Consider a time-harmonic plane wave $u_i = e^{ik\hat{k}\cdot x}$ (complex excess pressure) with time dependence $e^{-i\omega t}$ impinging on a bounded, but not necessary simply connected, scatterer with support $V \subset \mathbb{R}^3$ of arbitrary shape, see Fig. 1. The plane wave is impinging in the \hat{k} direction, and x denotes the position vector with respect to some origin. The scatterer V is assumed to be linear and time-translational invariant with passive material properties modeled by general anisotropic and heterogeneous constitutive relations. The analysis includes the impermeable case as well as transmission problems with or without losses. The scatterer V is embedded in the exterior region $\mathbb{R}^3 \setminus \bar{V}$, which is assumed to be a compressible homogeneous and isotropic fluid characterized by the wave number $k = \omega/c$. The material properties of $\mathbb{R}^3 \setminus \bar{V}$ are assumed to be lossless and independent of time.

Let $u = u_i + u_s$ denote the total field in $\mathbb{R}^3 \setminus \bar{V}$, where the time-dependent physical excess pressure p is related to u via $p = \text{Re}\{ue^{-i\omega t}\}$. The scattered field u_s represents the disturbance of the field in the presence of V . It satisfies the Helmholtz equation in the exterior of V , see Ref. 2, i.e.,

^{a)}Electronic mail: christian.sohl@eit.lth.se

^{b)}Electronic mail: mats.gustafsson@eit.lth.se

^{c)}Electronic mail: gerhard.kristensson@eit.lth.se

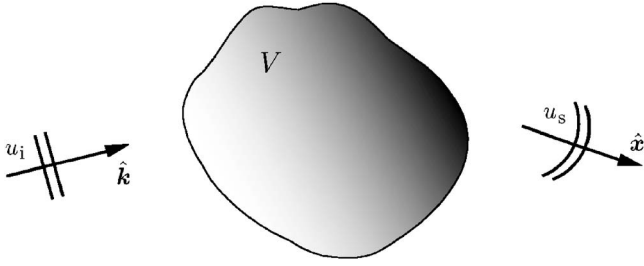


FIG. 1. Illustration of the direct scattering problem: The scatterer V is subject to a plane wave $u_i = e^{i\mathbf{k}\cdot\mathbf{x}}$ impinging in the $\hat{\mathbf{k}}$ direction. The incident field is perturbed by V and a scattered field u_s is detected in the $\hat{\mathbf{x}}$ direction.

$$\nabla^2 u_s + k^2 u_s = 0, \quad \mathbf{x} \in \mathbb{R}^3 \setminus \bar{V}. \quad (1)$$

The boundary condition imposed on u_s at large distances $x = |\mathbf{x}|$ is the Sommerfeld radiation condition

$$\lim_{x \rightarrow \infty} x \left(\frac{\partial u_s}{\partial x} - i k u_s \right) = 0, \quad (2)$$

which is assumed to hold uniformly in all directions $\hat{\mathbf{x}} = \mathbf{x}/x$. The condition (2) establishes the outgoing character of u_s , and provides a condition for a well-posed exterior boundary value problem. For a discussion of various boundary conditions imposed on V , see Sec. III.

From the integral representations in Ref. 9 it is clear that every solution to Eq. (1) satisfying Eq. (2) has an asymptotic behavior of an outgoing spherical wave, i.e.,

$$u_s = \frac{e^{ikx}}{x} S(k, \hat{\mathbf{x}}) + \mathcal{O}(x^{-2}) \quad \text{as } x \rightarrow \infty.$$

The scattering amplitude S is independent of x and describes the interaction of V with the incident field. From a time-domain description of the problem it follows that S is the Fourier transform of some temporal scattering amplitude S_t . Assume S_t is causal in the forward direction in the sense that $S_t(\tau, \hat{\mathbf{k}}, \hat{\mathbf{k}}) = 0$ for $\tau < 0$, where $\tau = ct - \hat{\mathbf{k}} \cdot \mathbf{x}$. Based on this condition, the Fourier transform of S_t reduces to an integral over $\tau > 0$, i.e.,

$$S(k, \hat{\mathbf{k}}) = \int_0^\infty S_t(\tau, \hat{\mathbf{k}}, \hat{\mathbf{k}}) e^{ik\tau} d\tau. \quad (3)$$

The convergence of Eq. (3) is improved by extending its domain of definition to complex-valued k with $\text{Im } k > 0$. Such an extension defines a holomorphic function S in the upper half plane $\text{Im } k > 0$, see Sec. I in Ref. 8. Note that S in general is not a holomorphic function at infinity for $\text{Im } k > 0$ in the absence of the causality condition.

The description of broadband scattering is simplified by introducing a weighted function ϱ of the scattering amplitude in the forward direction. For this purpose, let ϱ denote

$$\varrho(k) = S(k, \hat{\mathbf{k}})/k^2, \quad \text{Im } k > 0.$$

Since S_t is real-valued it follows from Eq. (3) that ϱ is real-valued on the imaginary axis, and that it satisfies the cross symmetry $\varrho(-k^*) = \varrho^*(k)$ (the asterisk denotes complex conjugation) for complex-valued k with $\text{Im } k \geq 0$.

An important measure of the total energy that V extracts from the incident field in the form of radiation or absorption is given by the extinction cross section σ_{ext} . The extinction cross section is related to ϱ via the optical theorem, see Ref. 8,

$$\sigma_{\text{ext}} = 4\pi k \text{Im } \varrho, \quad (4)$$

where $k \in [0, \infty)$. The optical theorem is a direct consequence of energy conservation (or conservation of probability in the scattering theory of the Schrödinger equation) and states that the total energy removed from the incident field is solely determined by $\text{Im } \varrho$. The extinction cross section is commonly decomposed into the scattering cross section σ_s and the absorption cross section σ_a , i.e.,

$$\sigma_{\text{ext}} = \sigma_s + \sigma_a. \quad (5)$$

Here, σ_s and σ_a are defined as the scattered and absorbed power divided by the incident power flux. The scattering and absorption cross sections are related to u_s and u on the boundary ∂V via, see Ref. 10,

$$\sigma_s = \frac{4\pi}{k} \text{Im} \int_{\partial V} u_s^* \frac{\partial u_s}{\partial n} dS, \quad \sigma_a = \frac{4\pi}{k} \text{Im} \int_{\partial V} u \frac{\partial u^*}{\partial n} dS,$$

where the normal derivative $\partial/\partial n$ is evaluated with respect to the outward pointing unit normal vector. In the permeable and lossy case, the absorption cross section σ_a represents the total energy absorbed by V . For a lossless scatterer, $\sigma_a = 0$.

Assume that ϱ vanishes uniformly as $|k| \rightarrow \infty$ for $\text{Im } k \geq 0$. This assumption is justified by the argument that the high-frequency response of a material is nonunique from a modeling point of view. The assumption is also supported by the extinction paradox, which states that σ_{ext} approaches twice the projected cross-section area in the high-frequency regime, i.e., $\text{Im } \varrho(k) = \mathcal{O}(k^{-1})$ as $k \rightarrow \infty$ for real-valued k . For a discussion of the extinction paradox for electromagnetic waves, see Ref. 11. Under this assumption, it follows from the analysis in Refs. 7 and 8 that ϱ satisfies the Hilbert transform or the Plemelj formula

$$\text{Re } \varrho(k') = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\text{Im } \varrho(k)}{k - k'} dk, \quad (6)$$

where k' is real-valued and \mathcal{P} denotes Cauchy's principal value. It is particularly interesting to evaluate Eq. (6) in the static limits as $k' \rightarrow 0$. For this purpose, assume that $\text{Re } \varrho(k') = \mathcal{O}(1)$ and $\text{Im } \varrho(k') = \mathcal{O}(k')$ as $k' \rightarrow 0$, and that ϱ is sufficiently regular to interchange the principal value and the static limit. Based on these assumptions, Eq. (4) yields

$$\lim_{k \rightarrow 0} \text{Re } \varrho(k) = \frac{2}{\pi} \int_0^\infty \frac{\text{Im } \varrho(k)}{k} dk, \quad (7)$$

where it has been used that $\text{Im } \varrho(k) = -\text{Im } \varrho(-k)$ for real-valued k . The optical theorem (4) inserted into Eq. (7) finally yields

$$\int_0^\infty \frac{\sigma_{\text{ext}}(k)}{k^2} dk = 2\pi^2 \lim_{k \rightarrow 0} \text{Re } \varrho(k). \quad (8)$$

The left-hand side of Eq. (8) is referred to as the integrated extinction. The identity provides a forward dispersion relation for the extinction cross section as a direct consequence of causality and energy conservation. Due to the lack of any length scale in the static or low frequency limit, the right-hand side of Eq. (8) is proportional to the volume of V since ϱ has the dimension of volume, cf. the discussion in Ref. 4. Furthermore, the right-hand side of Eq. (8) depends only on the static material properties of V , and is presented in Sec. III for a large class of homogeneous and isotropic scatterers.

The weak assumptions imposed on ϱ in the above-presented derivation are summarized as follows: $\varrho(k) \rightarrow 0$ uniformly as $|k| \rightarrow \infty$ for $\text{Im } k \geq 0$, and $\text{Re } \varrho(k) = \mathcal{O}(1)$ and $\text{Im } \varrho(k) = \mathcal{O}(k)$ as $k \rightarrow 0$ for real-valued k . In general, the integrated extinction (8) is not valid if any of these assumptions are violated, as illustrated in Sec. III C. In fact, the above-noted requirements can be relaxed by the introduction of the Plemelj formula for distributions. The integrated extinction (8) can also be derived using Cauchy's integral theorem, see Ref. 4.

The integrated extinction (8) may be used to establish physical bounds on broadband scattering by acoustic waves. Since σ_{ext} is defined as the sum of the scattered and absorbed power divided by the incident power flux, it is by definition non-negative. Hence, the left-hand side of Eq. (8) is estimated from below by

$$|K| \min_{k \in K} \frac{\sigma(k)}{k^2} \leq \int_K \frac{\sigma(k)}{k^2} dk \leq \int_0^\infty \frac{\sigma_{\text{ext}}(k)}{k^2} dk, \quad (9)$$

where $|K|$ denotes the absolute bandwidth of any frequency interval $K \subset [0, \infty)$, and σ represents either σ_{ext} , σ_s , or σ_a . By combining the left-hand side of Eq. (9) with the right-hand side of Eq. (8), one obtains the fundamental inequality

$$|K| \min_{k \in K} \frac{\sigma(k)}{k^2} \leq 2\pi^2 \lim_{k \rightarrow 0} \text{Re } \varrho(k). \quad (10)$$

The interpretation of Eq. (10) is that it yields an upper bound on the absolute bandwidth $|K|$ for a given scattering and/or absorption cross section $\min_{k \in K} \sigma(k)/k^2$. From Eq. (10), it is seen that the static limit of $\text{Re } \varrho$ bounds the total amount of power extracted by V within the frequency interval K . The electromagnetic analogy to Eq. (10) is, *inter alia*, central for establishing upper bounds on the performance of antennas of arbitrary shape, see Ref. 6.

III. THE EFFECT OF VARIOUS BOUNDARY CONDITIONS

In the following, the static limit of $\text{Re } \varrho$ is examined for various boundary conditions and applied to the integrated extinction (8). For this purpose, V is assumed to be homogeneous and isotropic with sufficiently smooth boundary ∂V to guarantee the existence of boundary values in the classical sense.

A. The Neumann or acoustically hard problem

The Neumann or acoustically hard problem corresponds to an impermeable scatterer with boundary condition $\partial u / \partial n = 0$ for $x \in \partial V$. The physical interpretation of the Neumann boundary condition is that the velocity field on ∂V is zero since no local displacements are admitted. From the fact that u_s only exists in $\mathbb{R}^3 \setminus \bar{V}$, it follows that the corresponding scattered field in the time-domain cannot precede the incident field in the forward direction, i.e., the causality condition imposed on S_t in Sec. II is valid for the Neumann problem. The static limit of S is derived in Refs. 10 and 12 from a power series expansion of u_i and u_s . The result in terms of $\text{Re } \varrho$ reads

$$\lim_{k \rightarrow 0} \text{Re } \varrho(k) = \frac{1}{4\pi} (\hat{\mathbf{k}} \cdot \boldsymbol{\gamma}_m \cdot \hat{\mathbf{k}} - |V|), \quad (11)$$

where $|V|$ denotes the volume of V . Here, $\boldsymbol{\gamma}_m$ models the scattering of acoustic waves in the low frequency limit. In analogy with the corresponding theory for electromagnetic waves in Ref. 4, $\boldsymbol{\gamma}_m$ is termed the magnetic polarizability dyadic. The magnetic polarizability dyadic is proportional to $|V|$, and closed-form expressions of $\boldsymbol{\gamma}_m$ exist for the ellipsoids.

An expression of the integrated extinction for the Neumann problem is obtained by inserting Eq. (11) into Eq. (8), viz.

$$\int_0^\infty \frac{\sigma_{\text{ext}}(k)}{k^2} dk = \frac{\pi}{2} (\hat{\mathbf{k}} \cdot \boldsymbol{\gamma}_m \cdot \hat{\mathbf{k}} - |V|). \quad (12)$$

Note that Eq. (12) is independent of $\hat{\mathbf{k}}$ when $\boldsymbol{\gamma}_m$ is isotropic, i.e., $\boldsymbol{\gamma}_m = \gamma_m \mathbf{I}$ where \mathbf{I} denotes the unit dyadic, corresponding to a scatterer which is invariant under certain point groups, see Ref. 4 and references therein. The product $\hat{\mathbf{k}} \cdot \boldsymbol{\gamma}_m \cdot \hat{\mathbf{k}}$ on the right-hand side of Eq. (12) can be estimated from above by the largest eigenvalue of $\boldsymbol{\gamma}_m$, and associated upper bounds on these eigenvalues are extensively discussed in Ref. 4. The static limit of $\text{Re } \varrho$ in Eq. (11) can also be inserted into the right-hand side of Eq. (10) to yield an upper bound on the scattering and absorption properties of V within any finite frequency interval K .

The integrated extinction (12) takes a particularly simple form for the sphere. In this case, $\boldsymbol{\gamma}_m$ is isotropic with $\gamma_m = 3|V|/2$, see Refs. 4 and 12, and the right-hand side of Eq. (12) is reduced to $\pi|V|/4$. This result for the sphere has numerically been verified using the classical Mie-series expansion in Ref. 1.

B. The transmission or acoustically permeable problem

In addition to the exterior boundary value problem (1) and (2), the transmission or acoustically permeable problem is defined by the interior requirement that $\nabla^2 u_s + k_\star^2 u_s = 0$ for $x \in V$ with the induced boundary conditions $u^+ = u^-$ and $\rho_\delta \partial u^+ / \partial n = \partial u^- / \partial n$. Here, $k_\star = \omega / c_\star$ denotes the wave number in V , and u^+ and u^- represent the limits of u from $\mathbb{R}^3 \setminus \bar{V}$ and V , respectively. The quantity ρ_δ is related to the relative mass density $\rho_{\text{rel}} = \rho_\star / \rho$ via $\rho_\delta = \rho_{\text{rel}} / (1 - i\omega \delta_\star \kappa_\star)$, where κ_\star and ρ_\star

denote the compressibility and the mass density of V , respectively. The conversion of mechanical energy into thermal energy due to losses in V are modeled by the compressional viscosity $\delta_\star > 0$, which represents the rate of change of mass per unit length. In the lossless case, $\delta_\star = 0$, the phase velocity is $c_\star = 1/\sqrt{\kappa_\star \rho_\star}$ and $\rho_\delta = \rho_{\text{rel}}$.

The causality condition introduced in Sec. II is valid for the transmission problem provided $\text{Re } c_\star \leq c$, i.e., when the incident field precedes the scattered field in the forward direction. If V does not fulfill this requirement, ϱ is not holomorphic for $\text{Im } k > 0$ and the analysis in Sec. II does not hold. Hence, the integrated extinction (8) is not valid if $\text{Re } c_\star > c$. This defect can partially be justified by replacing the definition of ϱ by $\varrho = e^{2ika} S(k, \hat{\mathbf{k}})/k^2$, where $a > 0$ is sufficiently large to guarantee the existence of causality in the forward direction. The compensating factor e^{2ika} corresponds to a time-delayed scattered field, and for homogeneous and isotropic scatterers, a sufficient condition for a is that $2a > \text{diam } V$, where $\text{diam } V$ denotes the diameter of V . A drawback of the introduction of the factor e^{2ika} in the definition of ϱ is that the optical theorem no longer can be identified in the derivation. Instead, the integrated extinction reduces to an integral identity for $\text{Re } \varrho$ and $\text{Im } \varrho$. Unfortunately, in this case the integrand does not have a definite sign and therefore the estimate (10) is not valid.

The static limit of the scattering amplitude S for the transmission problem is derived in Refs. 10 and 12. The result in terms of $\text{Re } \varrho$ reads

$$\lim_{k \rightarrow 0} \text{Re } \varrho(k) = \frac{1}{4\pi} ((\kappa_{\text{rel}} - 1)|V| - \hat{\mathbf{k}} \cdot \boldsymbol{\gamma}(\rho_{\text{rel}}^{-1}) \cdot \hat{\mathbf{k}}), \quad (13)$$

where $\kappa_{\text{rel}} = \kappa_\star/\kappa$ denotes the relative compressibility of V , and $\boldsymbol{\gamma}$ represents the general polarizability dyadic. In the derivation of Eq. (13), it has been used that possible losses $\delta_\star > 0$ in V do not contribute in the static limit of $\text{Re } \varrho$, which supports that the argument in $\boldsymbol{\gamma}$ is ρ_{rel} rather than ρ_δ . Analogous to $\boldsymbol{\gamma}_m$, the general polarizability dyadic is proportional to $|V|$, and closed-form expressions for $\boldsymbol{\gamma}$ exist for the ellipsoids, see Refs. 4, 10, and 12. From the properties of $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_m$ in Refs. 4, 10, and 12, it follows that $\boldsymbol{\gamma}(\rho_{\text{rel}}^{-1}) \rightarrow -\boldsymbol{\gamma}_m$ as $\rho_{\text{rel}} \rightarrow \infty$, and hence the static limit of $\text{Re } \varrho$ reduces to Eq. (11) for the Neumann problem as $\kappa_{\text{rel}} \rightarrow 0+$ and $\rho_{\text{rel}} \rightarrow \infty$. Another interesting limit corresponding to vanishing mass density in V is given by $\boldsymbol{\gamma}(\rho_{\text{rel}}^{-1}) \rightarrow \boldsymbol{\gamma}_e$ as $\rho_{\text{rel}} \rightarrow 0+$, where $\boldsymbol{\gamma}_e$ is termed the electric polarizability dyadic in analogy with the low frequency scattering of electromagnetic waves, see Refs. 4, 10, and 12.

The integrated extinction for the transmission problem is given by Eq. (13) inserted into Eq. (8). The result is

$$\int_0^\infty \frac{\sigma_{\text{ext}}(k)}{k^2} dk = \frac{\pi}{2} ((\kappa_{\text{rel}} - 1)|V| - \hat{\mathbf{k}} \cdot \boldsymbol{\gamma}(\rho_{\text{rel}}^{-1}) \cdot \hat{\mathbf{k}}), \quad (14)$$

Note that Eq. (14) is independent of any losses $\delta_\star > 0$, and that the directional character of the integrated extinction only depends on the relative mass density ρ_{rel} . For $\rho_{\text{rel}} \rightarrow 1$, i.e., identical mass densities in V and $\mathbb{R}^3 \setminus \bar{V}$, the integrated extinction is independent of the incident direction $\hat{\mathbf{k}}$, depending only on the relative compressibility κ_{rel} . Furthermore, the

integrated extinction (12) vanishes in the limit as $\kappa_{\text{rel}} \rightarrow 1$ and $\rho_{\text{rel}} \rightarrow 1$, corresponding to identical material properties in V and $\mathbb{R}^3 \setminus \bar{V}$. Due to the non-negative character of the extinction cross section, this limit implies that $\sigma_{\text{ext}} = 0$ independent of the frequency, as expected. Analogous to the Neumann problem, Eq. (14) is also independent of the incident direction $\hat{\mathbf{k}}$ for scatterers with $\boldsymbol{\gamma} = \gamma \mathbf{I}$ for some real-valued γ . The product $\hat{\mathbf{k}} \cdot \boldsymbol{\gamma} \cdot \hat{\mathbf{k}}$ on the right-hand side of Eq. (14) is bounded from above by the largest eigenvalue of $\boldsymbol{\gamma}$, and associated upper bounds on these eigenvalues are discussed in Ref. 4. The static limit of $\text{Re } \varrho$ in Eq. (11) can also be inserted into the right-hand side of Eq. (10) to yield an upper bound on the scattering and absorption properties of V over any finite frequency interval K .

For the simple case of an isotropic and homogeneous sphere, $\gamma = 3|V|(1 - \rho_{\text{rel}})/(2\rho_{\text{rel}} + 1)$, and the right-hand side of Eq. (13) is independent of the incident direction as required by symmetry. Also this result for the sphere has been verified numerically to arbitrary precision using the classical Mie-series expansion.

C. Boundary conditions with contradictions

The integrated extinction (8) and the analysis in Sec. II are not applicable to the Dirichlet or acoustically soft problem with $u = 0$ for $\mathbf{x} \in \partial V$. The physical interpretation of the Dirichlet boundary condition is that the scatterer offers no resistance to pressure. The Dirichlet problem defines an impermeable scatterer for which u_s only exist in $\mathbb{R}^3 \setminus \bar{V}$. Hence, the causality condition introduced in Sec. II is valid. However, the assumption that $\text{Re } \varrho(k) = \mathcal{O}(1)$ as $k \rightarrow 0$ for real-valued k is not valid in this case. Instead, Refs. 10 and 12 suggest that

$$\text{Re } \varrho(k) = \mathcal{O}(k^{-2}) \quad \text{as } k \rightarrow 0$$

for real-valued k . The conclusion is therefore that the integrated extinction (8) is not valid for the Dirichlet problem.

The same conclusion also holds for the Robin problem with impedance boundary condition $\partial u / \partial n + ik\nu u = 0$ for $\mathbf{x} \in \partial V$. The Robin problem models an intermediate behavior between the Dirichlet and Neumann problems, see Ref. 10. The real-valued parameter ν is related to the exterior acoustic impedance η (defined by the ratio of the excess pressure and the normal velocity on ∂V) via $\eta\nu = \sqrt{\rho/\kappa}$, where κ and ρ denote the compressibility and mass density of $\mathbb{R}^3 \setminus \bar{V}$, respectively. In the limits $\nu \rightarrow 0+$ and $\nu \rightarrow \infty$, the Robin problem reduces to the Neumann and Dirichlet problems, respectively. For the Robin problem, the static limit of $\text{Re } \varrho$ for $\nu \neq 0$ reads, see Refs. 10 and 12,

$$\text{Re } \varrho(k) = \mathcal{O}(k^{-1}) \quad \text{as } k \rightarrow 0$$

for real-valued k . Hence, the assumption in Sec. II that $\text{Re } \varrho(k) = \mathcal{O}(1)$ as $k \rightarrow 0$ is not valid for the Robin problem either. The question whether a similar identity to the integrated extinction exists for the Dirichlet and Robin problems with other weight functions than $1/k^2$ in Eq. (8) is presently investigated.

IV. CONCLUSION

The static limits of $\text{Re } \varrho$ in Sec. III can be used in Eq. (10) to establish physical bounds on the amount of energy a scatterer can extract from a known incident field in any frequency interval $K \subset [0, \infty)$. Both absorbed and radiated energy are taken into account. From the analysis of homogeneous and isotropic scatterers in Sec. III, it is clear that the integrated extinction holds for both Neumann and transmission problems. However, the present formulation of the integrated extinction fails for the Dirichlet and Robin problems since the assumption in Sec. II that $\text{Re } \varrho(k) = \mathcal{O}(1)$ as $k \rightarrow 0$ for real-valued k is violated for these boundary conditions.

The eigenvalues of the polarizability dyadics γ , γ_e , and γ_m are easily calculated using either the finite element method or the boundary element method. Some numerical results of these eigenvalues are presented in Refs. 4 and 5 together with comprehensive illustrations of the integrated extinction for electromagnetic waves.

The integrated extinction (8) can also be used to establish additional information on the inverse scattering problem of linear acoustics. One advantage of the integrated extinction is that it only requires measurements of the scattering amplitude in the forward direction. The theory may also be used to obtain additional insights into the possibilities and limitations of manufactured materials such as acoustic metamaterials in Ref. 13. However, the main importance of the integrated extinction (8) is that it provides a fundamental knowledge of the physical processes involved in wave interaction with matter over any bandwidth. It is also crucial to the understanding of the physical effects imposed on a system by the first principles of causality and energy conservation.

ACKNOWLEDGMENTS

Financial support by the Swedish Research Council is gratefully acknowledged. The authors are also grateful for

fruitful discussions with Anders Karlsson at the Department of Electrical and Information Technology, Lund University, Sweden.

- ¹P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
- ²V. V. Varadan and V. K. Varadan, "Acoustic, electromagnetic and elastodynamic fields," in *Field Representations and Introduction to Scattering*, edited by V. V. Varadan, A. Lakhtakia, and V. K. Varadan, Handbook on Acoustic, Electromagnetic and Elastic Wave Scattering, Vol. 1 (Elsevier Science, Amsterdam, 1991), Chap. 1, pp. 1–35.
- ³E. M. Purcell, "On the absorption and emission of light by interstellar grains," *Astrophys. J.* **158**, 433–440 (1969).
- ⁴C. Sohl, M. Gustafsson, and G. Kristensson, "Physical limitations on broadband scattering by heterogeneous obstacles," *J. Phys. A: Math. Theor.* **40**, 11165–11182 (2007).
- ⁵C. Sohl, M. Gustafsson, and G. Kristensson, "Physical limitations on metamaterials: restrictions on scattering and absorption over a frequency interval," *J. Phys. D* **40**, 7146–7151 (2007).
- ⁶M. Gustafsson, C. Sohl, and G. Kristensson, "Physical limitations on antennas of arbitrary shape," *Proc. R. Soc. London, Ser. A* **463**, 2589–2607 (2007).
- ⁷C. Sohl, "Dispersion relations for extinction of acoustic and electromagnetic waves," Licentiate thesis, Dept. of Electrical and Information Technology, Lund University, Lund, Sweden, 2007, <http://www.eit.lth.se> (last viewed 10/24/07).
- ⁸H. M. Nussenzveig, *Causality and Dispersion Relations* (Academic, London, 1972).
- ⁹S. Ström, "Introduction to integral representations and integral equations for time-harmonic acoustic, electromagnetic and elastodynamic wave fields," in *Field Representations and Introduction to Scattering*, edited by V. V. Varadan, A. Lakhtakia, and V. K. Varadan, Handbook on Acoustic, Electromagnetic and Elastic Wave Scattering, Vol. 1 (Elsevier Science, Amsterdam, 1991), Chap. 2, pp. 37–141.
- ¹⁰G. Dassios and R. Kleinman, *Low Frequency Scattering* (Oxford University Press, Oxford, 2000).
- ¹¹H. van de Hulst, *Light Scattering by Small Particles* (Wiley, New York, 1957).
- ¹²R. E. Kleinman and T. B. A. Senior, "Rayleigh scattering," in *Low and high frequency asymptotics*, edited by V. V. Varadan and V. K. Varadan, Handbook on Acoustic, Electromagnetic and Elastic Wave Scattering, Vol. 2 (Elsevier Science, Amsterdam, 1986), Chap. 1, pp. 1–70.
- ¹³J. Li and C. T. Chan, "Double-negative acoustic metamaterial," *Phys. Rev. E* **70**, 055602 (2004).

Acoustoelasticity in soft solids: Assessment of the nonlinear shear modulus with the acoustic radiation force

J.-L. Gennisson,^{a)} M. Rénier, S. Catheline, C. Barrière, J. Bercoff, M. Tanter, and M. Fink
*Laboratoire Ondes et Acoustique, ESPCI, CNRS UMR 7587, INSERM, Université Paris VII,
10 rue Vauquelin, 75231 Paris cedex 05, France*

(Received 4 July 2007; revised 28 August 2007; accepted 29 August 2007)

The assessment of viscoelastic properties of soft tissues is enjoying a growing interest in the field of medical imaging as pathologies are often correlated with a local change of stiffness. To date, advanced techniques in that field have been concentrating on the estimation of the second order elastic modulus (μ). In this paper, the nonlinear behavior of quasi-incompressible soft solids is investigated using the supersonic shear imaging technique based on the remote generation of polarized plane shear waves in tissues induced by the acoustic radiation force. Applying a theoretical approach of the strain energy in soft solid [Hamilton *et al.*, *J. Acoust. Soc. Am.* **116**, 41–44 (2004)], it is shown that the well-known acoustoelasticity experiment allowing the recovery of higher order elastic moduli can be greatly simplified. Experimentally, it requires measurements of the local speed of polarized plane shear waves in a statically and uniaxially stressed isotropic medium. These shear wave speed estimates are obtained by imaging the shear wave propagation in soft media with an ultrafast echographic scanner. In this situation, the uniaxial static stress induces anisotropy due to the nonlinear effects and results in a change of shear wave speed. Then the third order elastic modulus (A) is measured in agar-gelatin-based phantoms and polyvinyl alcohol based phantoms. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2793605]

PACS number(s): 43.25.Dc, 43.25.Ed [OAS]

Pages: 3211–3219

I. INTRODUCTION

For almost two decades, the study of shear wave propagation in soft media became of a peculiar interest in the ultrasonic medical imaging community, especially in the field of dynamic elastography.^{1–3} Methods such as sonoelastography^{4,5} or transient elastography^{6,7} have shown their efficiency to determine the elastic properties of biological soft tissues. However, to date, advanced techniques in that field have been concentrated on the estimation of the second order elastic modulus (μ). In order to better understand pathologies, new refinements were developed to study other mechanical parameters such as anisotropy,⁸ viscoelasticity,^{9,10} or nonlinearity.¹¹ Transient elastography is one of the most efficient approaches, as it images in real time the transient propagation of shear waves in soft media using ultrafast echographic scanners. Based on this concept, three possible ways to quantify nonlinearity in soft solids: Propagation of shocked shear waves,¹² nonlinear interaction between shear waves¹³ and acoustoelasticity.¹¹

The acoustoelasticity theory in solids has been first established by Hugues and Kelly in 1953¹⁴ using the second order and third order elastic coefficient. These parameters come from the expansion of the strain energy density as a function of the invariants of the strain tensor in “Lagrangian” coordinates.¹⁵ The third order moduli can also be expressed as a function of the Landau coefficients (A , B , C).¹⁶ Experimentally, acoustoelasticity consists of measuring the speed of acoustic waves in stressed solids. The third order moduli are

deduced from the slope of the ultrasonic waves velocities as a function of the uniaxial stress or of the hydrostatic pressure applied to the sample. However, acoustoelastic analysis has usually been applied to compressible engineering materials. From our knowledge, regarding soft solids, biological tissues or “nearly” incompressible materials, only few works have been developed on the acoustoelasticity theory for hyperelastic media^{17,18} or on the nonlinear properties of tissue-like phantoms.¹⁹

In a recent paper, shear wave propagation theory in soft solids brought new ideas in the field of elasticity theory. Hamilton *et al.*²⁰ suggested a new expression of the elastic energy density. Applied to nonlinear shear wave experiments reported in Ref. 12, Zabolotskaya *et al.*²¹ obtained a straightforward nonlinear equation where only three elastic coefficients are involved. In this paper, this approach is applied to the field of acoustoelasticity. It results in evolution equations for shear waves in stressed incompressible media that depend on only two elastic coefficients.

In a previous paper,¹¹ we proposed the estimation of the shear wave speed in a medium submitted to an uniaxial stress, using the one dimensional (1D) transient elastography technique.⁸ In these soft solid acoustoelasticity experiments, there were two main difficulties. First, it was difficult to obtain a uniform and constant uniaxial stress, as the stress field in the samples was highly dependent on the boundary conditions. The experimental setup has been improved by using a cylindrical phantom and a mechanical guide to insure a vertical direction of the applied loads. Second, in the 1D transient elastography approach, cylindrical shear waves were generated by giving a low frequency pulse (~ 100 Hz) at the surface of the phantom with a point source

^{a)}Electronic mail: jl.gennisson@espci.fr

(the transducer itself). Thus, in the near field coupling between longitudinal waves and shear waves were nonnegligible. Moreover, model equations were established for plane waves, so experimental assessments were biased by strong diffraction effects. Consequently the experimental set-up has been enhanced by replacing the mechanical vibrator by the acoustic radiation force technique as a shear wave source.²² This technique enables the controlled generation of quasi-plane shear waves in the medium by focusing an ultrasound beam successively at several depths in the medium thus generating several in-depth sources of shear waves. The electronic steering of the ultrasonic beam permits to move the shear source at a higher speed than resulting shear waves propagation speed leading to a supersonic regime. Thus, contrary to the 1D transient elastography, shear diffraction effects were strongly minimized and interaction between longitudinal waves and shear deformations were excluded.

In this paper, the theoretical background of the acoustoelasticity technique is introduced in the first section to deduce the nonlinear shear coefficient A of soft solids. Theoretical equations derived in this section are compared with the classical approach in solids. In the next section, the experimental setup and experimental results are described. Two soft materials used as phantoms in medical imaging are tested: agar-gelatin (AG) phantoms and polyvinyl alcohol cryogel (PVA) phantoms. Finally, the experimental assessments of the nonlinear shear coefficient are discussed.

II. THEORETICAL BACKGROUND

The general principle of acoustoelasticity is based on the expressions of the speed of elastic waves in a uniaxially stressed lossless solid. Basic equations are summarized as follows. In “Lagrangian” coordinates (\mathbf{a}, t) (\mathbf{a} is the equilibrium position of the particle, t is the time), equations of motion are

$$\rho_0 \ddot{u}_i = \frac{\partial P_{ik}}{\partial a_k} \quad \text{for } 1 \leq i, \quad k \leq 3. \quad (1)$$

where ρ_0 , P_{ik} , and $\ddot{\mathbf{u}} = \partial^2 \vec{\mathbf{u}} / \partial t^2$ designate the density, the Piola–Kirchhoff stress tensor (assuming Einstein summation convention of repeated indices) and the particle acceleration (second time derivative of the displacement \mathbf{u}), respectively. The Piola–Kirchhoff stress tensor is given in Lagrangian coordinates by²³

$$P_{ik} = \frac{\partial e}{\partial \left(\frac{\partial u_i}{\partial a_k} \right)}. \quad (2)$$

where e designates the strain energy density which can be developed up to the third order:

$$e = \mu I_2 + \frac{\lambda}{2} I_1^2 + \frac{A}{3} I_3 + B I_1 I_2 + C I_1^3. \quad (3)$$

where (μ, λ) are Lamé coefficients involved in the linear behavior of the solid and (A, B, C) are third order elastic coefficients of Landau, describing the quadratic nonlinear response of the deformed solid. (I_1, I_2, I_3) are invariants of the

Lagrangian strain tensor defined by Landau and Lifshitz.²³

Now let us consider the medium as incompressible. In such a media, an alternative expansion of the strain energy that permit separation of effects due to compressibility and shear deformation was recently defined by Hamilton *et al.*²⁰ and Zabolotskaya *et al.*²¹ Rather than express the strain energy in terms of invariants I_1, I_2 and I_3 as in Eq. (3), one choose instead to express the strain energy in terms of the set of independent invariants III_c, I_2 , and I_3 , where $III_c = \rho_0^2 / \rho^2$ is the third principal invariant of the Green deformation tensor. Thus here in incompressible media, $III_c = 1$, and the strain energy is

$$e = \mu I_2 + \frac{A}{3} I_3 + D I_2^2, \quad (4)$$

where D is the fourth order elastic constant. In the following sections, only propagation of plane waves of small amplitude will be considered, so fourth order terms will be neglected. Developing the invariants using the strain tensor \bar{S} defined by

$$S_{ik} = \frac{1}{2} \left(\frac{\partial u_i}{\partial a_k} + \frac{\partial u_k}{\partial a_i} + \frac{\partial u_l}{\partial a_i} \frac{\partial u_l}{\partial a_k} \right), \quad (5)$$

and according to Eq. (1), one obtains the following expression of the equation of motion:

$$\begin{aligned} \rho_0 \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial P_{ik}}{\partial a_k} = & \mu \left(\frac{\partial^2 u_i}{\partial a_k^2} + \frac{\partial^2 u_k}{\partial a_i \partial a_k} \right) + \left(\mu + \frac{A}{4} \right) \\ & \times \left(\frac{\partial^2 u_l}{\partial a_i \partial a_k} \frac{\partial u_l}{\partial a_k} + \frac{\partial u_l}{\partial a_i} \frac{\partial^2 u_l}{\partial a_k^2} + \frac{\partial^2 u_k}{\partial a_l \partial a_k} \frac{\partial u_l}{\partial a_i} \right. \\ & + \frac{\partial^2 u_l}{\partial a_k^2} \frac{\partial u_l}{\partial a_i} + 2 \frac{\partial u_k}{\partial a_l} \frac{\partial^2 u_l}{\partial a_k \partial a_i} \left. \right) + \frac{A}{4} \left(\frac{\partial^2 u_k}{\partial a_k \partial a_l} \frac{\partial u_l}{\partial a_i} \right. \\ & \left. + \frac{\partial u_k}{\partial a_l} \frac{\partial^2 u_l}{\partial a_k \partial a_i} \right). \end{aligned} \quad (6)$$

In order to simplify further analytical developments, let us consider now some important points. First, in an acoustoelasticity experiment, two main displacements are involved. The static displacement due to the uniaxial stress applied and the dynamic displacement due to the propagation of shear waves induced here by the acoustic radiation force of the ultrasonic “palpating” beam. Moreover, shear strains must be of small amplitude compared to the static deformation induced by the uniaxial stress σ , which is the case experimentally here as ultrasonic radiation force induces very small (micrometric) displacements. Thus, the displacement can be expressed as the sum of the static displacement (\vec{u}^S) induced by the uniaxial stress and of the dynamic displacement (\vec{u}^D) due to the shear wave propagation:

$$\vec{u}^{\text{TOT}} = \vec{u}^D + \vec{u}^S. \quad (7)$$

Neglecting the nonlinear propagation of the shear wave and static deformations of higher order (i.e., $i = 1, 2, 3$), the wave equation is written as

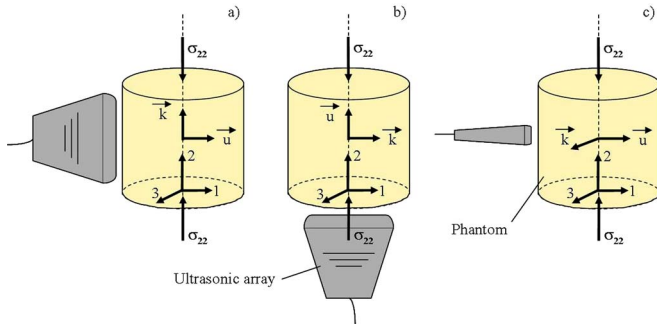


FIG. 1. (Color online) The three possible configurations generating (using the acoustic radiation pressure) and detecting (with an ultrafast scanner) the propagation of polarized shear waves in the phantom under uniaxial stress (σ_{22}). The three resulting shear waves are designated as (a) 12 or 32, (b) 21 or 23, and (c) 13 or 31.

$$\begin{aligned} \rho_0 \frac{\partial^2 u_i^D}{\partial t^2} = & \mu \left(\frac{\partial^2 u_i^D}{\partial a_k^2} + \frac{\partial^2 u_k^D}{\partial a_i \partial a_k} \right) + \left(\mu + \frac{A}{4} \right) \left(\frac{\partial^2 u_1^D}{\partial a_1 \partial a_k} \frac{\partial u_1^S}{\partial a_k} \right. \\ & + \frac{\partial u_1^S}{\partial a_i} \frac{\partial^2 u_1^D}{\partial a_k^2} + \frac{\partial^2 u_k^D}{\partial a_1 \partial a_k} \frac{\partial u_i^S}{\partial a_1} + \frac{\partial^2 u_1^D}{\partial a_k^2} \frac{\partial u_i^S}{\partial a_1} \\ & + 2 \frac{\partial u_k^S}{\partial a_1} \frac{\partial^2 u_i^D}{\partial a_k \partial a_1} \left. \right) + \frac{A}{4} \left(\frac{\partial^2 u_k^D}{\partial a_k \partial a_1} \frac{\partial u_1^S}{\partial a_i} \right. \\ & \left. + \frac{\partial u_k^S}{\partial a_1} \frac{\partial^2 u_1^D}{\partial a_k \partial a_i} \right). \end{aligned} \quad (8)$$

Second, for easier understanding and to simplify the calculations, one can deal with linearly polarized plane shear waves. For example, a plane wave that propagates along a_2 axis is considered, with a particle motion restricted to a_1 axis, according to those plotted on Fig. 1(a). The particle displacement in the wave is then expressed as $u_1^D(a_2, t)$. For sake of simplicity, the plane shear wave is characterized by the two indices according to the following convention 12. The first index corresponds to the direction of the shear displacement induced by radiation force (direction of polarization of the shear wave) and the second index corresponds to the axis of propagation of the shear wave. Let us chose the uniaxial stress directed along the a_2 axis. Then it comes:

$$\rho_0 \frac{\partial^2 u_1^D}{\partial t^2} = \mu \frac{\partial^2 u_1^D}{\partial a_2^2} + \left(2\mu + \frac{A}{2} \right) \left(\frac{\partial^2 u_1^D}{\partial a_2^2} \right) \left(\frac{\partial u_1^S}{\partial a_1} + \frac{\partial u_2^S}{\partial a_2} \right). \quad (9)$$

Third, because of the static uniaxial stress applied, three states and their coordinates need to be distinguished: the natural state, \mathbf{a} , the initial state, \mathbf{x} and the current state, \mathbf{y} related to shear wave propagation. It is useful to express the acoustoelastic equations in the initial coordinates, which, because of the infinitesimal nature of \mathbf{u}^D coincide with the laboratory (Eulerian) coordinates. For simplicity, one supposes that the initial stress and strain are uniform and defined by the static displacement \mathbf{u}^S : $\mathbf{x} = \mathbf{a} + \mathbf{u}^S$. The dynamic displacement \mathbf{u}^D , due to wave propagation is defined by: $\mathbf{y} = \mathbf{x} + \mathbf{u}^D$. The equation of motion for \mathbf{u}^D comes from development around the initial state of Eqs. (1) and (2). Then with the change of variable, $\mathbf{a} \rightarrow \mathbf{x}$, by neglecting again the terms of higher order and by use of the chain rule, it comes in the general case:

$$\frac{\partial^2 u_i^D}{\partial a_k^2} \cong \frac{\partial^2 u_i^D}{\partial x_k^2} \left(1 + 2 \frac{\partial u_k^S}{\partial x_k} \right), \quad \frac{\partial^2 u_k^D}{\partial a_i \partial a_k} \cong \frac{\partial^2 u_k^D}{\partial x_i \partial x_k}. \quad (10)$$

So for the plane wave $u_1^D(a_2, t)$, the wave equation in the new coordinates system becomes:

$$\begin{aligned} \rho_0 \frac{\partial^2 u_1^D}{\partial t^2} = & \frac{\partial^2 u_1^D}{\partial x_2^2} \left[\mu + 2\mu \left(\frac{\partial u_1^S}{\partial x_1} + 2 \frac{\partial u_2^S}{\partial x_2} \right) \right. \\ & \left. + \frac{A}{2} \left(\frac{\partial u_1^S}{\partial x_1} + \frac{\partial u_2^S}{\partial x_2} \right) \right]. \end{aligned} \quad (11)$$

Finally by using the Hooke's law²⁴ and assuming an incompressible medium (Poisson's ratio $\nu=0.5$), the spatial derivatives of the static displacement are:

$$\frac{\partial u_2^S}{\partial x_2} = -\frac{\sigma_{22}}{E} \approx -\frac{\sigma_{22}}{3\mu}, \quad \frac{\partial u_1^S}{\partial x_1} = \frac{\partial u_3^S}{\partial x_3} = \frac{\nu \sigma_{22}}{E} \approx \frac{\sigma_{22}}{6\mu}, \quad (12)$$

where E is Young's modulus. As one deals with compression (opposite of the dilatation), a negative stress was considered. Then, the nonlinear elastodynamic equation is rewritten as

$$\rho_0 V_{S12}^2 = \mu - \sigma_{22} \left(1 + \frac{A}{12\mu} \right), \quad (13)$$

with V_S being the shear wave velocity.

According to the axis defined by Fig. 1, the result is the same for the couple of indices 32. Regarding the other possible couples of indices, two more nonlinear elastodynamic equations are obtained.

For indices 21 or 23 the shear wave velocity is then:

$$\rho_0 V_{S21}^2 = \mu - \sigma_{22} \left(\frac{A}{12\mu} \right), \quad (14)$$

and for indices 13 or 31:

$$\rho_0 V_{S13}^2 = \mu + \sigma_{22} \left(1 + \frac{A}{6\mu} \right). \quad (15)$$

Note that Eqs. (13)–(15), if the medium is unstressed ($\sigma_{22}=0$), correspond to the classical shear wave propagation equation in an isotropic media. Thus, the medium reveals an anisotropic behavior resulting from the nonlinear effects.^{11,25} Moreover, changing the stress in Eqs. (13)–(15) is equivalent to modifying the shear modulus (μ) according to a linear dependence with a linear slope defined by the shear modulus (μ) and the third order nonlinear coefficient (A).

With this approach, theory of acoustoelasticity in incompressible media developed in this section of the present paper is fully consistent with classical acoustoelasticity theories defined in classical solids. For example, one can take as a comparison the theoretical framework developed by Hughes and Kelly.¹⁴ Hughes and Kelly establish expressions for velocities of elastic waves in stressed solids using Murnaghan's theory¹⁵ of finite deformations and third order terms in the energy. These expressions can also be expressed as a function of the Landau coefficient¹⁶ for plane shear waves under uniaxial stress. Assuming the material incompressible, one can use following statements: $\lambda \gg \mu$ and $B = -K = -(\lambda + \frac{2}{3}\mu)$,²⁶ the equations derived from Hughes and Kelly lead to Eqs.

(13)–(15). Similarly, using the theoretical approach of Thurston and Brugger²⁷ and summarized by Norris in Ref. 16, Eqs. (13)–(15) can be retrieved.

Thus, the application of both previous theories (Hugues and Kelly;¹⁴ Thurston and Brugger²⁷) to incompressible media is consistent with Hamilton *et al.*²⁰ and Zabolotskaya *et al.*²¹ papers. With this theoretical approach, in the case of soft solids, the separation of compressibility and shear deformations is possible and shear velocities depend only on the second and third order shear elastic moduli (μ, A). The influence of second and third order elastic coefficients related to longitudinal waves can be neglected. Therefore, compared to expressions in isotropic and compressible medium,¹¹ neither the third order Landau coefficients B and C [Eq. (3)] (B is defined as the nonlinear coefficient of the coupling term between shear and longitudinal waves and C is defined as the nonlinear coefficient for longitudinal waves) nor λ the second Lamé coefficient (λ is linked to the compressibility of the material) have any effect on the determination of shear velocities.

In the following section, as the supersonic shear imaging technique enables the controlled generation of quasiplane shear waves and the total discrimination in time between shear and compressional deformations during the acquisition due to the huge discrepancy between shear ($\sim 1 \text{ m s}^{-1}$) and compressional wave ($\sim 1500 \text{ m s}^{-1}$) speeds; we will take advantage of these relationships [Eqs. (13)–(15)] to estimate the linear and nonlinear elastic moduli (μ, A) in an acoustoelasticity experiment.

III. MEASUREMENT SYSTEM

The experiments were performed on two kinds of quasi-incompressible soft solids commonly used for the mimicking of soft tissues: five AG phantoms and three PVA phantoms.

The five AG phantoms were prepared with different concentrations of gelatin and agar diluted in water (initially heated at 50 °C) following a strict procedure earlier described in Ref. 28. Two AG phantoms (GEL_1 and GEL_4) were cylinders of 12.5 cm diameter and 15 cm high, with a concentration of 3% by weight of agar (Agar, Prolabo, Fontenay/Bois, France) and 5% and 15% by weight of gelatin (Gelatin, Prolabo, Fontenay/Bois, France), respectively. The three other ones (GEL_2, GEL_3, GEL_5) were cylinders of 12 cm diameter and 15 cm high with 3% by weight of agar and 7%, 10%, and 8.5% by weight of gelatin, respectively. The agar was used as scatterers and gelatin was used as matrix.

PVA phantoms were constituted of 10% by weight of PVA (polyvinyl alcohol hydrolyzed, Sigma-Aldrich, St Louis, MO) dissolved in water. The solution container was heated in hot water to 80 °C, and to minimize dehydration the container was covered. 3% by weight of Sigmacell (Sigmacell Cellulose type 20, Sigma-Aldrich, St Louis, MO) was added to serve as scatterers. The three PVA phantoms were cylinders of a 15 cm high and 12 cm diameter. PVA phantoms underwent two (PVA_1), three (PVA_2) and five (PVA_3) freezing–thawing cycle to polymerize them with a certain level of elasticity as described in previous papers.^{29,30}

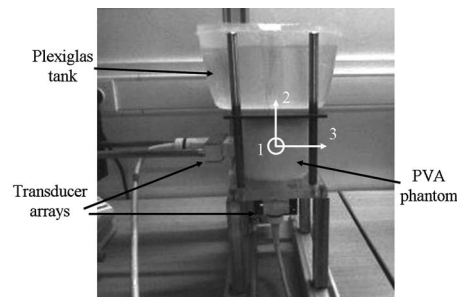


FIG. 2. Experimental set-up on a PVA phantom (PVA_2). Two transducer arrays are placed along the vertical axis of the phantom and at the bottom of the phantom, in order to measure the shear waves propagation with the 21 or 23 [Eq. (19)] and the 12 or 32 [Eq. (18)] indices respectively. During this experiment, the tank was incrementally filled with water in steps of 0.3 kg.

In experiments, a rigid Plexiglas tank was added on top of the phantoms and filled with different amounts of water [e.g., with the agar-gelatin phantom GEL_4, from 0.3 kg (weight of the Plexiglas plate and tank) to about 2.1 kg in incremental steps of 0.3 kg step] to control the uniaxial stress (Fig. 2). The loading charges were adapted depending of the phantom consistency. The Plexiglas tank and the phantom support were parallel to make the stress as uniaxial as possible. Moreover coupling gel was applied at the top and at the bottom of the phantoms to minimize friction effects between phantoms and supports.

Then, the 4.3 MHz ultrasonic array (128 elements, 0.33 mm pitch, 10 mm elevation, 60 mm elevation focus, Vermon, Tours, France) was applied at three different positions surrounding the phantoms as described on Fig. 1 to investigate the three possible shear wave propagations defined in part II. So, the ultrasonic beam was two times perpendicular [Figs. 1(a) and 1(c)] and one time parallel [Fig. 1(b)] to the uniaxial stress. The ultrasonic array was connected to an ultrafast scanner (the system was already describe in Ref. 31) and transmits ultrasound beams through a wide layer of coupling gel, which did not induce significant artifact on the applied uniaxial stress field.

To generate the radiation force, the ultrafast scanner was used to create an ultrasound-focused beam in the phantom at a chosen location (so called “supersonic push”). The typical ultrasound pulse was made of 500 oscillations at 4.3 MHz. This corresponded to a “pushing time” of $\sim 120 \mu\text{s}$. Just after the generation of the “pushing” beam, the scanner began an ultrafast imaging sequence by sending plane-wave insonifications at a high-frame rate (2000 Hz) in order to catch the shear wave created by the supersonic push.²²

In post acquisition, displacement fields were obtained by 1D cross correlation of beamformed RF signals stored with the reference echo image.³² The results were a set of images giving the displacement induced by the shear wave at each sample time (Fig. 3). In Fig. 3 one can also notice the effect of the attenuation on shear waves. Nevertheless, the attenuation is not relevant on the shear wave speed assessment because the signal to noise ratio is high enough to allow a good measurement of the displacements induced by shear waves along the imaged plane (about 25 mm, Fig. 3). The use of an inverse problem based on the Helmholtz equation, as described in previous papers,^{6,22} allowed us to map

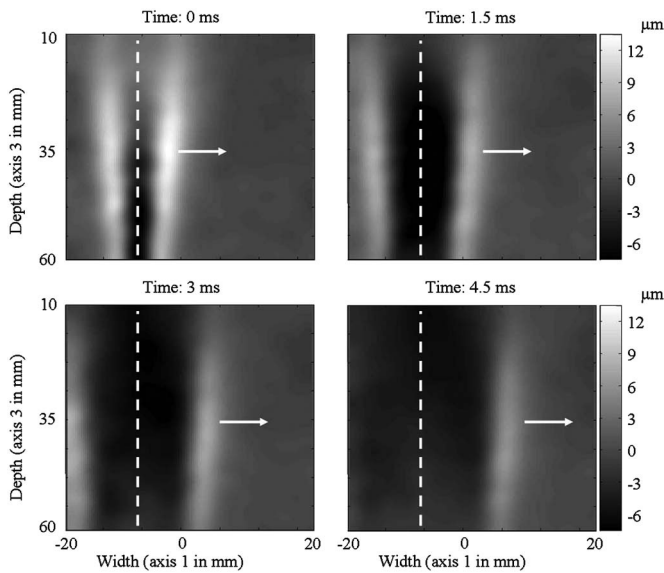


FIG. 3. Experimental displacement field in agar-gelatin phantom (GEL_1) at successive times after the “push”. The dotted line represents the axis of the polarization of the shear source (the supersonic push on axis 3) and the arrow the direction of propagation of the shear wave (axis 1).

the shear wave speed. Then knowing the density ($\rho_0 \approx 1000 \text{ kg/m}^3$), the shear modulus and its corresponding standard deviation, for each stress applied, were retrieved by taking the mean shear wave speed and its corresponding standard deviation at all depth between 0 and 20 mm in width (Fig. 4).

IV. RESULTS

The experiments on AG phantoms and on PVA phantoms were gathered in Figs. 5 and 6 respectively. Figures 5 and 6 showed an increasing speed versus applied stress for the shear wave propagating in the 13 plane and a decreasing speed in the 12 plane. The shear wave speed in the 21 plane was sometimes increasing and sometimes decreasing with respect to the applied stress, so it has an influence on both sign and value of the third order elastic coefficient (A). At the zero stress point for each data set, the shear Lamé coef-

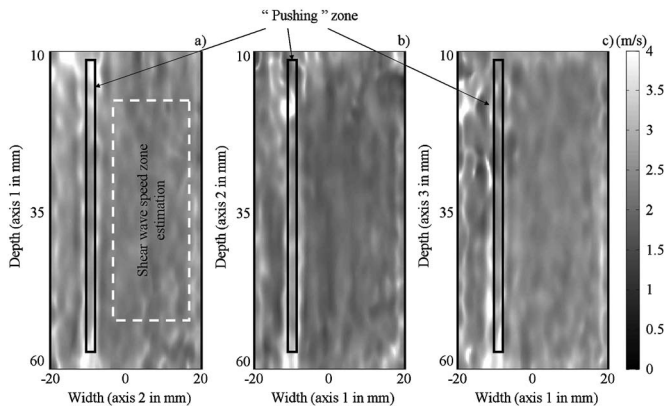


FIG. 4. Map of the shear wave speed for each position of the transducer array, (a) 12 (equivalent to 32), and (b) 21 (equivalent to 23), and (c) 31 (equivalent to 13) in AG phantom (GEL_1) at zero stress. The point line corresponds to the region of interest used for the estimation of the mean shear wave speed and its corresponding standard deviation.

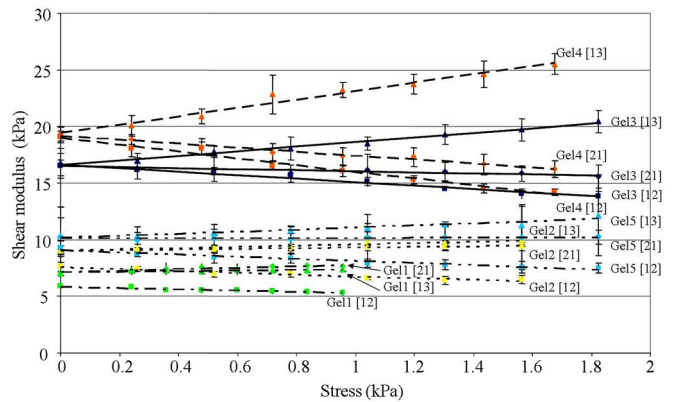


FIG. 5. (Color online) Experimental shear moduli as a function of applied uniaxial stress for each axis of propagation investigated in agar-gelatin phantoms. Four different phantoms with different gelatin concentration were investigated: Gel_1 (5% by weight of gelatin), Gel_2 (7%), Gel_3 (10%), Gel_4 (15%) and Gel_5 (8.5%). The slopes given by the dashed and solid lines were obtained from a least mean square fit.

ficient μ was computed as the mean value of the moduli measured for the three experimental configurations (Fig. 1). The values, summed up in Table I, show a good estimation of the linear shear modulus with a maximum 9% standard deviation in AG phantoms and 2% in PVA phantoms. As presented in Figs. 5 and 6, the shear modulus evolution clearly exhibits the appearance of an anisotropic behavior in all phantoms. Moreover, the shear modulus was linearly dependent with respect to the uniaxial stress in all materials as expected in the theoretical framework. From a quantitative point of view, the shear modulus affecting each quasi-plane shear wave (12, 21, 13) increased or decreased respectively with the following slopes (obtained from a least mean square fit of the data): GEL_1 (−0.56, 0.61, 0.26), GEL_2 (−0.79, 0.25, 0.59), GEL_3 (−1.55, −0.46, 2.03), GEL_4 (−3.02, −1.75, 3.71), GEL_5 (−0.94, 0.05, 0.92), PVA_1 (−0.48, 0.31, 0.55), PVA_2 (−1.11, −0.10, 1.26), and PVA_3 (−1.13, −0.23, 1.35). One can notice that when the linear shear modulus was higher, the value of slopes increased for the 13 plane and decreased for the 21 and 12 planes in both AG phantoms and PVA phantoms. Further, the slope for the

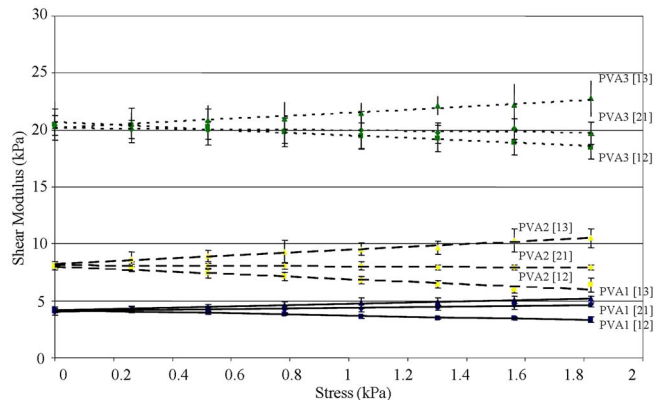


FIG. 6. (Color online) Experimental shear moduli as a function of applied uniaxial stress for each axis of propagation investigated in PVA phantoms. Three different phantoms, with different elasticity defined by the number of freezing-thawing cycles applied (2, 3, and 5 cycles), were investigated. The slopes given by the dashed and solid lines were obtained using a linear fit.

TABLE I. Elastic moduli measured in two different soft solids.

Materials	Linear shear modulus (Lamé coefficient μ (kPa) (mean value at 0 stress))	Non linear shear modulus [Landau coefficient A (kPa)]			
		Indice of propagation 12	Indice of propagation 21	Shear wave indices 13	Mean value
Agar-gelatin phantom 1	6.6±0.6	-35.0	-48.6	-29.4	-37.7±9.8
Agar-gelatin phantom 2	8.5±0.8	-21.5	-25.6	-21.0	-22.7±2.5
Agar-gelatin phantom 5	9.9±0.5	-7.1	-5.9	-4.8	-5.9±1.2
Agar-gelatin phantom 3	16.6±0.1	109.8	91.8	102.8	101.4±9.0
Agar-gelatin phantom 4	19.2±0.1	466.4	404.1	312.8	394.4±77.2
PVA phantom 1	4.1±0.1	-25.8	-15.4	-11.2	-17.5±7.5
PVA phantom 2	8.1±0.1	10.7	9.7	12.6	11.0±1.4
PVA phantom 3	20.4±0.1	31.8	56.3	42.8	43.6±12.2

[21](#) plane changed its sign with the stiffness of the material. In all experiments, correlation coefficients of estimated slopes using a linear fit were higher than 80% [except for two cases in AG phantoms (48% and 10%) and one case in PVA phantoms (51%)].

Using Eqs. [\(13\)–\(15\)](#), the experimental estimates of shear modulus versus axial stress slopes enable the calculation of the nonlinear Landau coefficient A for all soft solids. The mean value of A and its standard deviation were calculated from three independent values of A for each direction of propagation. One can notice that the standard deviation of A for all experiments in AG phantoms and PVA phantoms is less than 28% excluding for the one softer PVA phantom (43% in PVA_1).

Regarding the results presented in Table I, the nonlinear shear modulus A is plotted as a function of the shear modulus μ in Fig. 7. One can notice that A seems to evolve as an exponential as a function of μ for AG phantoms. For PVA phantoms, no specific variation can be observed.

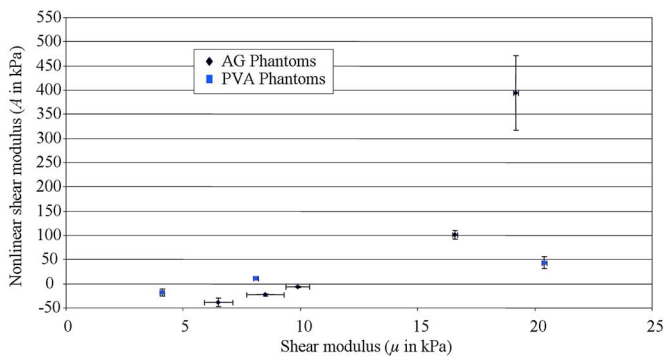


FIG. 7. (Color online) Experimental nonlinear shear moduli A as a function of experimental shear modulus μ for each phantom. In AG phantoms A seems to evolve as an exponential, but no specific variation can be noticed on PVA phantoms.

V. DISCUSSION

Results summarized in Table I confirm the fact that the experimental setup gives three independent ways to estimate the nonlinear coefficient A . Indeed, in the AG phantoms, estimates of A from the slopes obtained from shear wave velocities are in very good agreement for the four last phantoms (the standard deviation is less than 20%). AG phantom GEL_1 exhibits a 26% standard deviation for the A coefficient. In the PVA phantoms, the agreement for the different directions of propagation is also quite good: the standard deviation of the A coefficient estimates inferior to 28% for the two last phantoms. This can be explained by the difficulties to perfectly control the setup with such very soft phantoms, which cannot be submitted to large stresses in order to preserve their physical integrity, e.g., with a relatively large stress the phantoms are destroyed under the weight of the water tank. Nevertheless the same problem as in AG phantom appears for the very soft phantom PVA_1. Moreover, positioning the transducer array correctly with regards to the axis of the uniaxial stress is not trivial, since one cannot perfectly know the true spatial distribution of the uniaxial stress. However, compared to others results already reported in previous papers³³ for metals, crystals or polystyrene,³⁴ these experimental estimates of μ and A in soft solids confirm the fact that the A Landau coefficient exhibits values (kPa), consequently much lower than in classical solids.

Comparing results obtained in AG phantoms and in PVA phantoms, the shear modulus values are in the same range of elasticity for both type of phantoms. However, the behavior of the nonlinear coefficient A is quite different. For example, if one compares the AG phantom GEL_2 and the PVA phantom PVA_1, their shear moduli are practically the same (8.5±0.8 kPa and 8.15±0.1 kPa respectively). But the Landau coefficient A is negative in AG phantoms and positive in PVA phantoms. Moreover the range of elasticity of A is

TABLE II. Shear wave speed (13 direction) in an AG phantom (5% by weight of gelatin) at zero stress as a function of the height from the top of the phantom.

Shear modulus (kPa)	6.5±0.2	6.5±0.1	6.6±0.2	6.7±0.2	6.8±0.8
Height (cm)	2	4	6	8	10

larger in AG phantoms (from -37.7 to 394.4 kPa) than in PVA phantoms (from -17.5 to 43.6 kPa). Once again, comparing AG phantom GEL_4 and PVA phantom PVA_3, the shear modulus is quite identical but the A value is practically 10 times higher in AG phantom. Thus there is a big difference from the nonlinear point of view between these two kinds of material even though their linear behavior is quite similar.

The shear modulus at zero stress is determined by averaging the three measurements in each direction. On the one hand, the estimation of the shear modulus at zero stress in AG phantoms exhibits a quite small 7% standard deviation. The reason why the shear coefficient is not the same for all directions, as it should be in an ideally isotropic medium, is that the measurements were performed in the middle of the sample. For example, let us consider a horizontal plane at a given depth in the phantom with no uniaxial stress applied. Above this plane, the own weight of the phantom induces stress, which depends on the height of the column of the upper part of the material. Consequently, a uniaxial stress increasing with depth appears due to gravity effects. Thus, the speed of the shear waves measured should increase with depth. To confirm this argument, the shear wave speed was estimated at several depths within an AG phantom (GEL_1 with 5% by weight of gelatin and 3% of by weight of agar) from 2 cm top down to 10 cm with 2 cm steps. Resulting values for the 13 direction (the transducer array was horizontal) are summarized in Table II.

The shear wave speed increases as a function of the height and confirms the influence of the gravity force. Results presented on Fig. 5 show that this effect is much less pronounced as the gelatin concentration increases (and consequently shear modulus at zeros stress). On the other hand, in PVA phantoms the standard deviation also decreases to 2%. So, PVA phantoms seem to be less submitted to gravity effects. This is probably related to the internal structure of PVA phantoms³⁵ (polymer chains) more cross linked and stiffer compared to the internal structure of AG phantoms³⁶ (network of rigid gelatin rods). Nevertheless, these gravity effects do not affect drastically assessment of the third order elastic coefficient.

From Eqs. (13)–(15) one can see that starting from an isotropic solid the applied stress implies an anisotropic behavior of the material. In other words, modifying the stress is equivalent to induce an apparent anisotropy in the medium due to nonlinear effects. For a given stress one can define a specific nonlinear anisotropy corresponding to a specific set of different independent elastic constants (μ_{12} , μ_{21} , μ_{13}) depending of the nonlinear coefficient. At zero stress, velocities correspond to an isotropic solid ($\rho_0 V_{ij}^2 = \mu$; $i, j = 1, 2, 3$). It is

confirmed experimentally for PVA phantoms and the two stiffer AG phantoms, as all lines converge towards the same point at zero stress. However, for very soft solids, for example GEL_1, GEL_2, or GEL_5, an initial anisotropy was present due to gravity effects although the material was isotropic. Under their own weight net force, phantoms were naturally stressed and an apparent resulting anisotropy appears. In linear classical solids this kind of anisotropy was corresponding to a hexagonal system,²⁴ so called “transverse isotropy” because two shear waves speeds are identical in the plane defined by the polarization of the two waves; this plane can be defined as isotropic (The shear wave speeds assessed on the 21 and 13 direction were equivalent (GEL_1 (21 2.66 ± 0.06 m/s), 13 2.64 ± 0.03 m/s), GEL_2 (21 3.00 ± 0.08 m/s 13 3.00 ± 0.02 m/s), GEL_5 (21 3.20 ± 0.41 m/s, 13 3.21 ± 0.26 m/s)). Nevertheless, here the apparent anisotropy cannot be defined because anisotropy is defined with the symmetry of the second order elastic moduli. For instance the anisotropy induced by the stress effects must be distinguished from those caused by intrinsic material anisotropy.¹⁶ Now, for a higher uniaxial stress (e.g., $\sigma_{22} = 1.2$ kPa), in each material, three different velocities are observed (Figs. 5 and 6). One can notice that nonlinear effect increase the observed anisotropy and the behavior of the medium is more complicated. These experiments interestingly emphasize the fact that nonlinear effects significantly modify the anisotropic behavior of soft solids.

Moreover one can notice that when μ grows, A increases going from a negative value to a positive one for both AG phantoms and PVA phantoms. This is directly related to the slope variation of the shear wave speed along direction 21 which changes of sign, from positive to negative, as a function of the shear modulus increase. Consequently for certain shear modulus μ at zero stress, e.g., a certain concentration of agar-gelatin or a certain number of freezing–thawing cycles in PVA phantoms, the nonlinear shear modulus A is cancelled. Regarding Eqs. (13)–(15), anisotropy can be also retrieved if the nonlinear coefficient A is cancelled. In this specific configuration, the slope of the measurement for the couple of indices 21 [Eq. (14)] is null. The stress applied has no more effect on the shear wave velocity. But stress still induces anisotropy for both other couple of indices [Eqs. (13) and (15)]. Shear wave velocity in the 13 direction increases whereas shear wave velocity in 12 direction decreases with the applied stress. In fact, such an $A=0$ assumption means that the third invariant of the Lagrangian strain tensor I_3 has no effect in the development of the strain energy [Eq. (4)] (if one considers only the development of the strain energy at the third order). So, in that particular configuration, the only contribution to nonlinearity is the quadratic term from the strain tensor, also known as the geometric nonlinearity.

From Fig. 7, one can notice that the nonlinear parameter A seems to evolve as an exponential regarding to the shear modulus for the AG phantoms and it seems not to be the case for the PVA phantoms. These effects can be intuitively explained by the differences between microstructures of both materials. On the one hand, the elasticity results from the response of the polymeric filaments between cross-links,

from alterations in the network structure, or both.³⁷ On the other hand, the principles responsible for the nonlinear elasticity due to the microstructure are unknown, theories show that systems of filamentous proteins arranged in an open crosslinked mesh invariably stiffen at low strains without requiring specific architecture or multiple elements with different intrinsic stiffness.³⁸ Such behavior could be strongly interacting in such PVA phantoms known to contain crosslinked filaments. Moreover, as a comparison in classical solids such as glasses, the different constituent of the glasses used in the experiments can be related to velocity variations observed for both longitudinal and transversal waves in an acoustoelasticity experiments.^{39,40} However, the understanding of the behavior of the nonlinear shear modulus (A) as a function of the shear modulus (μ) needs further investigation.

VI. CONCLUSION

In this paper, a fast, simple, reproducible and robust method based on the supersonic shear imaging technique was used to revisit the acoustoelasticity experiment in order to characterize the nonlinear shear behavior of soft solids. This ultrasound-based technique enables both the remote generation in soft solids of quasiplane polarized shear waves and the ultrafast imaging of their resulting propagation. It enables the local assessment (or mapping) of the third order shear elastic coefficient A . Using an expression of the strain energy derived by Hamilton *et al.*²⁰ and Zabolotskaya *et al.*²¹ in quasiincompressible media, this third order coefficient can be estimated from a single ultrasonic experiment. Instead of measuring shear wave speeds for three different polarizations in order to completely determine the nonlinearity of standard solids, one single polarization experiment under a varying applied uniaxial stress is sufficient in soft solids to characterize the nonlinear shear elasticity. It means that *in vivo* assessment of this parameter could be feasible within the framework of the supersonic shear imaging approach.

ACKNOWLEDGMENTS

The authors would like to gratefully thank Patricia Daenens and Marc Yvert for technical assistance and Daniel Royer for illuminating discussion.

- ¹T. A. Krouskop, B. S. Dougherty, and F. S. Vinson, "A pulsed Doppler ultrasonic system for making noninvasive measurements of the mechanical properties of soft tissue," *J. Rehabil. Res. Dev.* **24**, 1–8 (1987).
- ²R. M. Lerner, K. J. Parker, J. Holen, R. Gramiak, and R. C. Waag, "Sonoelasticity: Medical elasticity images derives from ultrasound signals in mechanically vibrated targets," *Acoust. Imaging* **16**, 317–327 (1988).
- ³Y. Yamakoshi, J. Sato, and T. Sato, "Ultrasonic imagins of internal vibration of soft tissue under forced vibration," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **37**, 45–53 (1990).
- ⁴K. J. Parker and R. M. Lerner, "Sonoelasticity of organs: Shearwaves ring a bell," *J. Ultrasound Med.* **11**, 387–392 (1992).
- ⁵S. F. Levinson, M. Shinagawa, and T. Sato, "Sonoelastic determination of human skeletal muscle elasticity," *J. Biomech.* **28**, 1145–1154 (1995).
- ⁶J. Bercoff, S. Chaffai, M. Tanter, L. Sandrin, S. Catheline, M. Fink, J.-L. Gennisson, and M. Meunier, "*In vivo* breast tumor detection using transient elastography," *Ultrasound Med. Biol.* **29**, 1387–1396 (2003).
- ⁷L. Sandrin, B. Fourquet, J.-M. Hasquenoph, S. Yon, C. Fournier, F. Mal, C. Christidis, M. Ziol, B. Poulet, F. Kazemi, M. Beauprand, and R. Palau, "Transient elastography: a new non invasive method for assessment of

hepatic fibrosis," *Ultrasound Med. Biol.* **29**, 1705–1713 (2003).

- ⁸J.-L. Gennisson, S. Catheline, and M. Fink, "Transient elastography in anisotropic medium: application to the measurement of slow and fast shear wave speeds in muscles," *J. Acoust. Soc. Am.* **114**, 536–541 (2003).
- ⁹S. Catheline, J.-L. Gennisson, G. Delon, R. Sinkus, M. Fink, S. Abouelkaram, and J. Culioli, "Measurement of viscoelastic properties of homogeneous soft solid using transient elastography: An inverse problem approach," *J. Acoust. Soc. Am.* **116**, 3734–3741 (2004).
- ¹⁰J. Bercoff, M. Tanter, M. Muller, and M. Fink, "The role of viscosity in the impulse diffraction field of elastic waves induced by the acoustic radiation force," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **51**, 1523–1536 (2004).
- ¹¹S. Catheline, J.-L. Gennisson, and M. Fink, "Measurement of elastic nonlinearity of soft solids with transient elastography," *J. Acoust. Soc. Am.* **114**, 3087–3091 (2003).
- ¹²S. Catheline, J.-L. Gennisson, M. Tanter, and M. Fink, "Observation of shock transverse waves in elastic media," *Phys. Rev. Lett.* **91**, 43011–43014 (2003).
- ¹³X. Jacob, J.-L. Gennisson, S. Catheline, M. Tanter, C. Barrière, D. Royer, and M. Fink, "Study of elastic nonlinearity of soft solids with transient elastography," *Proc.-IEEE Ultrason. Symp.* **1**, 660–663 (2003).
- ¹⁴D. S. Hugues and J. L. Kelly, "Second-order elastic deformation of solids," *Phys. Rev.* **92**, 1145–1149 (1953).
- ¹⁵T. D. Murnaghan, *Finite Deformation of an Elastic Solid* (Wiley, New York, 1951).
- ¹⁶A. Norris, *Non Linear Acoustics: Finite Amplitude Waves in Solids*, edited by M. Hamilton (Academic, New York, 1998).
- ¹⁷H. Kobayashi and R. Vanderby, "New strain energy function for acoustoelasticity analysis of dilatational waves in nearly incompressible hyperelastic materials," *J. Appl. Mech.* **72**, 843–851 (2005).
- ¹⁸H. Kobayashi and R. Vanderby, "Acoustoelastic analysis of reflected waves in nearly incompressible, hyper-elastic materials: Forward and inverse problem," *J. Acoust. Soc. Am.* **121**, 879–887 (2007).
- ¹⁹R. Q. Erkamp, A. R. Skovoroda, S. Y. Emelianov, and M. O'Donnell, "Measuring the nonlinear elastic properties of tissue-like phantoms," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **51**, 410–419 (2004).
- ²⁰M. F. Hamilton, Y. A. Ilinskii, and E. A. Zabolotskaya, "Separation of compressibility and shear deformation in the elastic energy density (L)," *J. Acoust. Soc. Am.* **116**, 41–44 (2004).
- ²¹E. A. Zabolotskaya, M. F. Hamilton, Y. A. Ilinskii, and G. D. Meegan, "Modeling of non linear shear waves in soft solids," *J. Acoust. Soc. Am.* **116**, 2807–2813 (2004).
- ²²J. Bercoff, M. Tanter, and M. Fink, "Supersonic shear imaging: a new technique for soft tissues elasticity mapping," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **51**, 374–409 (2004).
- ²³L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 3rd ed. (Butterworth-Heinemann, Oxford, 2002).
- ²⁴D. Royer and E. Dieulesaint, *Elastic Waves in Solid* (Springer, Berlin, 1996), Vol. 1.
- ²⁵P. A. Johnson and P. Rasolofosaon, "Nonlinear elasticity and stress-induced anisotropy in rock," *J. Geophys. Res.* **101**, 3113–3124 (1995).
- ²⁶S. Kostek, B. K. Sinha, and A. N. Norris, "Third-order elastic constants for an inviscid fluid," *J. Acoust. Soc. Am.* **94**, 3014–3017 (1993).
- ²⁷R. N. Thurston and K. Brugger, "Third-order elastic constants and the velocity of small amplitude elastic waves in homogeneously stressed media," *Phys. Rev.* **133**, A1604–A1610 (1964).
- ²⁸J.-L. Gennisson and G. Cloutier, "Sol-gel transition in agar-gelatin mixtures studied with transient elastography," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **53**, 716–723 (2006).
- ²⁹J. Fromageau, E. Brusseau, D. Vray, G. Gimenez, and P. Delachartre, "Characterization of PVA cryogel for intravascular ultrasound elasticity imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **50**, 1318–1324 (2003).
- ³⁰J. Fromageau, J.-L. Gennisson, C. Schmitt, R. Maurice, R. Mongrain, and G. Cloutier, "Estimation of polyvinyl alcohol cryogel mechanical properties with 4 ultrasound elastography methods and comparison with gold standard testings," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **54**, 498–509 (2007).
- ³¹L. Sandrin, M. Tanter, S. Catheline, and M. Fink, "Shear modulus imaging using 2d transient elastography," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **49**, 426–435 (2002).
- ³²J. Ophir, I. Céspedes, H. Ponnekanti, Y. Yasdi, and X. Li, "Elastography: a quantitative method for imaging the elasticity of biological soft tissues," *Ultrason. Imaging* **13**, 111–134 (1991).

- ³³M. A. Breazeale, "Comparison of the non linear behavior of fluids and solids," *Proceedings of the 13th ISNA*, World Scientific, Singapore, 1993, p. 451.
- ³⁴K. Naugolnykh and L. Ostrovsky, *Nonlinear Wave Processes in Acoustics* (Cambridge University Press, New York, 1998), Chap. 1, pp. 15–16.
- ³⁵C. M. Hassan and N. A. Peppas, "Structure and applications of polyvinyl alcohol hydrogels produced by conventional crosslinking or by freezing/thawing methods," *Adv. Polym. Sci.* **153**, 37–65 (2000).
- ³⁶C. Joly-Duhamel, D. Hellio, A. Ajdari, and M. Djabourov, "All gelatin networks: 2. The master curve for elasticity," *Langmuir* **18**, 7158–7166 (2002).
- ³⁷P. R. Onck, T. Koeman, T. van Dillen, and E. vander Giessen, "Alternative explanation of stiffening in cross-linked semiflexible networks," *Phys. Rev. Lett.* **95**, 178102 (2005).
- ³⁸C. Storm, J. J. Pastore, F. C. MacKintosh, T. C. Lubensky, and P. A. Janmey, "Nonlinear elasticity in biological gels," *Nature (London)* **435**, 191–194 (2005).
- ³⁹R. J. Wang, W. H. Wang, F. Y. Li, L. M. Wang, Y. Zhang, P. Wen, and J. F. Wang, "The Gruneisen parameter for bulk amorphous materials," *J. Phys.: Condens. Matter* **15**, 603–608 (2003).
- ⁴⁰B. Zhang, R. J. Wang, and W. H. Wang, "Response of acoustic and elastic properties to pressure and crystallization of Ce-based bulk metallic glass," *Phys. Rev. B* **72**, 104205 (2005).

A two-dimensional pseudospectral model for time reversal and nonlinear elastic wave spectroscopy

Thomas Goursolle,^{a)} Samuel Callé, and Serge Dos Santos

Université François-Rabelais de Tours, LUSSE/CNRS FRE-2448, ENIVL rue de la Chocolaterie, BP 3410, 41034 Blois Cedex, France

Olivier Bou Matar

European Associated Laboratory in Nonlinear Magneto-Acoustic (LEMAC) - IEMN – DOAE, UMR CNRS 8520, Ecole Centrale de Lille, BP 48, 59651 Villeneuve D'ASCQ Cedex, France

(Received 7 February 2007; revised 7 September 2007; accepted 25 September 2007)

One way to characterize metallic materials in the presence of defects like dislocation networks is to measure their large dynamic nonlinear elastic response. In this numerical study, a new method combining the nonlinear elastic wave spectroscopy (NEWS) method with a time reversal (TR) process is proposed. This method, called NEWS-TR, uses nonlinear analysis as a pretreatment of time reversal and then consists of retrofocusing only nonlinear components on the defect position. A two-dimensional pseudospectral time domain algorithm is developed here to validate the NEWS-TR method as a potential technique for damage location. Hysteretic nonlinear behavior of the materials being studied is introduced using the Preisach-Mayergoyz model. Moreover, in order to extend this solver in two dimensions, the Kelvin notation is used to modify the elastic coefficient tensor. Simulations performed on a metallic sample show the feasibility and value of the NEWS-TR methodology for microdamage imaging. Retrofocusing quality depends on different parameters such as the filtering method used to keep only nonlinear components and the nonlinear effect measured. In harmonic generation, pulse inversion filtering seems to be a more appropriate filtering method than classical harmonic filtering for most defect positions, mainly because of its ability to filter all fundamental components. © 2007 Acoustical Society of America.
[DOI: 10.1121/1.2799900]

PACS number(s): 43.25.Dc, 43.40.Le, 43.60.Tj [DRD]

Pages: 3220–3229

I. INTRODUCTION

It has been known for several years that cracked materials contain small soft features (dislocations, grain contacts, cracks, etc.) within a hard matrix (crystals, grains, etc.) producing very large nonlinear response.¹ Dynamic nonlinear responses may manifest themselves in a variety of ways,² including resonance frequency shifts, harmonic generation, frequency mixing, nonlinear attenuation, and slow dynamic effects.³ These elastic behaviors, widely studied in geophysics,^{4–6} have had wide ranging implications in nondestructive evaluation for many years.⁷ One of the most frequently studied methods is harmonic analysis in the frequency domain,^{8,9} consisting of measuring the second and higher harmonic amplitudes versus the strain amplitude of the fundamental, which provides quantitative information about the nature of the nonlinearity. Another technique consists of parametric interactions between waves emitted in materials. The sample acts as a nonlinear frequency mixer, so that sum and difference frequency waves are created.¹⁰ In addition, modulation of amplitude¹¹ and phase¹² have been investigated in order to evaluate the classical nonlinear perturbation coefficient β coming from the Taylor expansion of stress-strain relation.¹³ The study of resonance frequency provides key information about nonlinear behavior. By plot-

ting the frequency shift as a function of the fundamental-mode strain amplitude, it is observed that resonance amplitude distortion increases significantly with defects.^{7,14,15} These methods are referred to as nonlinear elastic wave spectroscopy (NEWS) methods.⁷

It is well known for a long time that cracks or dislocation can induce, in fatigued metallic samples, some localized hysteretic deformation.^{16–18} This behavior, also observed in granular media,^{3,19} is generally called “nonclassical nonlinear effect.” Although NEWS methods allow characterization of this behavior, they do not provide information about defect localization. To overcome this problem, a method combining a time reversal (TR) process and a nonlinear treatment which has been first proposed by Bou Matar *et al.*²⁰ (jointly with Gliozzi *et al.*²¹) is explored here more in details.

Developed at the beginning of the 1990s,^{22,23} the TR process involves a propagating elastic wave that is reversed in time and reemitted back by receivers toward the source from which it was generated. As time reversal invariance is verified by the wave equation in nonattenuating media,²² the reemitted field is expected to retrofocus on the source and reconstruct the signal.²⁴ This method has been developed in other domains such as ocean acoustics,²⁵ particle chaos study,²⁶ and medicine,²⁷ and performed in granular media.²⁸ For further details, excellent reviews can be found on this subject.^{22,29} The TR process in the nonlinear regime has recently been demonstrated to be highly valuable for ultra-

^{a)}Electronic mail: thomas.goursolle@univ-tours.fr

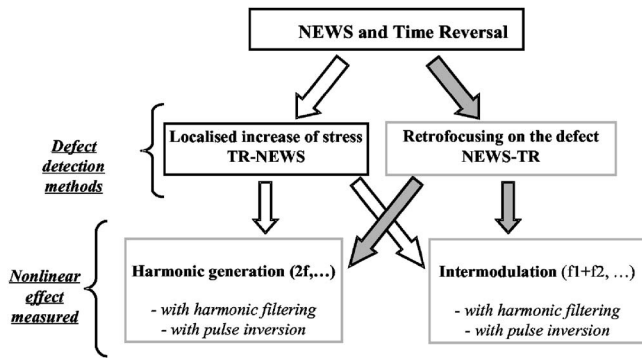


FIG. 1. Representation of two detection techniques: Localized increase in stress (TR-NEWS) and retrofocusing on the defect (NEWS-TR). For each technique the nonlinear effects measured can be harmonic generation and intermodulation with different kinds of filtering of nonlinear components (harmonic filtering and pulse inversion).

sound imaging of damage in solids.^{20,21,30} Moreover, similar ideas have already been used in fluids where wave phase conjugation (or time reversal) in nonlinear regime has been demonstrated highly useful in ultrasonic imaging.³¹

Two methods have been derived from the nonlinearity-based time reversal concept, depending on whether nonlinear treatment is performed before or after the TR process. As presented in Fig. 1, these two methodologies are defined as TR-NEWS, with nonlinear analysis as a posttreatment of time reversal, and as NEWS-TR, with nonlinear analysis as a pretreatment of time reversal.³²

The TR-NEWS method, which consists of locally increasing the stress field using properties of linear TR and subsequently applying nonlinear analysis, has been experimentally demonstrated by Sutin *et al.*³⁰ It seems to have a wide potential for application in solid ultrasound imaging for nondestructive testing.^{32,33} The NEWS-TR application is based on signal retrofocusing on the defect position when only the nonlinear components of the received signal are time reversed.^{34–36} This method, described for the first time by Bou Matar *et al.*²⁰ in 2005, which is the object of this study, has only been validated experimentally recently.³⁷ A numerical study based on a two-dimensional (2D) pseudospectral time domain algorithm is proposed here to validate the NEWS-TR method as a potential technique for damage location.

The nonlinear elastic wave solver used in this simulation study is described in Sec. II. The one-dimensional (1D) equations of wave propagation in heterogeneous media with nonclassical nonlinearity are presented. Hysteretic nonlinear behavior is introduced using the Preisach–Mayergoyz (PM) model.^{19,38,39} Section III shows the numerical results for NEWS-TR simulations in a metallic sample. After presentation of simulation parameters and linear TR validation, the nonlinear signature of harmonic generation is tested. Two filtering methods are investigated for these signatures, i.e., classical harmonic filtering and pulse inversion filtering. The retrofocusing qualities of these different configurations are compared.

II. THE NONLINEAR ELASTIC WAVE SOLVER

Here the pseudospectral wave solver used for 2D simulation of the nonlinear wave propagation in damaged plates

is described in detail. First, cracked zones, corresponding to sources of the nonlinear hysteretic wave propagation process occurring in the specimen, as introduced by Van den Abeele *et al.*¹⁵ will be introduced in 1D. It is then shown how this model can be extended to 2D with the use of the Kelvin notation. Finally, a brief description of the numerical technique is given.

A. 1D model equation

In a heterogeneous medium in which a compressional wave propagates, the propagation is modeled by the 1D hyperbolic system of conservation laws:

$$\frac{\partial v}{\partial t} = \frac{1}{\rho_0} \frac{\partial \tau}{\partial z}, \quad (1a)$$

$$\frac{\partial \tau}{\partial t} = K(t) \frac{\partial v}{\partial z}, \quad (1b)$$

where ρ_0 and $K(t)$ are the volumetric mass density and the elastic modulus of the material, respectively, v is the particle velocity, and τ the longitudinal stress. Here the modulus $K(t) = \partial \tau / \partial \varepsilon$, where ε is the strain, is considered as time dependent in order to introduce the nonlinearity, which can be considered of a different kind (e.g., quadratic, bimodular, or hysteretic). Here a hysteretic nonlinearity model, based on Preisach–Mayergoyz space (PM space),³⁸ first introduced by McCall and Guyer,^{1,39} has been used. In this model no analytical expression of the bulk modulus is given. It is calculated by summation of the strain contribution of a large number of elementary hysteretic elements. Each of these hysteretic element units (HEU) is described by two characteristic pore pressures, P_c and P_o , corresponding to the transition between two states when the pore pressure is increased or decreased, respectively. Here a pore pressure P , defined as the negative of the local stress, has been introduced in order to use the same formalism and notations as Guyer *et al.*^{1,39} One state corresponds to an “open” state and the other to a “closed” state. Our implementation of the PM space model is based on the multiscale approach developed by Van Den Abeele *et al.*¹⁵ For each cell of the calculation grid (representing a mesoscopic level of the medium description), N_0 hysteretic units are considered with different values of the two characteristic stresses. This representation is commonly termed “PM space” and can be described mathematically by its density distribution $f(P_c, P_o)$.

The elementary hysteretic elements, which have been considered, are shown on Fig. 2(a). It is an extension of the model of Scalerandi *et al.*⁴⁰ implemented with the LISA code shown in Fig. 2(b). In Fig. 2, K_1 and K_2 correspond to the bulk modulus of the HEUs in the open and closed states, respectively. Moreover γ_1 and γ_2 are the strain jumps, which occur at the HEUs opening (respectively, closing) when the pore pressure is decreased (respectively, increased). It can be shown that in this model the bulk modulus is given by¹⁵

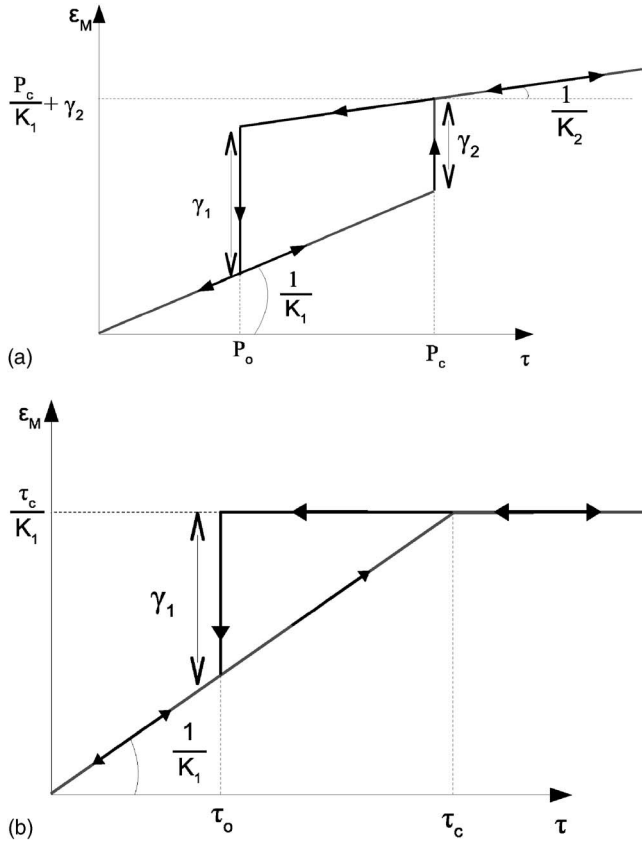


FIG. 2. Elementary hysteretic elements used in the computation of the PM space model. (a) “Elastic” two-state elements and (b) hysteretic element proposed by Scalerandi *et al.* (Ref. 40). In each case, one state corresponds to an open state and the other to a closed state, each with an associated transition stress, P_o and P_c , respectively.

$$\frac{1}{K} = \frac{1}{E(1 + \beta\tau + \delta\tau^2 + \dots)} + \frac{\partial \varepsilon_H}{\partial \tau}, \quad (2)$$

where ε_H is the strain contribution of the hysteretic elements, E the linear elastic modulus, and β and δ are “classical” quadratic and cubic nonlinear parameters. In order to calculate the strain variation as a function of stress induced by the HEUs, the method proposed by Aleshin *et al.*⁴¹ has been used. The first step is to write the strain rate in time as a function of stress as a weighted sum (over the PM space) of the strain contribution ε_M of each individual HEU:

$$\frac{\partial \varepsilon_H}{\partial \tau} = - \int_{-\infty}^{P_c} dP_o \int_{P_o}^{+\infty} dP_c f(P_o, P_c) \frac{\partial \varepsilon_M}{\partial P}. \quad (3)$$

As shown in Fig. 2(a), the strain variation ε_M for each elementary HEU (if the pore pressure increases [$\partial P / \partial t > 0$]) is given by

$$\begin{aligned} \frac{\partial \varepsilon_M}{\partial P} &= -\frac{1}{K_1} \quad \text{if HEU } M \text{ is open,} \\ \frac{\partial \varepsilon_M}{\partial P} &= -\frac{1}{K_2} \quad \text{if HEU } M \text{ is closed.} \end{aligned} \quad (4)$$

Introducing these expressions in Eq. (3), the strain variation induced by all the HEUs becomes

$$\begin{aligned} \frac{\partial \varepsilon_H}{\partial P} &= - \int_{-\infty}^P dP_o \gamma_2 f(P_o, P) \Omega - \frac{1}{K_2} \\ &\times \int_{-\infty}^{P_c} dP_o \int_{P_o}^{+\infty} dP_c f(P_o, P_c) - \left(\frac{1}{K_1} - \frac{1}{K_2} \right) \\ &\times \int_{-\infty}^{P_c} dP_o \int_{P_o}^{+\infty} dP_c u(P_c - P) f(P_o, P_c) \Omega, \end{aligned} \quad (5)$$

with $\Omega=1$ if the HEU is open, and $\Omega=0$ if it is closed. Here $u(x)$ is the step function. Finally,

$$\begin{aligned} \frac{\partial \varepsilon_H}{\partial P} &= - \int_{O_2}^P dP_o \gamma_2 f(P_o, P) \\ &- \frac{1}{K_2} \int_{-\infty}^{P_c} dP_o \int_{P_o}^{+\infty} dP_c f(P_o, P_c) \\ &- \left(\frac{1}{K_1} - \frac{1}{K_2} \right) \int_{O_2}^{P_c} dP_o \int_P^{+\infty} dP_c f(P_o, P_c), \end{aligned} \quad (6)$$

where O_2 is the P_o value for the HEU individuated by the point A_2 in the PM space (Fig. 3). The first double integral term is the inverse of the bulk modulus contribution when all the HEUs are in the closed state. The second double integral term is the modification of the inverse of the bulk modulus due to open HEUs, and the single integral term is the additional contribution to the inverse of the bulk modulus of the HEUs closing at the actual pore pressure P . When the pore pressure is decreased it can be shown similarly that

$$\begin{aligned} \frac{\partial \varepsilon_H}{\partial P} &= - \int_P^{C_2} dP_c \gamma_1 f(P, P_c) \\ &- \frac{1}{K_1} \int_{-\infty}^{P_c} dP_o \int_{P_o}^{+\infty} dP_c f(P_o, P_c) \\ &- \left(\frac{1}{K_2} - \frac{1}{K_1} \right) \int_{-\infty}^{P_c} dP_o \int_{P_o}^{C_2} dP_c f(P_o, P_c), \end{aligned} \quad (7)$$

where now C_2 is the P_c value for the HEU identified by the point A_2 in the PM space, and the first double integral term is the inverse of the bulk modulus contribution when all the HEUs are in the open state. The second double integral term is the modification of the inverse of the bulk modulus due to closed HEUs, and the single integral term is the additional contribution to the inverse of the bulk modulus of the HEUs opening at the actual pore pressure P . In cases of both increasing and decreasing stress, which is the inverse of the introduced pore pressure, the variations in $\partial \varepsilon_H / \partial \tau$ can be calculated by looking at the closing or opening HEUs in the PM space between the current and past value of the stress, respectively.

In the case shown in Fig. 2(b) the following equations have to be used to simplify Eqs. (6) and (7):

$$\gamma_1 = \gamma = \frac{P_c - P_o}{K_1}, \quad \gamma_2 = 0, \quad \frac{1}{K_2} = 0. \quad (8)$$

The choice of this kind of triangular (or trapezoidal) HEU is due to the fact that “classical” rectangular HEUs only pro-

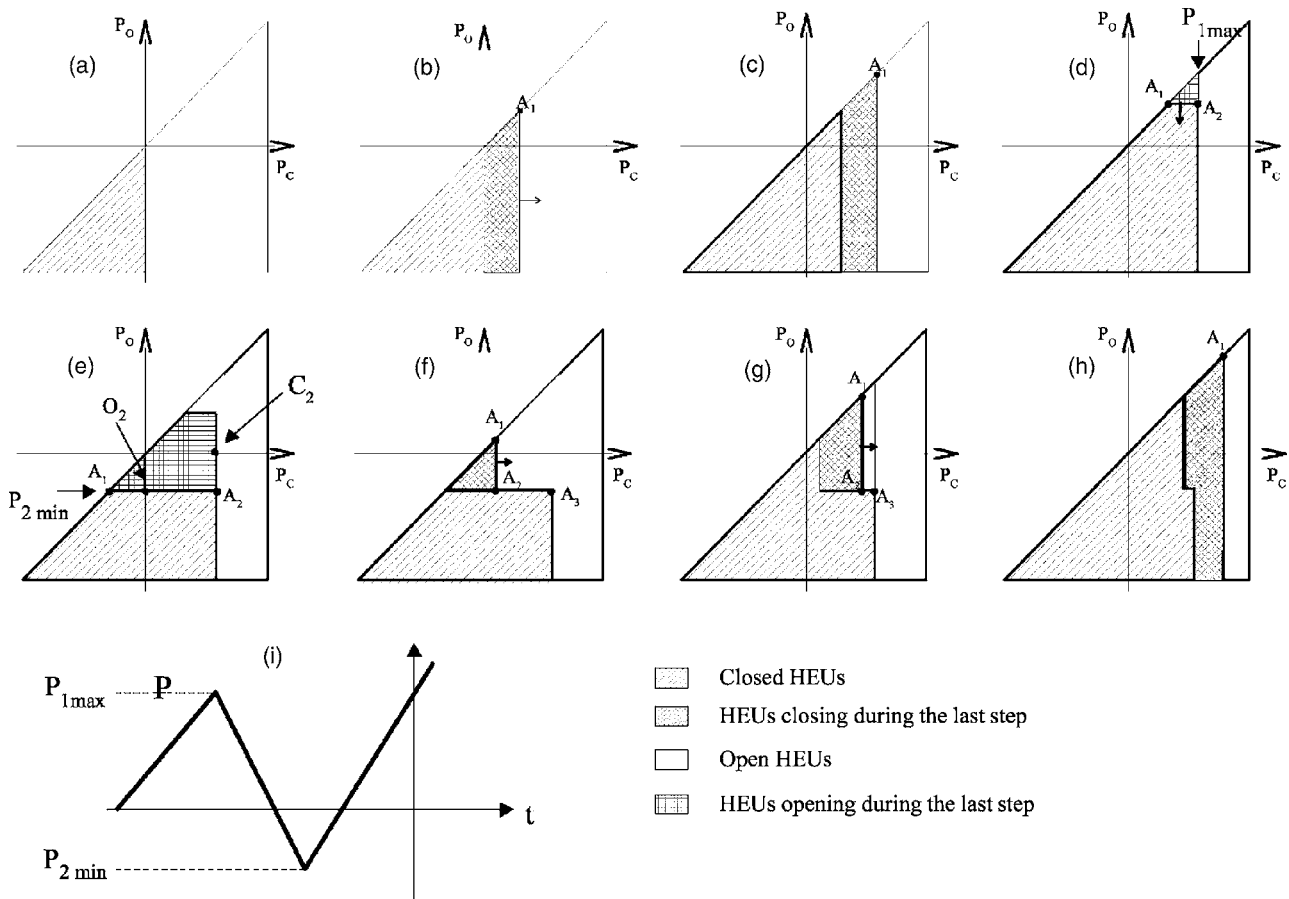


FIG. 3. (a)–(h) Evolution of the PM space domain during cycles of compression and rarefaction presented in (i).

duce odd harmonics when nonlinear wave distortions are simulated.¹⁹ However in this study the pulse inversion technique was also investigated, and as explained in Sec. III, this method is linked to the second harmonic content of the signal. Moreover, in experiments even harmonics are generally also measured, although their amplitudes are small in comparison with those of odd harmonics.⁴

B. 2D Kelvin notation model equation

When considering the propagation of elastic waves in a heterogeneous solid medium, Newton’s second law is written:

$$\frac{\partial v_i}{\partial t} = \frac{1}{\rho_0} \frac{\partial \tau_{ij}}{\partial x_j}, \quad (9)$$

where x_j are the components of the position vector, ρ_0 is the density, v_i are the components of the particle velocity vector, τ_{ij} are the components of the stress tensor, and t is time. The constitutive equation for a linear elastic solid is given by the Hooke’s law:

$$\tau_{ij} = C_{ijkl} \varepsilon_{kl}, \quad (10)$$

where ε is the strain tensor and C_{ijkl} are the elastic coefficients. Here the propagation of P – S_V waves in 2D transverse isotropic solid is considered where the stress–strain relation is given by

$$\begin{pmatrix} \tau_{xx} \\ \tau_{zz} \\ \tau_{xz} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{13} & 0 \\ C_{13} & C_{33} & 0 \\ 0 & 0 & C_{44} \end{pmatrix} \begin{pmatrix} \varepsilon_{xx} \\ \varepsilon_{zz} \\ \varepsilon_{xz} \end{pmatrix}, \quad (11)$$

where here the elastic matrix is given in the Voigt notation. The 2D elastic wave solver used in this study is based on a Kelvin notation method.^{42,43} Using this notation, the three eigenvectors of the elastic constants tensor (in 2D) correspond to three eigenstress/eigenstrain vectors. These vectors represent directions where applied stress and response strain are in the same direction. Doing so, it is possible to use for each of these three directions a scalar “PM space” model similar to the one used for 1D simulations. Each PM space associated with the different directions modifies the associated eigenstiffness. Finally, with these new eigenstiffness, the “effective” elastic coefficients are consequently calculated at each time step. Some details on the implementation procedure are described in the Appendix.

C. Pseudospectral wave solver

The numerical technique used to solve the system of differential equations presented in the preceding part is a pseudospectral wave solver developed in our laboratory.⁴⁴ This method consists of computing spatial derivatives in the Fourier domain by means of a fast Fourier transform (FFT):

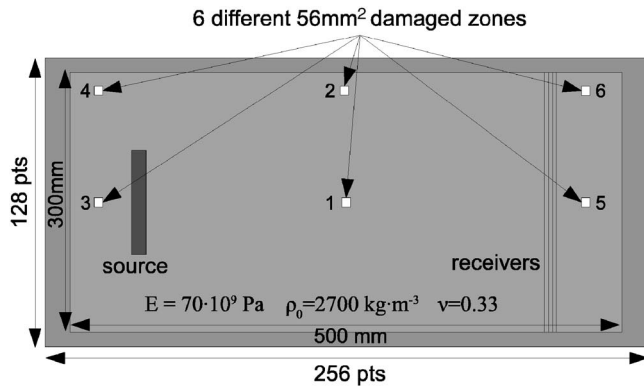


FIG. 4. Geometry of simulation domain with characteristics of source, receivers, defect positions, and material (where E is the Young modulus, ρ_0 the volumetric mass density, and ν the Poisson's ratio).

$$\frac{\partial f(x,y,t)}{\partial x} = \text{FFT}^{-1} \left[\left(jk_x \exp\left(\pm jk_x \frac{\Delta x}{2}\right) \right) \text{FFT}[f(x,y,t)] \right], \quad (12)$$

where FFT^{-1} is the inverse fast Fourier transform and k_x is the wave number in x direction. Δx is the spatial step of the numerical grid in x direction. To reduce numerical artifacts in pseudospectral simulations for a heterogeneous medium, a staggered spatial grid implementation is used. Moreover, a staggered fourth-order Adams-Bashforth method⁴⁵ is used to update stress and particle velocity at alternating half time steps. The main advantage of this method for nonlinear elastic wave simulation is that only two node points are needed per minimum wavelength, even for long time propagation.⁴⁶ In this case, where frequencies up to the fifth harmonic are considered, only ten node points per wavelength are enough. This can be compared to the almost 50 node points which are needed in finite element or finite difference time domain methods. More details on the pseudospectral wave solver implemented can be found in Ref. 44.

III. NUMERICAL RESULTS

Simulation domain and excitation parameters are first described and then the results obtained for NEWS-TR method are developed. Generation of the harmonic frequency signal (called harmonic generation) was tested for nonlinear signature. Two filtering methods were investigated for each signature, i.e., classical harmonic filtering and the pulse inversion method (also known as the phase-coded pulse sequence).

A. Simulation parameters and linear validation

1. Simulation parameters

Figure 4 shows the geometry (source, receivers and defect positions) and some of physical properties (Young modulus, volumetric mass density and Poisson's ratio) of the 2D specimen [aluminum sample (300 mm × 500 mm)]. The realization of virtual TRA experiments is finalized to prove the efficacy of the technique in retrofocusing on the defect position making use of only the nonlinear components of the wave fields:

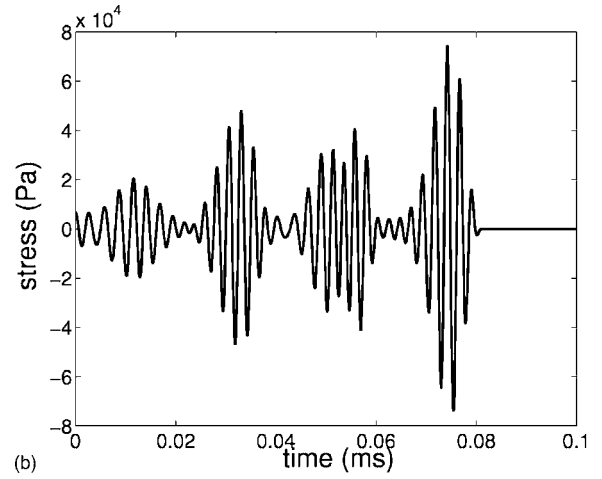
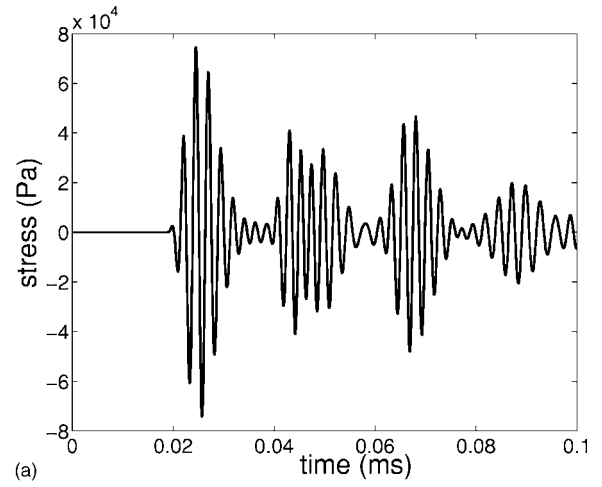


FIG. 5. Temporal signal received at one point on one line (a) and returned (b).

$$f(t) = \pm p_0 \sin(2\pi f_c t) e^{-(1/2)(t/w)^2} (1 - e^{-(t/s_w)^2}) \times e^{-(1/2)((x-x_s)/w_x)^2} e^{-(1/2)((y-y_s)/w_y)^2}, \quad (13)$$

where $f_c=200$ kHz is the source frequency, $p_0=50$ kPa the initial pressure, and $w=2/f_c$, $s_w=3/f_c$, $w_x=0.15$, $w_y=0.75$, $x_s=12.5$ cm and $y_s=15$ cm are emission parameters.

The receivers are composed of four different lines (1D array of 112 elements) and simulations with three different length receivers (300, 165, and 50 mm) have been done. The 56 mm² damaged zone (squares on bulks in Fig. 4) presents hysteretic nonlinear behavior and acoustical properties of the undamaged zone are linear. Six different defect positions are defined to make it possible to study their influence on the retrofocusing quality (Fig. 4).

2. Linear time reversal validation

The signal emitted by the source with the form presented is received by the four lines, time reversed and emitted back by the receivers (Fig. 5). During the retropropagation in the sample with a central defect (position 1), the maximum amplitude of stress along the z axis in time obtained for each point of the simulation grid is stored in a matrix.

This matrix presented in Fig. 6 shows that with the linear time reverse principle, the wave retrofocuses exactly on

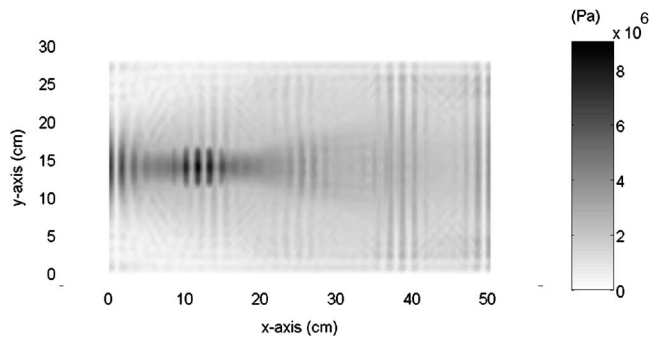


FIG. 6. Matrix of the maximum amplitude of stress along z -axis obtained during retropropagation of the returned signal (Figure 5(b)).

the source position and not on the defect.²⁹ The size of the focal spot can be directly linked with source size and form, and with frequency of the signal emitted.

3. Filtering methods

Two filtering methods were investigated to return only the nonlinear parts (harmonics) of the received signal for NEWS-TR method, i.e., harmonic filtering and pulse inversion.

a. Harmonic filtering. One option consists of selecting only the nonlinear/harmonic energy contained in the response signals and returning only this part back into the medium by the time reverse mirror. Hysteretic nonlinearity exhibiting high level of odd harmonics, the third harmonic is extracted, in the next simulations, from the received signal using a fourth-order Butterworth filter. The choice of this filter is justified by the perspective to use the NEWS-TR method in experiments.

In Fig. 7, the profiles of the maxima-in-time of the

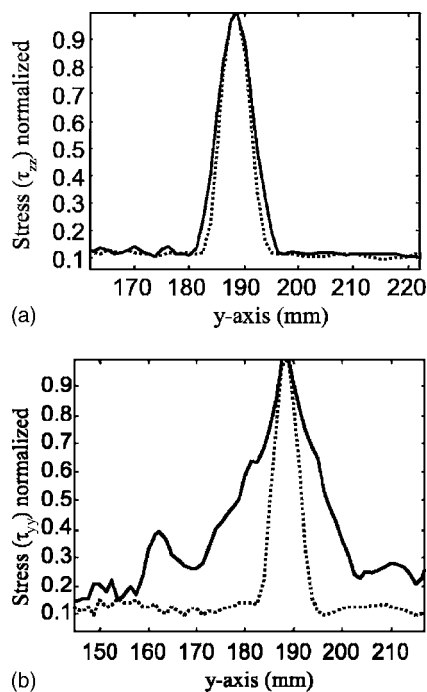


FIG. 7. Comparison of stress amplitude (along z axis) retrofocusing spot size (a) between a 600 kHz (third harmonic) (solid line) and a 1 MHz (fifth harmonic) (dotted line) returned signal for a 300 mm source and (b) between a 50 mm source (solid line) and a 165 mm source (dotted line).

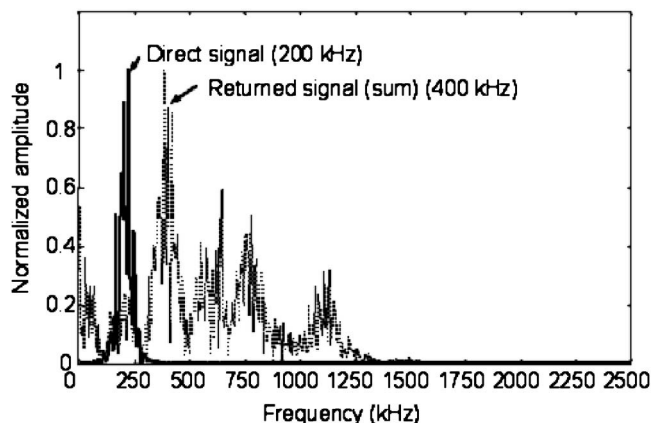


FIG. 8. Comparison between spectral representations of direct signal and returned signal (sum).

TR backpropagated wave field evaluated along the y axis across the point with coordinates (190; 160) mm is presented at two different frequencies [Fig. 7(a)] and for two different source sizes [Fig. 7(b)]. As expected, the higher the frequency, the greater the increase in retrofocusing quality (size of the focusing spot). Moreover, decreasing the source size reduces the retrofocusing quality.

b. Pulse inversion. An alternative filtering procedure is based on the fact that the phase inversion of a pulsed excitation signal (180° phase shift) will lead to the exact inverted phase response signal within a linear medium.⁴⁷ This is not the case in a nonlinear (or microdamaged) material due to the generation of harmonics. Advantage of this information is taken by adding the responses from two phase-inverted pulses (positive and negative) and sending back the sum to the receivers. This operation is called “phase-coded pulse-sequence (PC-PS) filtering” or “pulse inversion (PI).” Only the relevant information on the local nonlinearities is reversed and sent back into the material.

B. Nonlinear time reversal simulations of harmonic generation

In order to localize possible defects in the plate, it is essential to combine NEWS with TR. The response signal, which has been propagated in the sample and received by the four lines of receivers (see Sec. II A 1), is once again time reversed, but now only after filtering out the linear component and retaining the nonlinear part of the response. The signature of nonlinearity, which was studied, was classical harmonic generation.

1. Results

Simulations were performed with defects in different positions and for both methods of filtering. For the harmonic filtering method only the third-harmonic components in the response signals were returned into the bulk to retrofocus on the defect position.

For the PI method, the response signals for two opposite waves were calculated from the emitter to receivers and the sum of the two signals received was returned. Figure 8 shows the spectral representation of one of the received signals (direct signal) and the sum of the two received signals (returned signal). The received signal frequency was about

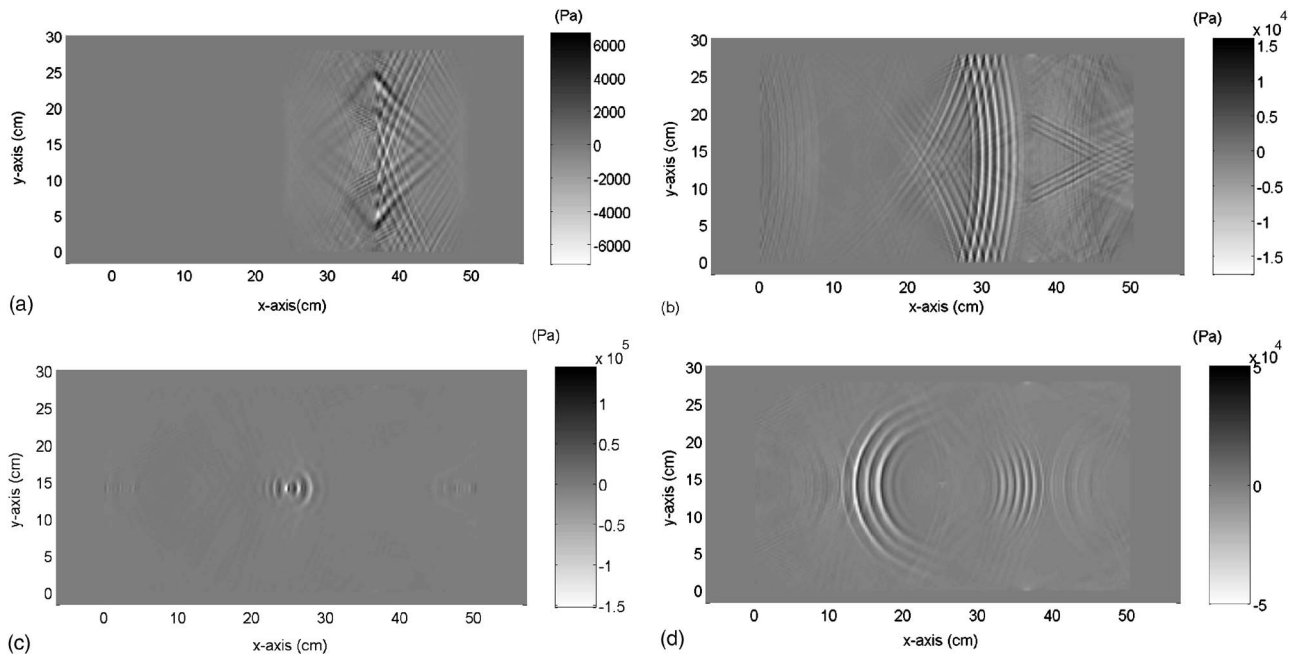


FIG. 9. Retrofocusing propagation in material. (a) $2 \mu\text{s}$: Emission of reversed signal with receivers. (b) $8 \mu\text{s}$: Reflection of reversed signal on edge. (c) $17 \mu\text{s}$: Retrofocusing on damaged zone. (d) $18.8 \mu\text{s}$: Diffraction of the retrofocusing signal from local damaged zone.

200 kHz, equal to f_c . Nevertheless, although the frequency of the summed signal reached a maximum pressure at about 400 kHz, i.e., around the second harmonic of the received signal, higher harmonics were also noticeable. The value of the PI method is in removing all odd order harmonics and hence linear components (here 200 kHz components).

Four different retrofocusing steps are shown in Fig. 9. At the beginning [Fig. 9(a)], receivers are excited in order to backpropagate the returned time signal. After reflection of this signal on the right of the sample edge [Fig. 9(b)], the

focused wave front is observable. After $17 \mu\text{s}$ backpropagation [Fig. 9(c)], the returned signal retrofocuses on the defect position (in this example, defect position 1). The signal after retrofocusing is presented in Fig. 9(d).

Harmonic filtering and PI were tested on the six different defect positions defined in Sec. III A. Figure 10 illustrates the maximum amplitude of stress (along z axis) in time for two different defect positions. Figures 10(a) and 10(c) represent the maximum amplitude plots in the case of a central defect (position 1) with PI and harmonic filtering meth-

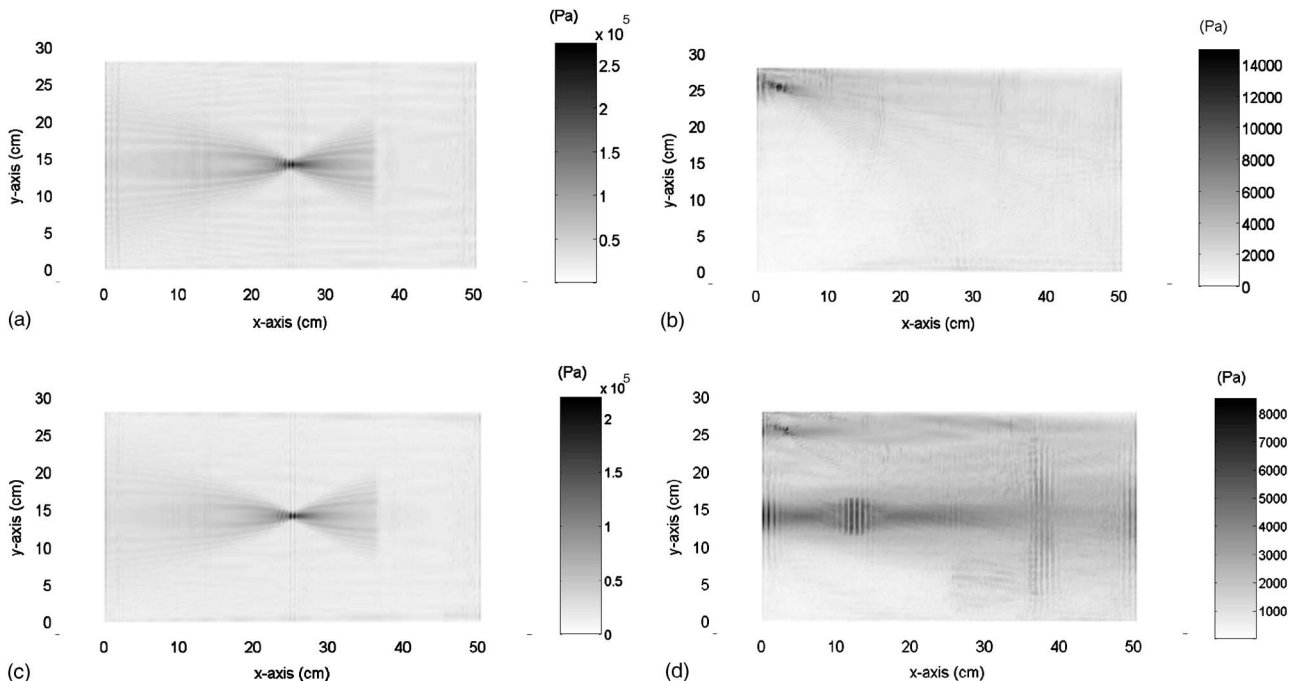


FIG. 10. Matrix of the maximum amplitude of stress along z axis for retrofocusing with the pulse inversion method (a), (b) and harmonic filtering method (c), (d) for two different defect positions.

TABLE I. Comparison between the two methods of filtering [harmonic and pulse inversion (PI)] for the six defect positions and three source sizes.

	$A_{PI} = \frac{\bar{m}(A_{\max}(\text{defect}))}{\bar{m}(A_{\max}(\text{bulk}))}$ (with PI)	$A_{Hfil} = \frac{\bar{m}(A_{\max}(\text{defect}))}{\bar{m}(A_{\max}(\text{bulk}))}$ (with harmonic filtering)	$\frac{A_{PI}}{A_{Hfil}}$
Position 1			
280 mm source	6.44	9.31	0.69
165 mm source	5.67	9.66	0.59
50 mm source	3.93	4.34	0.90
Position 2			
280 mm source	5.86	6.39	0.92
165 mm source	5.38	5.30	1.02
50 mm source	3.44	2.97	1.16
Position 3			
280 mm source	7.08	6.88	1.03
165 mm source	6.24	6.15	1.01
50 mm source	3.36	3.28	1.02
Position 4			
280 mm source	7.27	2.88	2.53
165 mm source	6.10	2.10	2.91
50 mm source	4.25	1.00	4.28
Position 5			
280 mm source	7.35	6.38	1.15
165 mm source	6.00	5.68	1.06
50 mm source	4.12	4.13	1.00
Position 6			
280 mm source	6.62	5.91	1.12
165 mm source	5.63	4.50	1.25
50 mm source	4.05	2.35	1.72

ods, respectively. Similarly, Figs.10(b) and 10(d) represent the maximum amplitude plots in the case of a defect near the edge (position 4) with these two filtering methods.

2. Comparison and discussion

The retrofocusing quality was different according to defect position and filtering method. All results for three different source sizes for the six defect positions are presented in Table I. They were obtained from the images of the maximum retrofocused signals. The first and second columns represent the ratio between the mean amplitude on the defect and the mean amplitude on the bulk with PI [$A_{PI} = \bar{m}(A_{\max}(\text{defect}))/\bar{m}(A_{\max}(\text{bulk}))$] and harmonic filtering [$A_{Hfil} = \bar{m}(A_{\max}(\text{defect}))/\bar{m}(A_{\max}(\text{bulk}))$], respectively.

The ratio of these two results was then calculated and is presented in column 3 of Table I. Careful study of the evolution of this ratio A_{PI}/A_{Hfil} , bearing in mind that if $A_{PI}/A_{Hfil} \leq 1$ then retrofocusing quality with harmonic filtering is better than with PI, leads to the following conclusions. If the defect is in the middle of the bulk (i.e., between the emitter and the receivers), the retrofocusing spot is smaller when applying the harmonic filtering method [Figs. 10(a) and 10(c)]. This can be explained by the dependence of the spot size on the frequency: The higher the frequency, the smaller the spot size (Fig. 4). Moreover, the larger the source, the smaller the focalization spot becomes. Using harmonic filtering, the third harmonic was returned, while with PI filtering it was mainly the second harmonic, which was

returned (Fig. 8). If the defect is near the edge of the sample, the retrofocusing is only on the defect when applying PI filtering [Figs. 10(b) and 10(d)]. This can have several explanations. First of all, the energy contained in the harmonics is smaller if the defect is in position 4 than when it is in the middle position. Moreover, when applying harmonic filtering, there is always a small part of the fundamental frequency which is not filtered out, whereas it is possible be sure that this part is not present with PI filtering. The ratio between the PI part and the filtered harmonic (third, for example) is higher in the case of an edge position.

IV. SUMMARY AND CONCLUSION

The purpose of this study was to provide a numerical model for validating the combination of nonlinear elastic wave spectroscopy with time reversal methodology (NEWS-TR method) as a defect imaging technique. The approach developed for this purpose was based on a 2D pseudospectral time domain algorithm. The elastic coefficient tensor was modified using the Kelvin notation, and hysteretic nonlinear behavior using the Preisach–Mayergoyz space model was introduced in the damaged area.

Simulations performed on a metallic sample showed the feasibility and value of the NEWS-TR methodology for microdamage imaging. Harmonic generation was investigated with two filtering methods: Harmonic filtering and pulse inversion filtering. Different parameters influenced retrofocusing quality. In these simulations, if the defect was in the middle of the bulk the retrofocusing spot was smaller when applying the harmonic filtering method in contrast, if the defect was near the edge of the sample the retrofocusing was only on the defect when applying pulse inversion filtering. Two-frequency wave mixing (intermodulation) simulations have been performed with the same numerical model and have given identical results.³⁵ Three-dimensional extension seems to be an efficient tool to apply the NEWS-TR method to real samples by combining simulations with experimental data, and this work is now in progress in our laboratory.

ACKNOWLEDGMENT

This study was supported by the EC Sixth Framework Program AERONEWS (*Health Monitoring of Aircraft by Nonlinear Elastic Wave Spectroscopy Specific Targeted Research: FP6-502927*).

APPENDIX

Based on the Kelvin notations, the calculation steps needed for the implementation of 2D elastic wave propagation in an initially transverse isotropic and nonlinear hysteretic medium (based on a PM space formalism) are as follows:

- (1). Calculation of the three-dimensional Kelvin stress vector:

$$\tilde{\tau} = \begin{pmatrix} \tau_{xx} \\ \tau_{yy} \\ \tau_{xy}\sqrt{2} \end{pmatrix};$$

- (2). Calculation of the stress projections along the three eigenstress directions:

$$\tilde{\tau}^{(1)} = \frac{X_1}{\sqrt{1+X_1^2}} \tau_{xx} + \frac{1}{\sqrt{1+X_1^2}} \tau_{yy},$$

$$\tilde{\tau}^{(2)} = \frac{X_2}{\sqrt{1+X_2^2}} \tau_{xx} + \frac{1}{\sqrt{1+X_2^2}} \tau_{yy}, \quad \tilde{\tau}^{(3)} = \tau_{xy} \sqrt{2}.$$

- (3). Modification of each associated eigenvalue $E^{(i)}$ using a scalar PM space model:

$$E^{(1)} = \frac{1}{\frac{1}{\alpha_1} + \gamma_1 \frac{\partial \tilde{\mathcal{E}}_H^{(1)}}{\partial \tilde{\tau}^{(1)}}}, \quad E^{(2)} = \frac{1}{\frac{1}{\alpha_2} + \gamma_2 \frac{\partial \tilde{\mathcal{E}}_H^{(2)}}{\partial \tilde{\tau}^{(2)}}},$$

$$E^{(3)} = \frac{1}{\frac{1}{2C_{44}} + \gamma_3 \frac{\partial \tilde{\mathcal{E}}_H^{(3)}}{\partial \tilde{\tau}^{(3)}}}.$$

- (4). Calculation of the modified elastic tensor using

$$C_{11} = E^{(1)} \frac{X_1^2}{1+X_1^2} + E^{(2)} \frac{X_2^2}{1+X_2^2},$$

$$C_{13} = E^{(1)} \frac{X_1}{1+X_1^2} + E^{(2)} \frac{X_2}{1+X_2^2},$$

$$C_{33} = E^{(1)} \frac{1}{1+X_1^2} + E^{(2)} \frac{1}{1+X_2^2},$$

$$C_{44} = \frac{E^{(3)}}{2}.$$

Here the following notations have been introduced:

$$X_1 = -\frac{1}{C_{13}} \left(\frac{C_{33}}{2} - \frac{C_{11}}{2} - \frac{1}{2} \sqrt{C_{33}^2 - 2C_{11}C_{33} + C_{11}^2 + 4C_{13}^2} \right),$$

$$X_2 = -\frac{1}{C_{13}} \left(\frac{C_{33}}{2} - \frac{C_{11}}{2} + \frac{1}{2} \sqrt{C_{33}^2 - 2C_{11}C_{33} + C_{11}^2 + 4C_{13}^2} \right),$$

$$\alpha_1 = \frac{C_{33}}{2} + \frac{C_{11}}{2} + \frac{1}{2} \sqrt{C_{33}^2 - 2C_{11}C_{33} + C_{11}^2 + 4C_{13}^2},$$

$$\alpha_2 = \frac{C_{33}}{2} + \frac{C_{11}}{2} - \frac{1}{2} \sqrt{C_{33}^2 - 2C_{11}C_{33} + C_{11}^2 + 4C_{13}^2}.$$

¹K. McCall and R. Guyer, "Equation of state and wave propagation in hysteretic nonlinear elastic materials," *J. Geophys. Res.* **99**, 23887–23897 (1994).

²L. A. Ostrovsky and P. A. Johnson, "Dynamic nonlinear elasticity in geomaterials," *Riv. Nuovo Cimento* **24**, 1–46 (2001).

³J. A. TenCate, E. Smith, and R. A. Guyer, "Universal slow dynamics in

granular solids," *Phys. Rev. Lett.* **85**, 1020–1023 (2000).

⁴R. A. Guyer, J. A. TenCate, and P. Johnson, "Hysteresis and the dynamic elasticity of consolidated granular materials," *Phys. Rev. Lett.* **82**, 3280–3283 (1999).

⁵G. D. Meegan Jr., P. A. Johnson, R. A. Guyer, and K. R. McCall, "Observations of nonlinear elastic wave behavior in sandstone," *J. Acoust. Soc. Am.* **94**, 3387–3391 (1993).

⁶K. Ishihara, *Soil Behaviour in Earthquake Geotechnics* (Clarendon, Oxford, 1996).

⁷K. E. Van Den Abeele, A. Sutin, J. Carmeliet, and P. Johnson, "Micro-damage diagnostics using nonlinear elastic wave spectroscopy (NEWS)," *NDT & E Int.* **34**, 239–248 (2001).

⁸V. Nazarov and A. Sutin, "Harmonic generation in the propagation of elastic waves in nonlinear solid media," *Sov. Phys. Acoust.* **35**, 410–413 (1989).

⁹G. Dace, R. Thompson, and O. Buck, "Measurement of the acoustic harmonic generation for materials characterization using contact transducers," *Rev. Prog. Quant. Nondestr. Eval.* **11**, 2069–2076 (1992).

¹⁰A. Moussatov, B. Castagnède, and V. Gusev, "Frequency up-conversion and frequency down-conversion of acoustic waves in damaged materials," *Phys. Lett. A* **301**, 281–290 (2002).

¹¹K. Van Den Abeele, P. A. Johnson, and A. Sutin, "Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage. I. Nonlinear wave modulation spectroscopy (NWMS)," *Res. Nondestruct. Eval.* **12**, 17–30 (2000).

¹²M. Vila, F. Vander Meulen, S. Dos Santos, L. Haumesser, and O. Bou Matar, "Contact phase modulation method for acoustic nonlinear parameter measurement in solid," *Ultrasonics* **42**, 1061–1065 (2004).

¹³R. T. Beyer, *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York, 1998), Chap. 2.

¹⁴K. Van Den Abeele, K. Van de Velde, and J. Carmeliet, "Inferring the degradation of pultruded composites from dynamic nonlinear resonance measurements," *Polym. Compos.* **22**, 555–567 (2001).

¹⁵K. Van Den Abeele, F. Schubert, V. Aleshin, F. Windels, and J. Carmeliet, "Resonant bar simulations in media with localized damage," *Ultrasonics* **42**, 1017–1024 (2004).

¹⁶A. Granato and K. Lücke, "Theory of mechanical damping due to dislocations," *J. Appl. Phys.* **27**, 583–593 (1956).

¹⁷V. E. Nazarov and A. B. Kolpakov, "Experimental investigations of nonlinear acoustic phenomena in polycrystalline zinc," *J. Acoust. Soc. Am.* **107**, 1915–1921 (2000).

¹⁸M. W. Barsoum, M. Radovic, T. Zhen, P. Finkel, and S. R. Kalidindi, "Dynamic elastic hysteretic solids and dislocations," *Phys. Rev. Lett.* **94**, 085501 (2005).

¹⁹R. A. Guyer and P. A. Johnson, "Nonlinear mesoscopic elasticity: Evidence for a new class of materials," *Phys. Today* **52**, 30–36 (1999).

²⁰O. Bou Matar, S. Dos Santos, J. Fortineau, T. Goursolle, L. Haumesser, and F. Vander Meulen, "Pseudo spectral simulations of elastic waves propagation in heterogeneous nonlinear hysteretic medium," in *Proceedings of the 17th International Symposium of Nonlinear Acoustics*, State College, PA, 2005, pp. 95–98.

²¹A. Gliozzi, M. Griffa, and M. Scalerandi, "Efficiency of time-reversed acoustics for nonlinear damage detection in solids," *J. Acoust. Soc. Am.* **120**, 2506–2517 (2006).

²²M. Fink, "Time reversal of ultrasonic fields. I, II and III," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **39**, 555–592 (1992).

²³M. Fink, "Time reversed acoustics," *Phys. Today* **50**, 34–40 (1997).

²⁴C. Prada, F. Wu, and M. Fink, "The iterative time reversal mirror: A solution to self-focusing in the pulse echo mode," *J. Acoust. Soc. Am.* **90**, 1119–1129 (1991).

²⁵H. C. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal, C. Ferla, and D. R. Jackson, "Iterative time reversal in the ocean," *J. Acoust. Soc. Am.* **101**, 3089–3089 (1997).

²⁶R. K. Snieder and J. A. Scales, "Time-reversed imaging as a diagnostic of wave and particle chaos," *Phys. Rev. E* **58**, 5668–5675 (1998).

²⁷M. Tanter, J. Thomas, and M. Fink, "Time reversal and the inverse filter," *J. Acoust. Soc. Am.* **108**, 223–234 (2000).

²⁸A. Sutin, J. A. TenCate, and P. A. Johnson, "Single-channel time reversal in elastic solids," *J. Acoust. Soc. Am.* **116**, 2779–2784 (2004).

²⁹M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J.-L. Thomas, and F. Wu, "Time-reversed acoustics," *Rep. Prog. Phys.* **63**, 1933–1995 (2000).

³⁰A. Sutin, J. A. TenCate, and P. A. Johnson, "Development of nonlinear time reversed acoustics (NLTRA) for applications to crack detection in

- solids,” in *Proceedings of the Fifth World Congress on Ultrasonics*, Paris, 2003, pp. 121–124.
- ³¹A. Brysev, L. Krutyansky, V. Preobrazhensky, and P. Pernod, “Nonlinear ultrasonic phase conjugate beams and their application in ultrasonic imaging,” *Acoust. Phys.* **50**, 623–640 (2004).
- ³²P.-Y. Le Bas, K. Van Den Abeele, S. Dos Santos, T. Goursolle, and O. Bou Matar, “Experimental analysis for nonlinear time reversal imaging of damaged materials,” in *Proceedings of the Ninth European Conference on Non-Destructive Testing*, Berlin, 2006.
- ³³T. J. Ulrich, P. A. Johnson, and A. Sutin, “Imaging nonlinear scatterers applying the time reversal mirror,” *J. Acoust. Soc. Am.* **119**, 1514–1518 (2006).
- ³⁴O. Bou Matar, S. Dos Santos, S. Callé, T. Goursolle, S. Vanaverbeke, and K. Van Den Abeele, “Simulations of nonlinear time reversal imaging of damaged materials,” in *Proceedings of the Ninth European Conference on Non-Destructive Testing*, Berlin, 2006.
- ³⁵T. Goursolle, S. Callé, O. Bou Matar, and S. Dos Santos, “Nonlinear time reversal acoustics for defect localization: Numerical study of retrofocusing properties,” *Proc.-IEEE Ultrason. Symp.*, Vancouver, 317–320 (2006).
- ³⁶G. Zumpano and M. Meo, “A new nonlinear elastic time reversal acoustic method for the identification and localisation of stress corrosion cracking in welded plate-like structures – A simulation study,” *Int. J. Solids Struct.* **44**, 3666–3684 (2007).
- ³⁷T. J. Ulrich, P. A. Johnson, and R. Guyer, “Interaction dynamics of elastic waves with a complex nonlinear scatterer through the use of a time reversal mirror,” *Phys. Rev. Lett.* **98**, 104301 (2007).
- ³⁸I. Mayergoyz, “Hysteresis models from the mathematical and control theory points of view,” *J. Appl. Phys.* **57**, 3803–3805 (1985).
- ³⁹R. A. Guyer, K. R. McCall, and G. N. Boitnott, “Hysteresis, discrete memory and nonlinear wave propagation in rock: A new paradigm,” *Phys. Rev. Lett.* **74**, 3491–3494 (1995).
- ⁴⁰M. Scalerandi, V. Agostini, P. P. Delsanto, K. Van Den Abeele, and P. Johnson, “Local interaction simulation approach to modelling nonclassical, nonlinear behavior in solids,” *J. Acoust. Soc. Am.* **113**, 3049–3059 (2003).
- ⁴¹V. Aleshin, V. Gusev, and V. Zaitsev, “Propagation of initially bi-harmonics sound wave in a 1D semi-infinite medium with hysteresis nonlinearity,” *Ultrasonics* **42**, 1053–1059 (2004).
- ⁴²K. Helbig and P. N. J. Rasolofosaon, “A theoretical paradigm for describing hysteresis and nonlinear elasticity in arbitrary anisotropic rocks,” in *Proceedings of the Ninth International Workshop on Seismic Anisotropy*, Society of Exploration Geophysicists, Tulsa, 2000.
- ⁴³J. Dellinger, D. Vasicek, and C. Sondergeld, “Kelvin notation for stabilizing elastic-constant inversion,” *Rev. Inst. Fr. Pet.* **53**, 709–719 (1998).
- ⁴⁴O. Bou Matar, V. Preobrazhensky, and P. Pernod, “Two-dimensional axisymmetric numerical simulation of supercritical phase conjugation of ultrasound in active solid media,” *J. Acoust. Soc. Am.* **118**, 2880–2890 (2005).
- ⁴⁵M. Ghrist, B. Fornberg, and T. A. Driscoll, “Staggered time integrators for wave equations,” *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **38**, 718–741 (2000).
- ⁴⁶B. Fornberg, *A Practical Guide to Pseudospectral Methods*, Cambridge Monographs Series on Applied and Computational Mathematics, Cambridge University Press, No. 1 (1998).
- ⁴⁷D. H. Simpson, C. T. Chin, and P. N. Burns, “Pulse inversion Doppler: A new method for detecting nonlinear echoes from microbubble contrast agents,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **46**, 372–382 (1999).

Accelerometer measurements of acoustic-to-seismic coupling above buried objects

Keith Attenborough^{a)} and Qin Qin

Department of Engineering, the University of Hull, Hull HU6 7RX, United Kingdom

Jonathan Jefferis

DSTL, Fort Halstead, Kent, TN14 7BP, United Kingdom

Gary Heald

DSTL, Winfrith, Dorset, DT2 8WX, United Kingdom

(Received 20 December 2006; revised 17 September 2007; accepted 17 September 2007)

The surface velocity of sand inside a large PVC container, induced by the sound pressure from either a large loudspeaker radiating into an inverted cone and pipe or a Bruel and Kjaer point source loudspeaker mounted with its axis vertical, has been measured using accelerometers. Results of white noise and stepped frequency excitation are presented. Without any buried object the mass loading of an accelerometer creates resonances in the spectral ratio of sand surface velocity to incident acoustic pressure, i.e., the acoustic-to-seismic (A/S) admittance spectra. The A/S responses above a buried compliant object are larger and distinctive. The linear A/S admittance spectra in the presence of a buried electronic components box have been studied as a function of burial depth and sand state. The nonlinear responses above the buried box have been studied as a function of depth, sand state, and amplitude. Predictions of a modified one-dimensional lumped parameter model have been found to be consistent with the observed nonlinear responses. Also the modified model has been used to explain features of the A/S responses observed when using an accelerometer without any buried object. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2799477]

PACS number(s): 43.25.Jh, 43.25.Dc, 43.20.Gp, 43.25.Zx [RMW]

Pages: 3230–3241

I. INTRODUCTION

Donskoy¹ has shown that a nonlinear detection method for buried landmines is able to give better discrimination from clutter than the linear method based on spatial anomalies in the acoustically induced soil surface velocity,² albeit at the cost of a weaker signal. It has been suggested³ that the nonlinear method could be used in conjunction with the linear method to reduce the probability of false alarms and that “performing experiments in sifted loess soil, a weathered soil, or another granular material that is made homogeneous in a limited size concrete rectangular vessel should prove useful in modeling the vibro-acoustic nonlinear interaction.” Most reported experiments on acoustic-to-seismic (A/S) landmine detection have used a laser Doppler velocimeter (LDV). As well as avoiding the loading and interface effects associated with accelerometers, use of a LDV has clear advantages in terms of rapidity of measurements and spatial scanning. On the other hand, to the authors’ knowledge, studies of the effects of accelerometer loading and the contact region between an accelerometer and a granular medium have not yet been published.

The results of laboratory experiments using small and large accelerometers on sand in a container are reported here. These data demonstrate the effects of accelerometer loading without any buried object as well as the distinctive linear and nonlinear effects resulting from the presence of a buried compliant object. Donskoy *et al.*^{1,4} have proposed a lumped

parameter “engineering” model for the (*two-tone*) nonlinear response of soil and of soil plus buried object. The model implies that it is important to use probing frequencies for obtaining nonlinear responses at or near the frequencies of the resonances in the (linear) dynamic response of the soil plus buried object system. This single degree of freedom model has been refined to allow for higher order modes of vibration of the compliant lid of a landmine.⁵ Although there are clear indications of multimodal behavior in the data reported here, it is found that the one-dimensional (1D) lumped parameter model can be used to fit the relative nonlinear response amplitudes observed above a buried compliant object and also to explain observed characteristics of the admittance response obtained using the larger accelerometer without any buried object.

The usual downward shift in the main soil/mine resonance frequency with burial depth has been observed to reverse as depth is increased. This has been attributed either to soil stiffening⁵ or to a transition between 1D and three-dimensional effects.⁶ In the data reported here the resonance frequency decreases at a decreasing rate with depth between 5 and 40 mm. However, it is found necessary to decrease the sand shear stiffness with increasing depth to fit the lumped parameter model to the data.

Section II describes the equipment and calibration. Sections III and IV describe results obtained for admittance spectra and stepped tone inputs, respectively. Section V discusses the application of the lumped parameter model to data obtained without and with a buried compliant object and its

^{a)}Electronic mail: k.attenborough@hull.ac.uk

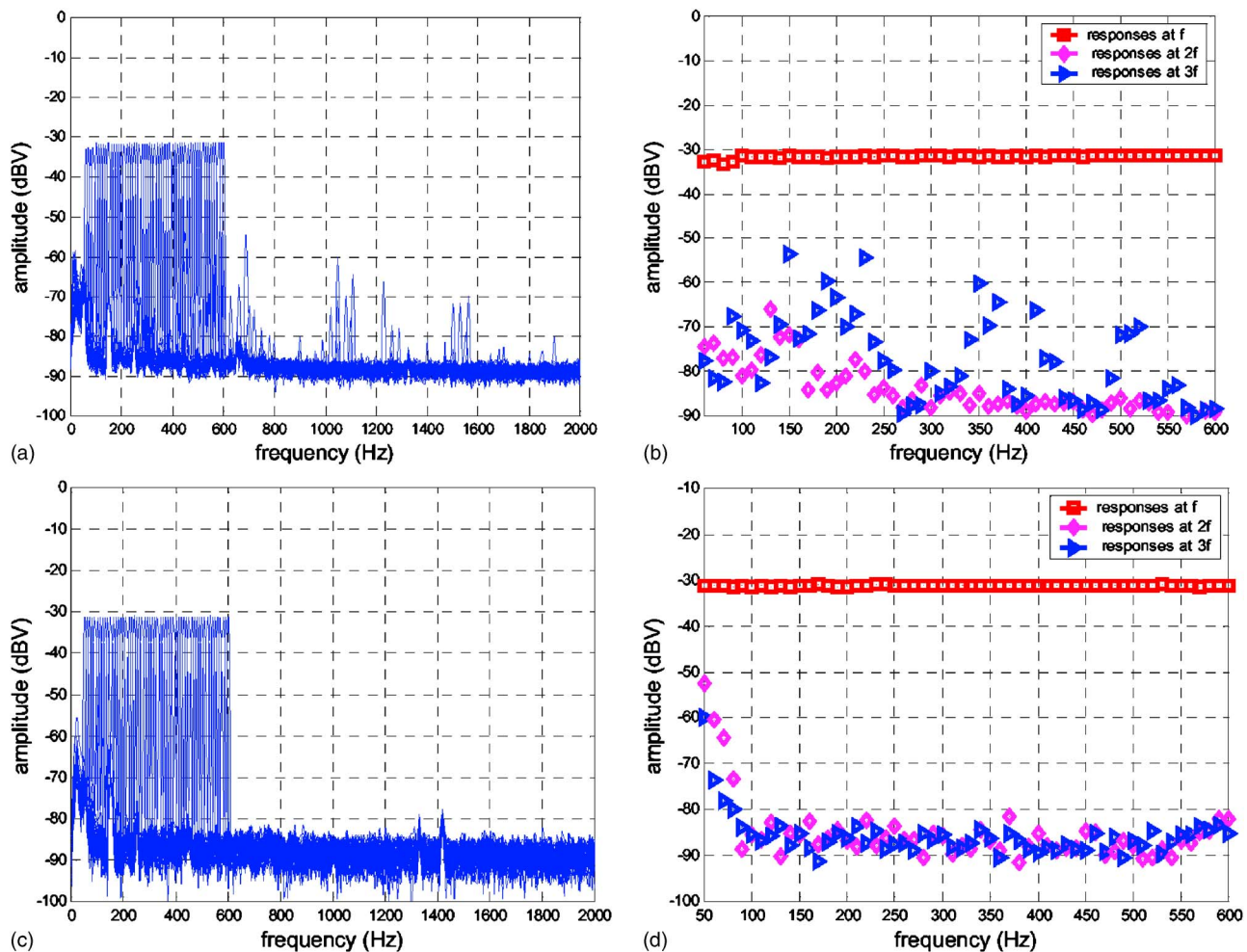


FIG. 1. (Color online) (a) Sound fields near sand surface during stepped tone excitation from the large loudspeaker and inverted cone arrangement, (b) corresponding spectra at primary frequency (f) and at second and third harmonics ($2f, 3f$) plotted at the associated primary frequencies. (c) Sound fields near sand surface during stepped tone excitation due to the B&K loudspeaker (d) corresponding spectra at primary frequency (f) and at harmonic frequencies ($2f$ and $3f$) plotted at the associated primary frequencies.

use in a modified form to reproduce effects observed when using an accelerometer without and with a buried object. Section VI contains concluding remarks.

II. MEASUREMENT SYSTEMS, CHARACTERISTICS, AND CALIBRATION

A. Sources and microphone

Two different loudspeaker sources (of white noise and stepped tones) have been used. The sound field from a Colossus 18B-600 loudspeaker was located above building sand in a thick-walled PVC tank in the shape of a truncated cone and above the center of buried objects by means of an inverted cone and pipe arrangement. The surface velocity of the sand, 85–110 cm above the base of the PVC container, was measured with an accelerometer located either at a horizontal distance of 70 mm from the outlet pipe axis or directly beneath the pipe outlet. Measurements have been made also using a Bruel and Kjaer (B&K) 4295 point source loudspeaker. When using this source, the accelerometer was

placed on the sand surface on the speaker axis directly below the speaker.

The acoustic signal close to the sand surface was detected using a Bruel and Kjaer 4134 $\frac{1}{2}$ in., condenser microphone attached to a B&K 2609 amplifier. Examples of the incident sound fields obtained when using stepped tones from the large loudspeaker and cone arrangement below the outlet and close to the sand surface are shown in Fig. 1(a). The measured levels at the input frequencies and the corresponding input frequency components at $2f$ and $3f$ are shown in Fig. 1(b) plotted at the primary input frequencies. Although there is a difference of at least 20 dB between the responses at the primary frequency and at the harmonics, to avoid spurious effects of harmonics in the exciting sound field, the measured surface velocity at the harmonic frequencies has been normalized using the sound field at these frequencies. Example stepped tone sound fields due to the B&K point source loudspeaker and the harmonic components are shown in Figs. 1(c) and 1(d). The harmonic outputs, particularly those at $3f$, are reduced greatly compared with those observed with the large loudspeaker and cone.

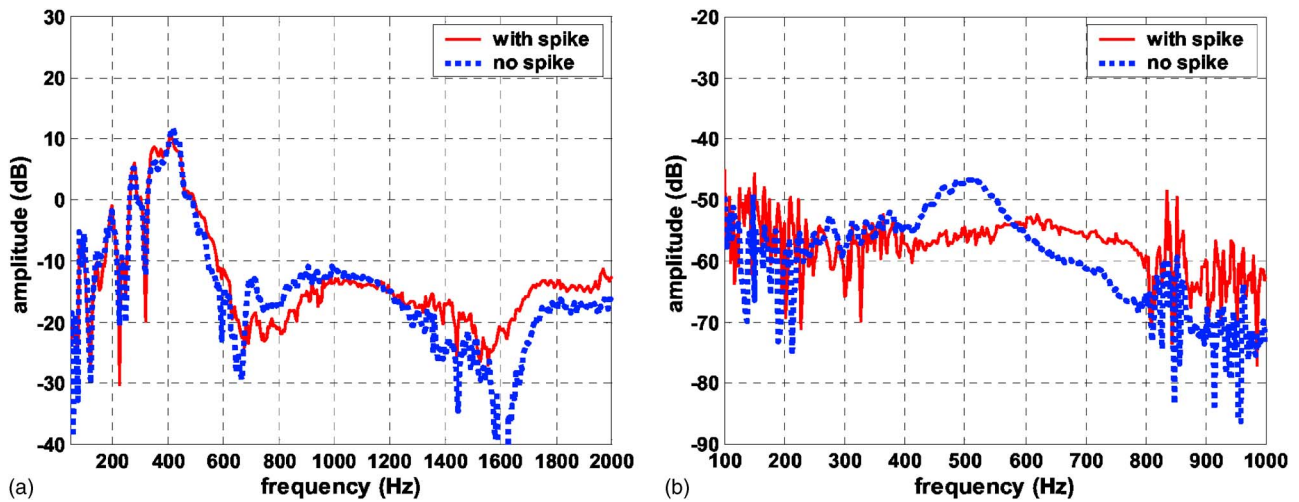


FIG. 2. (Color online) Admittance spectra measured on fairly dry (2 days) consolidated sand without a buried object (a) using the larger accelerometer and white noise at a surface SPL of 81 dB without and with a spike and (b) using a smaller accelerometer and white noise at 108 dB without and with an 18 mm (10 g) spike.

B. Accelerometers

Two accelerometers have been used for the measurements. The smaller accelerometer (DJB A/20, mass 18 g and sensitivity 35 pC/g) has a resonant frequency of 28 kHz. The larger accelerometer (B&K 7703A-1000, mass 120 g, and sensitivity 985.6 pC/g) has a resonant frequency of 8 kHz. The larger accelerometer is considerably more sensitive but results in greater mass loading effects. A B&K calibrator 4294 was used to calibrate the DJB A/20 accelerometer. Since the maximum load allowed on the B&K calibrator 4294 is 70 g, an indirect two stage process, involving a B&K 4809 shaker and the smaller accelerometer, was used to calibrate the system that contained the B&K 7703 (120 g) accelerometer. The resulting calibration has been used to obtain the absolute values of the admittance and surface velocity responses reported in Secs. III–V.

III. LINEAR A/S RESPONSES

A. A/S admittance spectra with no buried object

Use of the larger accelerometer resting on sand has been found to result in admittance spectra on moist or wet sand that have pronounced peaks between 200 and 400 Hz (see Fig. 2).

These peaks are not observed when using this accelerometer on a metal block but have been observed on all of the measurements made with the accelerometer on sand. Use of a spike on the larger accelerometer seems to affect the measured admittance spectra only at higher frequencies. On the other hand, a spike (mass 10 g, length 18 mm) makes a noticeable difference to the frequency content of A/S spectra measured on sand using the smaller accelerometer (see Fig. 2). Without a spike, there is a peak near 400 Hz followed by a significant fall off in the admittance spectrum above 600 Hz. The admittance peaks are affected more by the sand condition and the white noise amplitude than by the use of a spike. The moisture content of the sand affects the admittance spectra obtained at the same surface SPL with the larger accelerometer without a spike (see Fig. 3). As dis-

cussed in Sec. IV C, a change in moisture content is accompanied by a change in the sand shear strength.

Figure 4(a) shows admittance spectra obtained by using a small accelerometer with spike on different days at the surface of sand in the PVC “dustbin” without any buried object. These spectra are fairly similar. Any difference may be due to different sand moisture content. Although admittance spectra measured using the smaller accelerometer with spike [Figs. 2(b) and 4(b)] appear flatter than those obtained using the large accelerometer, they contain several peaks. As with the larger accelerometer it has been found that these give rise to nonlinear effects without any buried object.

As well as reducing the signal-to-noise ratio, use of a small accelerometer has been found to result in different peaks in the A/S admittance spectra from those obtained using the large accelerometer, i.e., there are spectral differences. Figure 4(b) shows the admittance spectra obtained with different accelerometers consecutively on wet sand without any buried object. Later sections show examples of

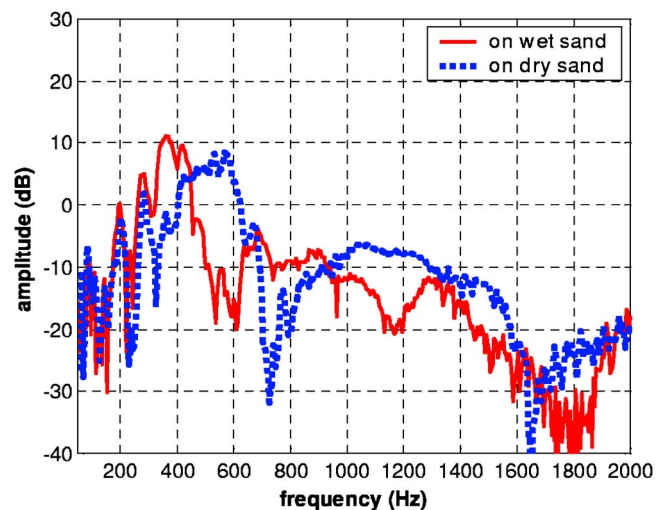


FIG. 3. (Color online) Admittance spectra measured using white noise at a surface SPL of 81 dB with the larger accelerometer (without spike) on moist consolidated sand and on dry consolidated sand without a buried object.

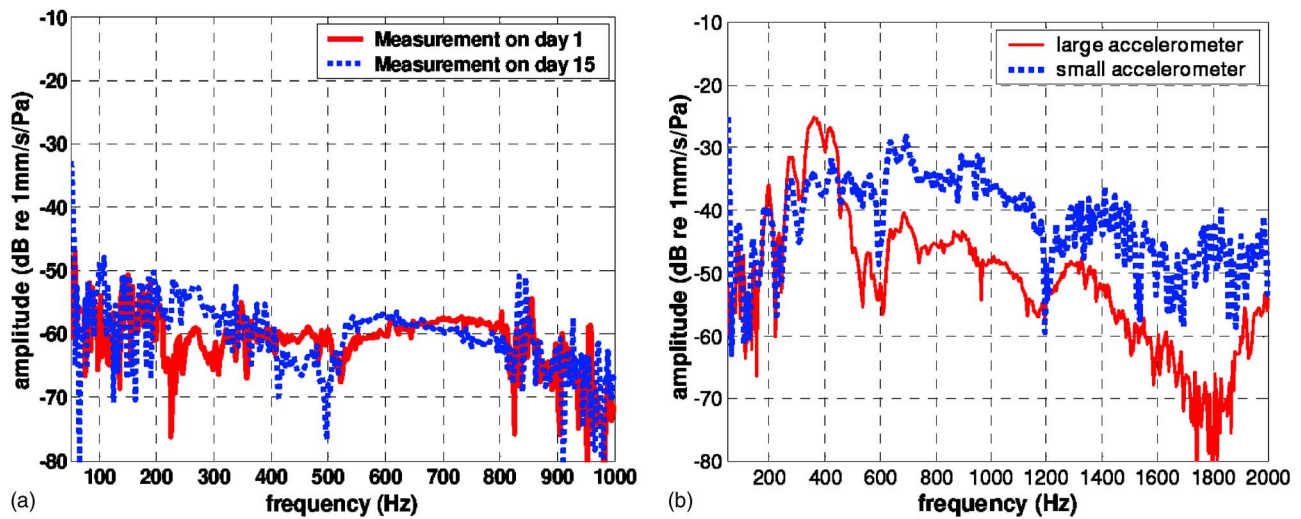


FIG. 4. (Color online) (a) Admittance spectra (calibrated) obtained 15 days apart on moist consolidated sand using the small accelerometer with spike and sound from the large loudspeaker with cone: (b) Admittance spectra obtained on wet sand subject to white noise at 81 dB SPL using the larger accelerometer without spike and the smaller accelerometer with a spike, respectively.

the differences obtained with buried objects. As well as the spectral differences, the smaller accelerometer results in noisier data.

B. A/S admittance spectra over buried RS box

The RS box is a die cast aluminum electronic components box with dimensions $187 \times 119 \times 52$ mm. Figure 5(a) shows the admittance spectrum obtained by using the large accelerometer directly on the top surface of the RS box when it was resting on sand. There are major admittance spectral peaks near 300 and 1200 Hz. Figure 5(a) shows also the admittance spectrum obtained with the RS box flush buried, i.e., with the top of the box exposed, and a thin (0.5-mm-thick) layer of sand beneath the accelerometer [Fig. 5(b)]. Similar admittance spectra are obtained with the box resting or sand or flush buried using either sand or double-sided tape beneath the accelerometer.

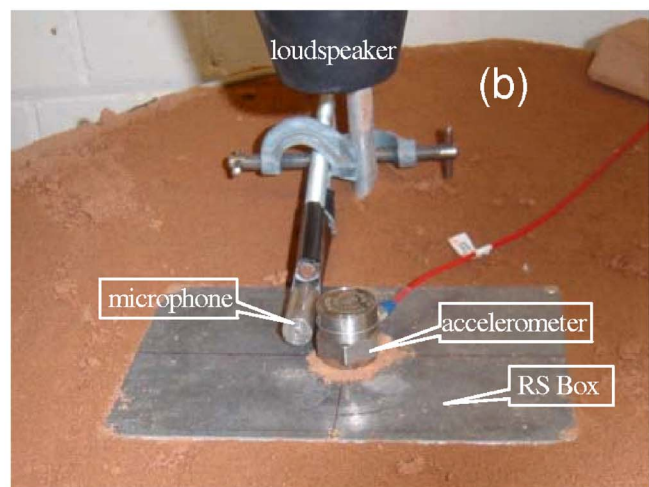
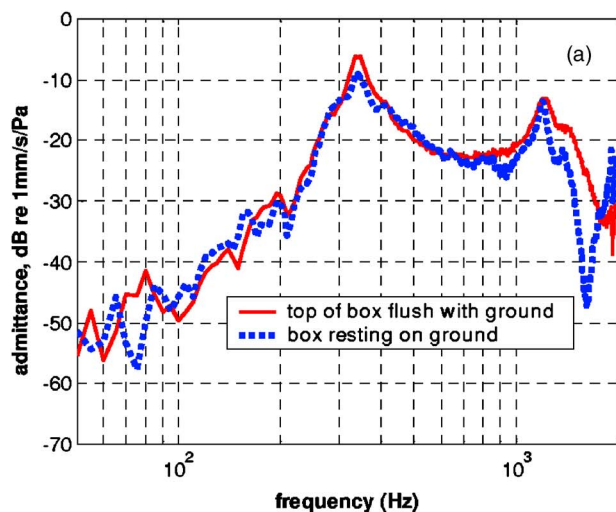


FIG. 5. (Color online) (a) Admittance spectra obtained on the top surface of the RS box when the box was resting on a sand surface and when the box was buried with the top surface exposed and a thin layer of sand beneath the accelerometer. (b) Photograph of measurement arrangement corresponding to the latter situation.

Figure 6 shows admittance spectra obtained with the large accelerometer placed on the surface of sand with the RS box buried at various depths in moist consolidated sand. The measured variation of the first (main) peak in the admittance spectra with increasing RS box burial depth is consistent with predictions of a 1D model,^{1,4} i.e., there are decreases in the frequency and magnitude of the main resonance which can be attributed to the depth-dependent additional mass of the soil column and the stiffness of the soil.

IV. LINEAR AND NONLINEAR STEPPED TONE RESULTS

A. Stepped tone data obtained without a buried object

Attempts to observe nonlinear acoustic responses using the speaker/cone system have been made difficult by the lack

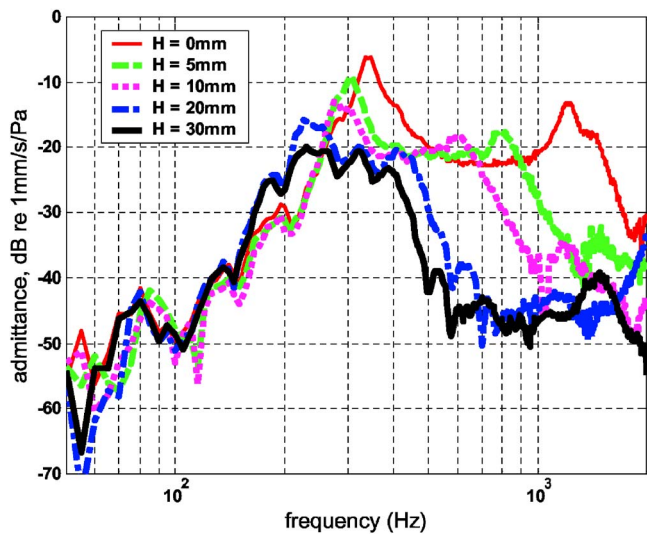


FIG. 6. (Color online) Admittance spectra obtained for various burial depths of the RS box in moist consolidated sand using white noise at a surface SPL of 87 dB.

of a “clean” source output. At the majority of frequencies, significant second and third harmonics are produced by this system. The nonlinearity in the acoustic loading induces corresponding seismic responses. If the A/S admittance obtained using stepped tones is equal to that obtained using white noise then it can be concluded that no significant nonlinearity is induced by the presence of the buried object. If the A/S response ratio obtained with stepped tones is greater than that obtained with white noise, then there must be some additional cause of nonlinearity in the system, potentially attributable to the buried object as long as the accelerometer signals used in the admittance calculations are above the noise floor.

A method of representing data from the stepped frequency experiments has been developed. The linear white noise admittance, the “fundamental admittance” at the stepped tone frequencies, the “second harmonic admittance,” derived by dividing the surface velocity response at the sec-

ond harmonic by the incident second harmonic sound level, and the correspondingly calculated “third harmonic admittance” are plotted on the same axes. If these traces track each other throughout the spectrum, then clearly there is no nonlinearity present and nonlinear responses due to the buried object have not been demonstrated. If either of the second or third harmonic admittance spectra show peaks above the linear (white noise) admittance, then something is causing nonlinearity, possibly the buried object. Although the B&K loudspeaker used for the experiments has a “cleaner” output than the speaker/cone system, the same method of analysis of the corresponding data has been adopted. Figures 7(a) and 7(b) show stepped tone A/S spectra obtained using the smaller and larger accelerometers, respectively, on dry consolidated sand without any buried object. The smaller accelerometer data were obtained for input frequencies up to 300 Hz whereas the larger accelerometer was used to obtain data for primary frequencies up to 600 Hz. Although the second and third harmonic surface velocities measured with the accelerometers are near the noise floor, these Figs. 7(a) and 7(b) indicate nonlinear responses associated with the linear admittance spectral peaks below 300 Hz. The $2f$ and $3f$ admittances obtained with the smaller accelerometer are not reliable, since the velocity responses at $2f$ and $3f$ are in the noise floor.

B. Stepped tone data as a function of burial depth in moist consolidated sand

Figure 8 shows linear and nonlinear surface velocity responses and admittances measured with the B&K loudspeaker and the larger accelerometer on the RS box flush buried in moist consolidated sand using a surface SPL of 87 dB. The data are presented in two ways. The left-hand figures show the surface velocity responses and noise floors recorded by the accelerometer for each tone. The right-hand figures show the admittance spectra calculated at f , $2f$, and $3f$ compared with that obtained using white noise. Figures 9(a)–9(d) show admittance spectra and noise floor spectra obtained with the RS box buried at depths from

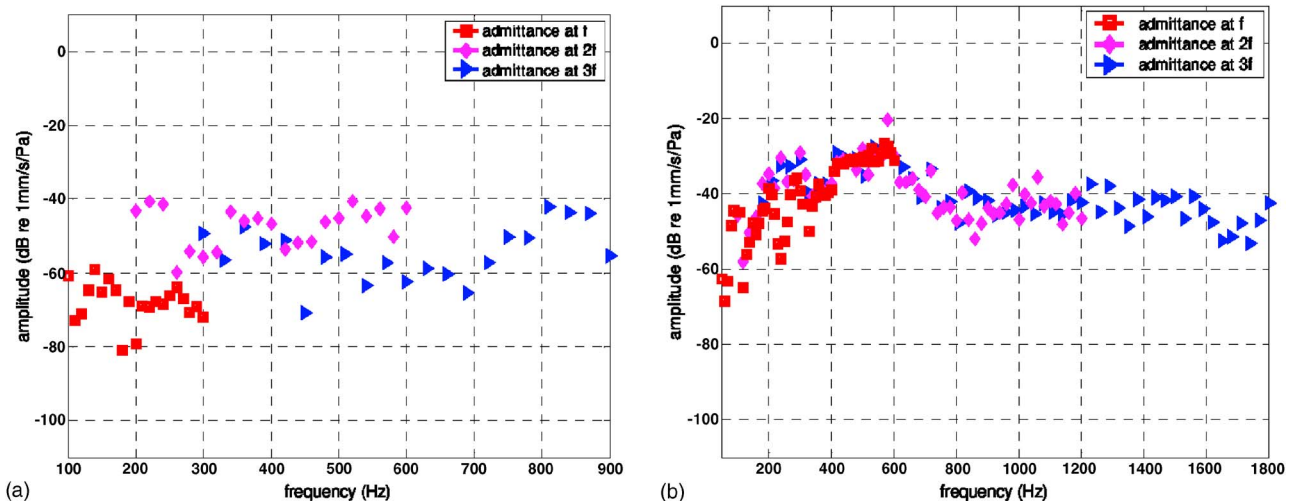


FIG. 7. (Color online) Ratios of the (dry consolidated) sand surface velocity to the corresponding stepped tone pressures at f , $2f$, and f , $3f$, respectively, without any buried object (a) measured using the smaller accelerometer with spike and (b) measured using the larger accelerometer with spike.

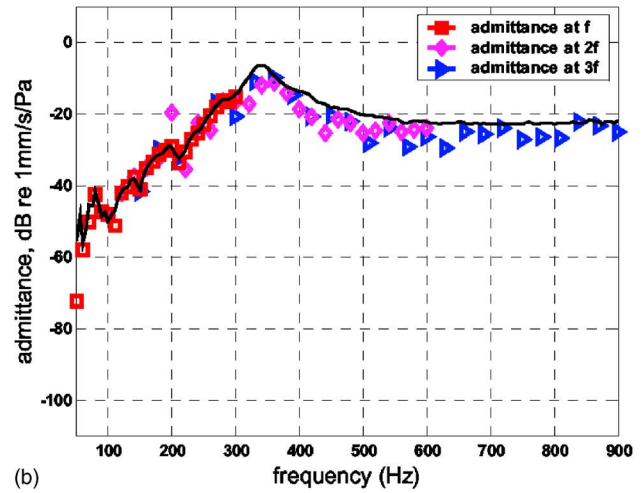
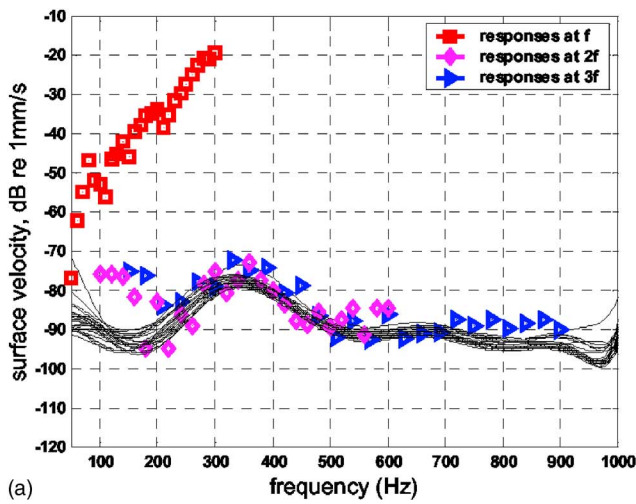


FIG. 8. (Color online) Measurements obtained by using large accelerometer with the RS box buried in moist consolidated sand with its surface exposed (burial depth 0 mm): (a) Velocity responses at f , f , $2f$, and f , $3f$, the continuous lines are the noise floors at each frequency (b) admittance spectra at f , $2f$, and $3f$, the continuous line is the admittance spectrum measured using white noise.

10 to 40 mm. The measurements for the 40 mm burial depth were made with a spike on the accelerometer.

Inspection of Figs. 8 and 9 indicates that, as predicted by Donskoy *et al.*,¹⁵ significant nonlinear responses associated

with the buried object are evident only when there is a sand layer above the buried compliant object: In this case if the RS box is buried at a depth of greater than 5 mm. Between depths of 10 and 30 mm the nonlinear responses diminish.

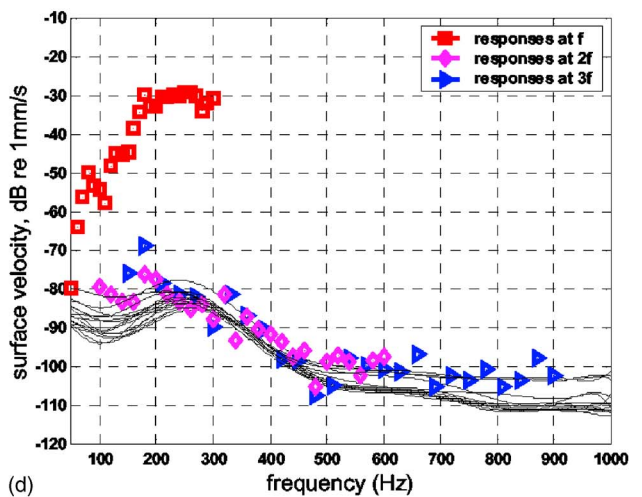
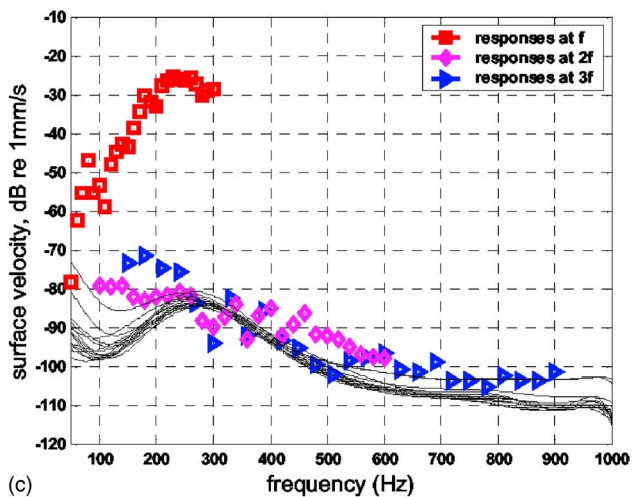
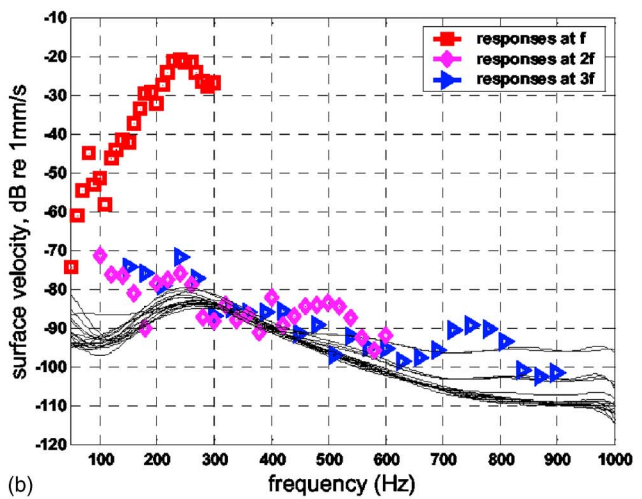
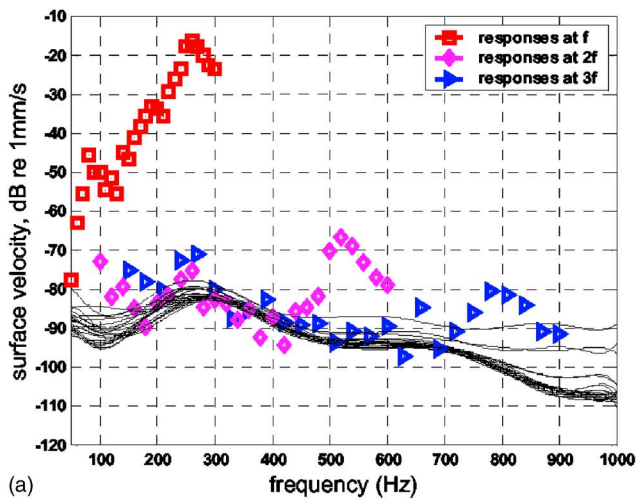


FIG. 9. (Color online) Surface velocity response spectra measured with the RS box buried in moist consolidated sand at depths of (a) 10 mm, (b) 20 mm, (c) 30 mm, and (d) 40 mm. The continuous lines represent noise floors at the various stepped frequencies.

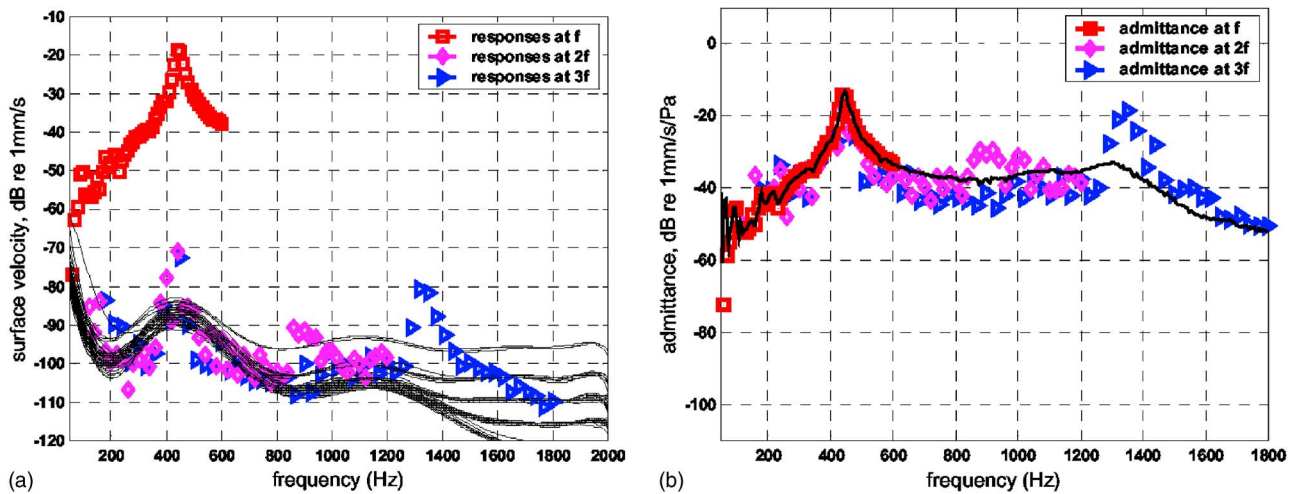


FIG. 10. (Color online) (a) Linear (squares) and nonlinear (diamonds and triangles) surface velocity responses to stepped tones and noise floors (continuous lines) obtained with the large accelerometer and the B&K loudspeaker source at a surface sound pressure level of 87 dB in dry consolidated sand containing the RS box buried at a depth of 10 mm and (b) corresponding admittance spectra.

The anomalously high nonlinear responses shown in Fig. 9(d) for 40 mm depth may be a consequence of use of a spike for those measurements. Experience with other buried objects (not reported here) showed that the nonlinear responses were affected by the condition of the backfill. Greatest nonlinear responses were obtained only if care was taken to use sand in a relatively “undisturbed” state above the buried object. This was achieved by setting some sand aside in a plastic bag to use for backfilling.

C. Variation in stepped tone admittance spectra with sound amplitude

Figure 10(a) shows spectra of the linear (white noise) and nonlinear surface velocity responses to stepped tones from the B & K loudspeaker. Figure 10(b) shows the admittance spectra calculated by dividing these surface velocity responses by the surface sound pressure levels at the corresponding linear and harmonic frequencies. The surface pressure level was 87 dB and the electronics components box was buried at 10 mm depth in dry consolidated sand, i.e., with the sand in a drier and stiffer state than for the measurements reported in Figs. 8 and 9. Figure 11 shows the corresponding surface velocity response spectra and accelerometer noise floors with surface pressure levels between 90 and 96 dB.

The differences in the sand shear strengths have been explored using a simple method [see Fig. 12(a)]. A block of sand in the state of interest resting on two adjacent metal blocks is subject to a gradually increasing shear force by raising the larger metal block on an extendable ramp. The force is measured by a balance placed beneath the larger metal block. The maximum force per unit area before fracture [see Fig. 12(b)] is taken as an indication of the shear strength. The moisture content has been obtained by weighing and drying. The moisture content (by weight) and shear strength indicator values for the two sand conditions are listed in Table I.

Comparison of Figs. 9(a) and 10(a) indicates that a decrease in sand moisture content and the corresponding in-

crease in shear strength produce a significant change in the nonlinear responses. In moist consolidated sand [Fig. 9(a)] the nonlinear responses are greater than those in dry consolidated sand [Fig. 10(a)]. Moreover the quadratic nonlinear response in moist sand exceeds the cubic response [Figure 9(a)]. Conversely, at all amplitudes in the dry consolidated sand, the cubic nonlinear response exceeds the quadratic nonlinear response [Figs. 10(a) and 11].

Figure 13 shows the peak surface responses in dry consolidated sand near 440 Hz as a function of amplitude.

As predicted by the suggested forms of nonlinearity, the nonlinear surface velocity responses at $2f$ and $3f$ increase with amplitude more rapidly than the linear (f) response. The 9.2 dB increase in the linear (f) response is associated with a 19 dB increase in the $2f$ response and a 20 dB increase in the $3f$ response. The rate of increase in the quadratic response is approximately twice that of the linear response, as it should be if it is quadratic. However after the expected initial increase of approximately 9 dB in the cubic nonlinear response, i.e., after the incident level is increased by 2.6 dB, the cubic response appears to “saturate.” Subsequently it increases at approximately the same rate as the quadratic response.

V. APPLICATION AND DEVELOPMENT OF A 1D ENGINEERING MODEL

A. Description of model

In the 1D engineering model,^{1,4} the admittance of the soil and mine system is determined from the lumped mechanical element arrangement shown in Fig. 14(a). The masses in the system are the soil column above the buried object and the buried object. The nonlinear response is assumed to be a consequence of the nonlinear behavior of the interface between the buried object and the surrounding soil and involves soil shear. The model results from analyzing the response of the equivalent electrical circuit [Fig. 14(b)] assuming a voltage drop across the points A and B in the form:

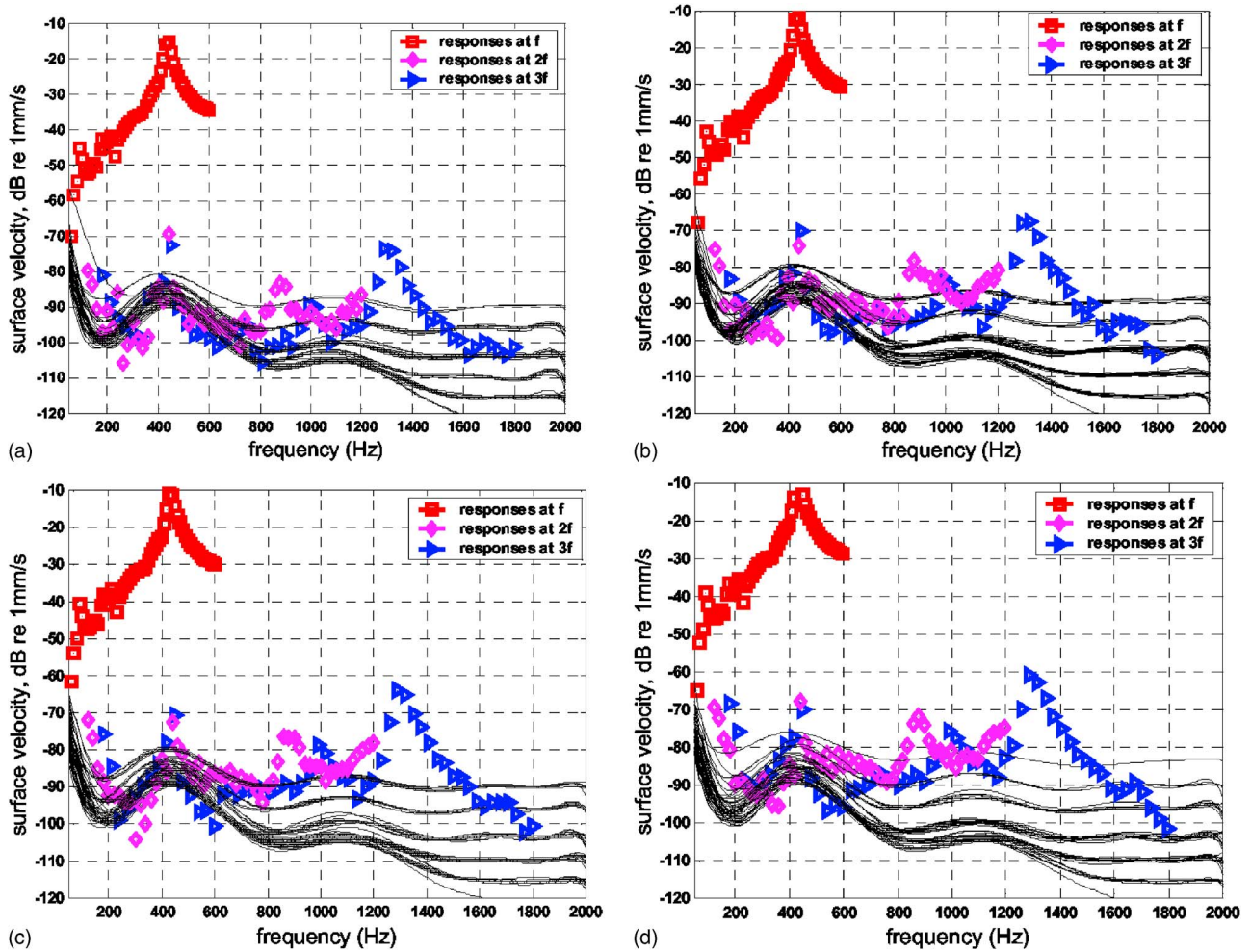


FIG. 11. (Color online) Linear and nonlinear surface velocity responses to stepped tone data and noise floors obtained with the large accelerometer and B&K loudspeaker source in *dry* consolidated sand at a surface sound pressure level of (a) 90 dB, (b) 92.6 dB, (c) 94.6 dB, and (d) 96.2 dB.

$$U(Q) = \frac{Q}{C_m} (1 + \alpha Q + \beta Q^2), \quad (1)$$

where $C_m = 1/k_m$, where k_m is the dynamic stiffness of the buried object, Q (charge) is the analogue of displacement, and α , β are the quadratic and cubic nonlinearity parameters, respectively. The response of the circuit to input (1) is obtained by calculating zeroth-, first-, and second-order perturbations. Details of the perturbation analysis are given elsewhere^{1,4} for two-tone excitation. For the single tone excitation considered here, expressions for quadratic and cubic nonlinear responses have been modified by assuming identical input frequencies.

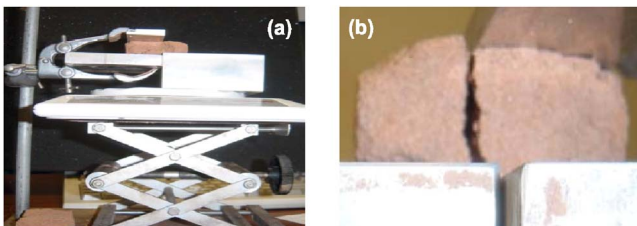


FIG. 12. (Color online) (a) Apparatus used for obtaining comparative indications of sand shear strength and (b) sand block after shearing fracture.

The quadratic nonlinear response to a sinusoidal input with angular frequency ω and amplitude U_0 is calculated from

$$I\Sigma(\omega) = \frac{2\alpha k_m}{\omega^2} I^2(\omega) \frac{z\Sigma^2(\omega)}{z_m^2(\omega)z_1(2\omega)}, \quad (2)$$

where

$$I(\omega) = \frac{U_0}{z_0(\omega)}, \quad z_0(\omega) = z_{s1}(\omega) + z\Sigma(\omega), \quad (3)$$

$$z\Sigma(\omega) = \frac{zS_2(\omega)z_m(\omega)}{zS_2(\omega) + z_m(\omega)}, \quad z_m(\omega) = r_m + i\left(m_m\omega - \frac{k_m}{\omega}\right), \quad (4)$$

TABLE I. Properties of two sand states.

Sand state	Moisture content (% by weight)	Shear strength indicator value (kPa)
Moist consolidated	8	3.6
Dry consolidated	0.9	12

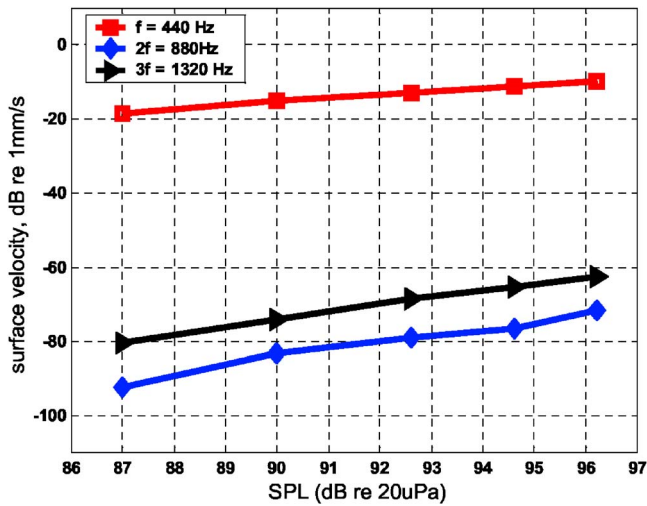


FIG. 13. (Color online) Variation of surface velocity response at the 440 Hz peak with incident sound level above the RS box buried at 10 mm depth in dry consolidated sand and the nonlinear responses at harmonics (2*f*, and 3*f*).

$$z_1(\omega) = z_{S1}(\omega) + z_m(\omega) \left[1 + \frac{z_{S1}(\omega)}{z_{S2}(\omega)} \right], \quad z_{S1}(\omega) = r_{S1} + i \left(m_s \omega - \frac{k_{S1}}{\omega} \right), \quad z_{S2}(\omega) = r_{S2} - i \left(\frac{k_{S2}}{\omega} \right). \quad (5)$$

α is the quadratic nonlinearity parameter, k_m , m_m , and r_m are, respectively, the stiffness, dynamic mass, and damping factor for the buried object, k_{s1} , r_{s1} , k_{s2} , r_{s2} , are, respectively, the shear stiffness, damping factor, normal stiffness, and damping factor for the sand, and m_s is the mass per unit area of the soil column.

The cubic response is calculated from

$$IM(\omega) = QM(\omega) + CM(\omega), \quad (6)$$

where

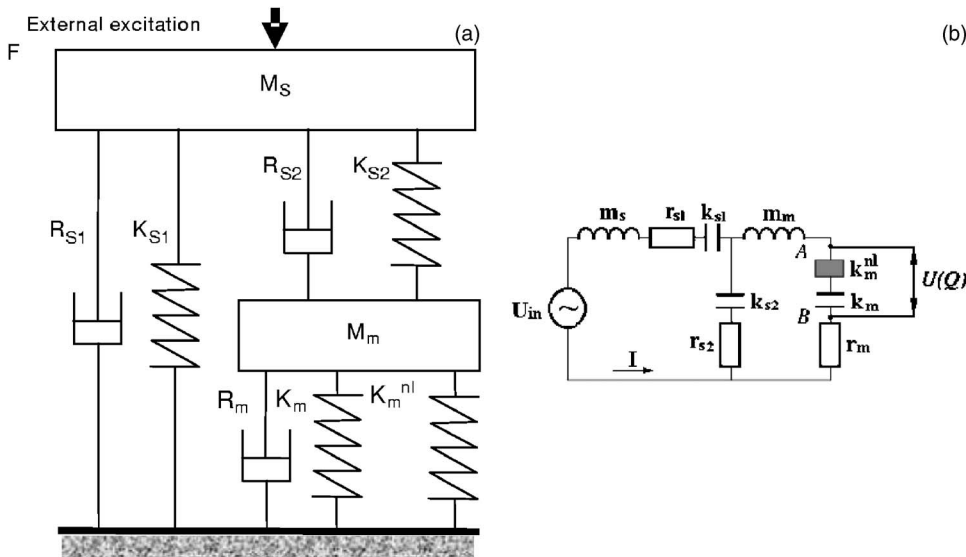


FIG. 14. (a) Lumped element model proposed by Donskoy *et al.* (Refs. 1 and 4) where M_m and M_s represent the masses of mine and soil column, respectively, R_{s1} and K_{s1} represent the dynamic shear damping and stiffness of the soil, R_{s2} and K_{s2} represent the dynamic compression damping and stiffness of the soil, and K_m and K_{mnl} represent the linear and nonlinear dynamic stiffness of the buried object/soil interface; (b) the equivalent circuit.

$$QM(\omega) = \frac{-\alpha^2 k_m^2 I^2(\omega) \overline{I(\omega)}}{\omega^4} \left(\frac{z_{S1}(\omega)}{z_m^2(\omega)} \right) \left(\frac{z_{S1}(\omega)}{z_m(\omega)} \frac{1}{z_1(\omega)} \right) \times \left(\frac{1}{z_2(2\omega)} \right), \quad z_2(\omega) = z_m + \frac{z_{s2}}{1 + (z_{s1}/z_{s2})}, \quad (7)$$

$$CM(\omega) = \frac{6i\beta^2 k_m I^2(\omega) \overline{I(\omega)}}{\omega^3} \left(\frac{z_{S1}(\omega)}{z_m(\omega)} \right)^2 \left(\frac{z_{S1}(\omega)}{z_m(\omega)} \right) \left(\frac{1}{z_1(\omega)} \right). \quad (8)$$

β is the quadratic nonlinearity parameter and the overbars denote complex conjugates.

The simplest way to include the effects of accelerometer loading in the absence of a buried object is to treat the accelerometer load as an object buried at zero depth. When a buried object is present, the accelerometer mass per unit area may be treated as additional to that exerted by the soil column above the object.

B. Best fit parameters for the box buried in consolidated sand

The model assumes that only the top of the buried object contributes to the dynamic responses. The mass of the electronic components box is 210 g. The thickness of the top of the box is 2 mm and the box area is $187 \times 119 \text{ mm}^2$ so assuming the density of aluminum to be 2700 kg/m^3 , the mass per unit area of top of box (m_m) is 5.4 kg/m^2 . Estimates of the stiffness and damping factor of the top of the box [$k_m = 2.38 \times 10^7 \text{ Pa/m}$, $r_m = 650 \text{ kg/(s m}^2)$] have been obtained by fitting the frequency and magnitude (335 Hz and $-8.9 \text{ dB re } 1 \text{ mm/s/Pa}$) of the first resonant peak shown in Fig. 6 corresponding to an input amplitude of 0.23 Pa (81 dB).

There are several peaks in the A/S admittance spectra measured with the larger accelerometer on consolidated sand without any buried object [see Figs. 3 and 4(b)]. The highest of these peaks is rather broad. A good fit to the stepped tone data is obtained by assuming a main admittance resonance peak of -23 dB at 460 Hz. The mass per unit area (m_{s1})

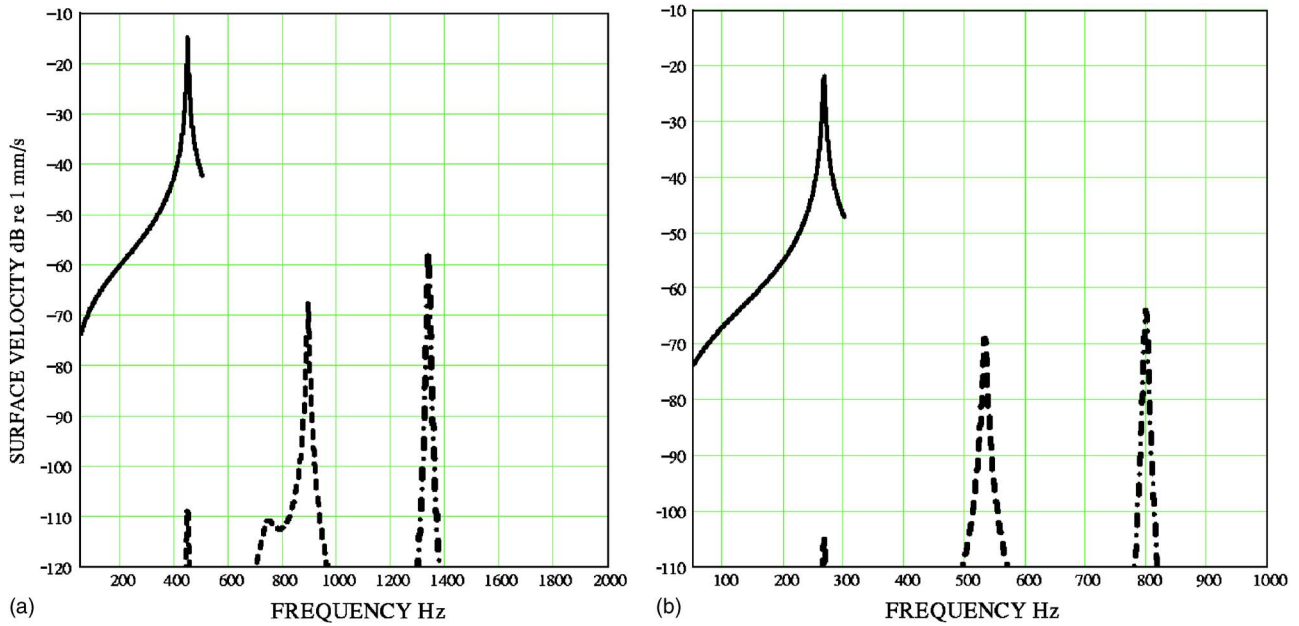


FIG. 15. (Color online) Predictions of the 1D engineering model, assuming that the accelerometer simply adds to the soil column mass loading, for the linear (continuous lines) and nonlinear (broken and dash-dot lines) A/S responses above the electronics components box buried at 10 mm depth (a) in dry consolidated sand, surface SPL 96.2 dB and (b) in moist consolidated sand, surface SPL at 87 dB.

exerted by the larger accelerometer is 236.8 kg/m^2 . Values for the dry consolidated sand shear stiffness and damping consistent with the assumed (single) peak frequency and magnitude are $k_{s1} = 2 \times 10^9 \text{ Pa/m}$, $r_{s1} = 6000 \text{ kg/(s m}^2)$. The admittance spectra measured without any buried object are discussed again in Sec. V C.

With the electronics components box buried so that its top surface is at 10 mm depth, the mass of sand column per unit area above RS box at 10 mm depth $m_{s2} = 15.9 \text{ kg/m}^2$. If the mass per unit area of the accelerometer (m_{s1}) is simply added to the mass per unit area of the soil column then there is a total mass per unit area (m_s) of 252.7 kg/m^2 . The remaining parameters (the normal stiffness and damping of the sand and the quadratic and cubic nonlinearity parameters of the sand/buried object interface) have been treated as free parameters for the purpose of obtaining the predictions in Fig. 15 corresponding to two sand states. For an input level of 1.29 Pa (96.2 dB), values of $k_{s2} = 0.9 \times 10^8 \text{ Pa/m}$, $r_{s2} = 1000 \text{ kg/(sm}^2)$, $\alpha = 0.8 \mu\text{m}^{-1}$, and $\beta = 0.1 \mu\text{m}^{-1}$ give the predictions in Fig. 15(a). The relative magnitudes of the linear, quadratic, and cubic stepped tone responses compare well with the data shown in Fig. 11(d). For moist consolidated sand values of $k_{s1} = 7 \times 10^8 \text{ Pa/m}$, $r_{s1} = 5000 \text{ kg/(sm}^2)$, $k_{s2} = 10^8 \text{ Pa/m}$, $r_{s2} = 1000 \text{ kg/(sm}^2)$, $\alpha = 1.5 \mu\text{m}^{-1}$, and $\beta = 0 \mu\text{m}^{-1}$ give the predictions in Fig. 15(b). That the best fit stiffness of the moist sand is approximately a third of that for the dry sand is consistent with the values in Table I. The frequencies and relative magnitudes of the linear, quadratic, and cubic stepped tone responses compare fairly well with the data for the linear and quadratic components shown in Fig. 9(a) but, even with $\beta = 0$, the cubic component is overpredicted.

C. A phenomenological model for the larger accelerometer on sand

In addition to direct mass loading by the accelerometer it is conjectured that, at the accelerometer/sand interface, there are nonlinear interactions involving shear. The 1D engineering model can be extended to include the accelerometer/sand interface as if it were a buried object/sand interface at zero depth. The accelerometer is assumed to have a dynamic mass ($m_a = km_{s1}$) which is a proportion (k) of the accelerometer mass per unit area and an associated dynamic stiffness (k_a) comparable to that of the RS box and damping (r_a). The impedance of the accelerometer/sand interface is given by

$$z_a(\omega) = r_a + i \left(m_a \omega - \frac{k_a}{\omega} \right). \quad (9)$$

The dynamic impedances due to the accelerometer and buried object interfaces are assumed to act in parallel (see Fig. 16) i.e.,

$$z_{ma}(\omega) = \frac{z_a(\omega)z_m(\omega)}{z_a(\omega) + z_m(\omega)}. \quad (10)$$

Expressions equivalent to Eqs. (5)–(8) are used after substituting z_a or z_{ma} for z_m as appropriate and adding the remainder of the accelerometer mass per unit area $[(1-k)m_{s1}]$ to m_s . Additional nonlinear parameters (α_a, β_a) are introduced to allow for nonlinearity at the accelerometer/sand interface. This nonlinearity is considered to be simultaneous with that due to the buried/object sand interface and in the same branch of the equivalent circuit (see Fig. 16) so that the perturbation results are given by Eqs. (1)–(8). The dynamic mass proportion (k), dynamic stiffness (k_a), damping (r_a), and nonlinear parameters of the accelerometer/sand interface

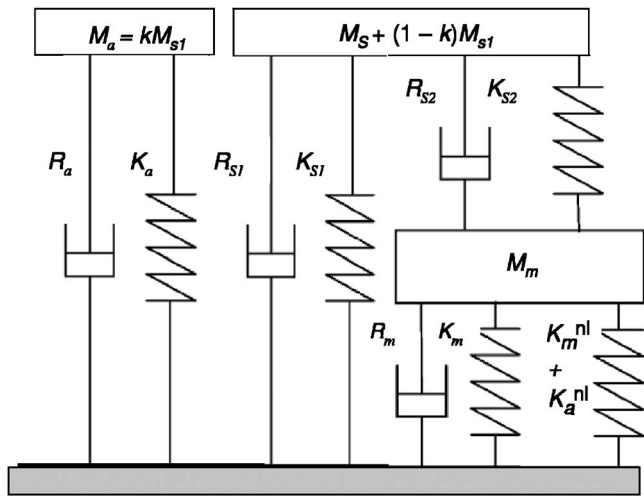


FIG. 16. Lumped parameter model modified to include the accelerometer/sand and buried object/sand interfaces in parallel. A proportion of the mass per unit area is treated as the dynamic mass involved at the accelerometer/sand interface (characterized by damping R_a , stiffness, K_a , and nonlinear parameters α_a and β_a) and the remaining accelerometer mass per unit area is included as an extra loading effect.

(α_a and β_a) are adjusted to give a tolerable visual fit of the data.

The parameters that reproduce the peaks near 200 and 300 Hz in the admittance spectra measured by the larger accelerometer on moist consolidated sand without any buried object (Fig. 3) are: $k_{s1} = 3 \times 10^8$ Pa/m, $r_{s1} = 3000$ kg/(sm²), mass loading of accelerometer = $0.6 \times m_{s1}$, $k_a = 1 \times 10^8$ Pa/m, $m_a = 0.4 \times m_{s1}$, $r_a = 6000$ kg/(sm²), $k_{s2} = 10^8$ Pa/m, and $r_{s2} = 1000$ kg/(sm²). However the highest peak near 400 Hz is not predicted. Assumed values of $\alpha_a = 0.5 \mu\text{m}^{-1}$, $\beta_a = 0.5 \mu\text{m}^{-1}$ give the predictions of the linear and nonlinear responses shown in Fig. 17(b), which may be compared with the data in Fig. 17(a). The two peaks near 200 and 300 Hz in

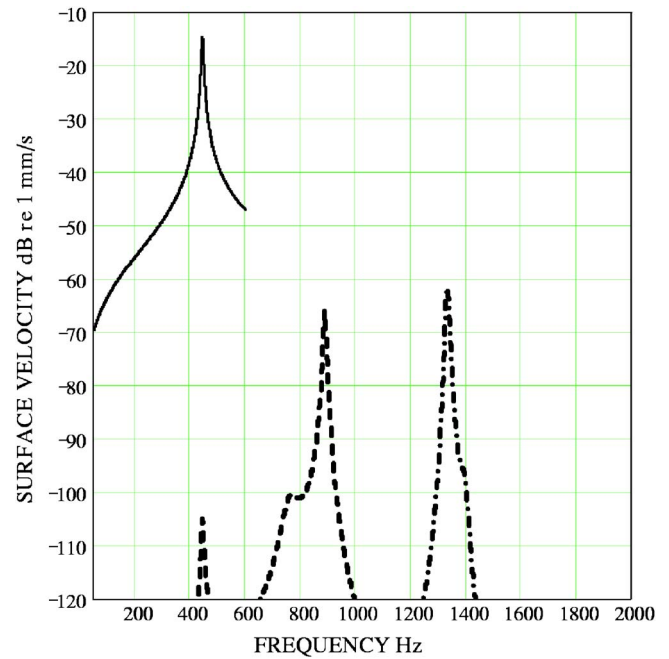
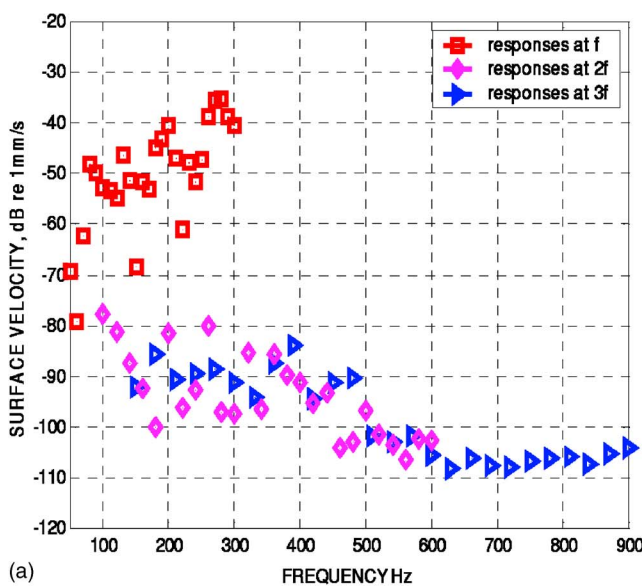


FIG. 18. (Color online) Predicted linear (continuous line) and nonlinear (broken and dot-dash lines) surface velocity responses above an RS box buried at 10 mm depth in dry consolidated sand assuming that the impedances of accelerometer/sand and buried object/sand interfaces act in parallel.

the linear stepped tone data and the relative magnitudes of the linear and nonlinear responses are reproduced fairly well.

Figure 18 shows predictions for dry sand containing an object buried at 10 mm depth. These have been obtained by assuming the accelerometer interface and nonlinearity parameters used for Fig. 17 but reducing the sand shear stiffness to $k_{s1} = 1.25 \times 10^9$ Pa/m rather than the value of 2×10^9 Pa/m used for the predictions in Fig. 15(a). These dry sand predictions can be compared with those of the simple model in Fig. 15(a) and with the data in Fig. 11(d).

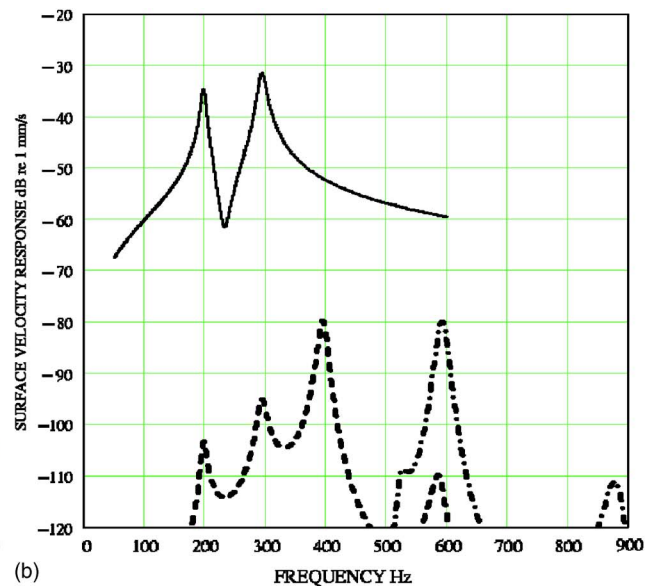


FIG. 17. (Color online) Measured stepped tone spectra of surface velocity on moist consolidated sand without any buried object obtained with the larger accelerometer and the B&K 4295 loudspeaker (b) corresponding predictions (linear—continuous line, quadratic—broken line, cubic—dash-dot line) including accelerometer mass loading and accelerometer/sand interface effects.

TABLE II. Assumed variation of parameter values with depth.

Depth/mm	k_{s1} (Pa/m)	r_{s1} (kg/(sm ²))	k_{s2} (Pa/m)	r_{s2} (kg/(sm ²))
0	6×10^8	0	1.5×10^8	3000
5	5.8×10^8	500	1.3×10^8	3000
10	5.6×10^8	1000	1.1×10^8	3000
20	5.2×10^8	1500	0.7×10^8	3000
30	4.8×10^8	4000	0.3×10^8	3000

D. Variation of admittance with burial depth

According to Donskoy *et al.*¹ and as expected intuitively, the shear stiffness increases and the normal (compressional) stiffness decreases with increasing burial depth. However, the data shown in Fig. 6 are best fitted by a linear decrease in the shear stiffness with depth when using the modified 1D engineering model (see Table II). Also, as listed in Table II, an assumed variation of the shear damping parameter values with depth is important to obtain tolerable visual agreements with measured magnitudes of the main peaks. A similar result is obtained with the simpler accelerometer model, so these results may reflect a limitation of the lumped parameter model in the presence of accelerometer loading. The depth-dependent values shown in Table II have been used to obtain the predictions shown in Fig. 19. While the measured change in frequency and magnitude of the main admittance peak are predicted fairly well, the Q 's of the main peaks are overpredicted.

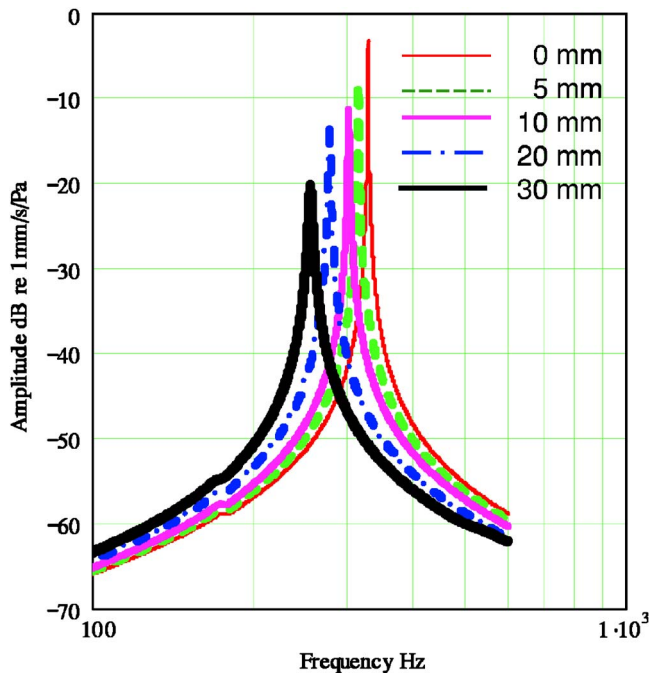


FIG. 19. (Color online) Admittance spectra as a function of burial depth predicted by using the parameter values listed in Table II.

VI. CONCLUSIONS

Accelerometers with different masses result in different A/S admittance spectra without or with a buried compliant object in sand, i.e., there are spectral as well as signal-to-noise differences. Peaks have been found in all accelerometer-measured A/S admittance spectra without a buried object. To some extent these peaks can be predicted by a 1D lumped parameter model modified, phenomenologically, to allow not only for loading from the accelerometer mass but also for an accelerometer/sand interface effect. Above a buried compliant object (an electronics components box) the A/S admittance is increased and the peaks in the admittance spectra depend on burial depth. This distinctive nonlinear behavior appears only after the box is buried to a depth of at least 5 mm. The modified lumped parameter model predicts the relative magnitudes of the linear and nonlinear responses in the dry sand state but overpredicts the cubic response in the moist sand state. The measured variation in admittance with depth is fitted by a decrease in the assumed shear stiffness with depth rather than the intuitively expected increase. In addition the shear damping has to be assumed to increase with depth to fit the measured decrease in the magnitude of the (first) admittance resonance. These results may indicate limitations of the lumped parameter model in the presence of accelerometer loading.

Nonlinear responses at $2f$ and $3f$ obtained with a buried electronic components box are substantially above the background noise level and larger than those observed with no buried object. The moisture content and corresponding shear strength of the sand are found to influence the strength of the nonlinear response and the relative strengths of the quadratic and cubic responses. Reasonable agreement between predictions and linear and nonlinear stepped frequency data with the RS box buried at 10 mm depth has been achieved using either a simple mass loading or a more complicated loading and interface model for the accelerometer. The quadratic nonlinear response is found to vary as expected with amplitude but the cubic nonlinear response saturates with amplitude.

ACKNOWLEDGMENT

Work supported by DSTL, UK.

¹D. Donskoy, A. Ekimov, N. Sedunov, and M. Tsionskiy, "Nonlinear seismo-acoustic landmine detection and discrimination," *J. Acoust. Soc. Am.* **111**, 2705–2714 (2002).

²N. Xiang and J. M. Sabatier, "An experimental study on antipersonnel landmine detection using acoustic-to-seismic coupling," *J. Acoust. Soc. Am.* **113**, 1333–1341 (2003).

³M. Korman and J. M. Sabatier, "Nonlinear acoustic techniques for landmine detection," *J. Acoust. Soc. Am.* **116**, 3354–3369 (2004).

⁴D. Donskoy, A. Reznik, A. Zagrai, and A. Ekimov, "Nonlinear vibrations of buried landmines," *J. Acoust. Soc. Am.* **117**, 690–700 (2005).

⁵A. Zagrai, D. Donskoy, and A. Ekimov, "Structural vibrations of buried land mines," *J. Acoust. Soc. Am.* **118**, 3619–3628 (2005).

⁶W. C. K. Alberts II, J. M. Sabatier, and R. Waxler, "Resonance frequency shift saturation in land mine burial simulation experiments," *J. Acoust. Soc. Am.* **120**, 1881–1886 (2006).

On the one-dimensional acoustic propagation in conical ducts with stationary mean flow

Ana Barjau^{a)}

Department of Mechanical Engineering, E.T.S.E.I.B., Polytechnic University of Catalunya (UPC), Diagonal 647, 08028 Barcelona, Spain

(Received 15 December 2006; revised 8 June 2007; accepted 18 September 2007)

This paper proposes a direct time-domain calculation of the time-domain responses of anechoic conical tubes with steady weak mean flow. The starting point is the approximated linear one-dimensional wave equation governing the velocity potential for the case of steady flow with low Mach number. A traveling solution with general space-dependent propagation velocity is then proposed from which the inward and outward pressure and velocity impulse responses can be obtained. The results include the well-known responses of conical and cylindrical ducts with zero mean flow. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2799478]

PACS number(s): 43.28.Py, 43.20.Mv [LLT]

Pages: 3242–3251

I. INTRODUCTION

The understanding of acoustic propagation in fluids with nonzero mean flow is of importance in many different problems, and therefore has been addressed by many authors from several points of view. The most general problem is that of three-dimensional (3D) nonlinear acoustic propagation in an inhomogeneous medium carrying a nonuniform unsteady mean flow. Some authors have studied it directly from the complete set of fluid-dynamic equations,¹ while others have derived first a general wave equation² and then proceeded to its application to different situations. In any case, the complexity of the problem does not allow for fully analytical solutions. Various numerical techniques have been applied to a variety of acoustic systems with mean flow (see, for example, Refs. 3 and 4).

There are a number of problems in which the situation can be idealized, thus leading to simpler equations with fully analytical solutions. Among them, the design of mufflers and silencers.

A usual idealization consists of assuming that the duct taper angle is small enough to avoid separation at the walls and that the mean flow Mach number is stationary and small enough so that M^2 can be neglected. Lung and Doige⁵ show that these hypotheses are valid up to a 20° taper for $M < 0.2$. Under these simplifications, the analytical frequency-domain solution of the corresponding 3D wave equation has been obtained for a conical duct by Willatzen.⁶

Further simplification can be obtained if just planar propagation is considered, which is equivalent to assuming that only the fundamental acoustic mode can exist, and if the acoustic variables are averaged over the cross section. As discussed by Campos,⁷ this hypothesis holds whenever the duct is not curved and has no internal obstacles; the ratio of the wave front area to the duct cross section is roughly constant; far from throats (where the acoustic problem becomes highly nonlinear).

All these assumptions are often accepted in the aforementioned design problems of mufflers and silencers.^{8,9} They highly simplify the problem while allowing the exploration of different important features such as the existence of cut-off frequencies, and the changes in amplitude and phase as the waves propagate.

The analytical solution of the resulting one-dimensional (1D) wave equation for the low frequency range has been obtained for a few regular duct shapes. The first authors to propose a frequency-domain analytical solution for a few duct shapes are Eisenberg and Kao.¹⁰ Their starting point is the set of basic equations for perfect fluids. Later on, Davies and Doak⁸ solve the 1D linear wave equation in the frequency domain. Easwaran and Munjal^{9,11} perform the same kind of calculation for exponential ducts and give the transfer matrix for hyperbolic and parabolic ducts, while Dokumaci¹² and Zhenlin¹³ determine if for general varying cross-section ducts.

Campos extensively treats this problem, also in the frequency domain. In Ref. 14 he shows that there is no elementary solution for any shape. For some shapes (parabolic and hyperbolic) it is possible to give a frequency domain solution in terms of Hankel functions.

Analytical time-domain calculations under the same simplifying hypothesis are rare. Shapiro¹⁵ treats the propagation of pressure pulses in open and closed cylindrical tubes. The total acoustic field is obtained by summing the multiple reflected and transmitted waves generated at the discontinuities. The final solution is then expressed as a series.

Compact analytical time-domain response functions have not been published either for general axisymmetric tubes or for the simple case of the conical duct, the main reason being the greater difficulty implied in time-domain calculations as compared to frequency-domain ones. Knowledge of such analytical response functions in cones with mean flow sets the path for the fully time-domain analytical calculation of the local reflection and transmission functions associated with diameter and taper discontinuities found in mufflers and silencers. Thus, the transient behavior of such systems could be more exactly analyzed than from

^{a)}Electronic mail: ana.barjau@upc.edu

frequency-domain descriptions. Moreover, the time domain is more suitable to describe mufflers when coupled to nonlinear systems. If the output of the nonlinear system (or the input of the linear muffler) is a complex flow $v(t)$, the muffler pressure response $p(t)$ can be obtained through a convolution by means of the corresponding pressure impulse response of the muffler, $p(t)=h(t)*v(t)$.

When calculating the local reflection and transmission functions associated with a particular discontinuity, both its left- and its right-hand side have to be considered as anechoic tubes (in order not to include the influence of other discontinuities). Whenever the left- or the right-hand side are divergent cones, this anechoic behavior is guaranteed if the cones are semi-infinite. However, convergent cones can only be physically extended to the apex, and any acoustic signal reaching it would necessarily be reflected. The convergent anechoic termination can either be treated analytically¹⁶ or implemented through an active system able to cancel any impinging wave.¹⁷

This paper proposes a direct time-domain calculation of the impulse response of anechoic conical tubes under the assumption of linear behavior, planar propagation, low Mach number, and no flow separation. The conical element is often encountered in muffler design (see, for instance, Ref. 18). More complex geometries can always be discretized as a series of conical spans with various tapers, which is always a better approximation than that obtained by means of cylindrical ones with various diameters.

Authors working on practical applications of acoustical propagation with nonzero mean flow could claim that the flow behavior at diameter or taper discontinuities is often complex, and thus the previous assumptions are uninteresting. Gerges *et al.*¹⁹ raise the question of the validity of such simplified treatments when dealing with reactive mufflers, and they show that for the low frequency range there is a good agreement between the predictions obtained with such 1D models and the experimental measurements.

Section II briefly presents the approximated linear 1D wave equation governing the velocity potential for the case of steady flow with a low Mach number. Henceforth, traveling solutions characterized by a general space-dependent propagation velocity are derived for it. Further developments lead to a system of differential equations whose compatibility is only guaranteed for cylindrical and conical tubes. Section III solves the differential system and gives the analytical expressions for the inward and outward propagation velocities. Then the formal expression of the acoustic pressure and velocity as a function of the velocity potential can be obtained. Sections IV and V deal with the analytical calculation of the time-domain responses of conical ducts to velocity and pressure impulses. The results include the well-known responses of conical and cylindrical ducts with zero mean flow.²⁰ Finally Sec. VI introduces future development.

II. GENERAL TIME-DOMAIN FORMULATION

The continuity and momentum equations governing the loss-less plane-wave propagation in ducts with variable cross-sectional area $A(x)$ (Fig. 1) are¹⁵

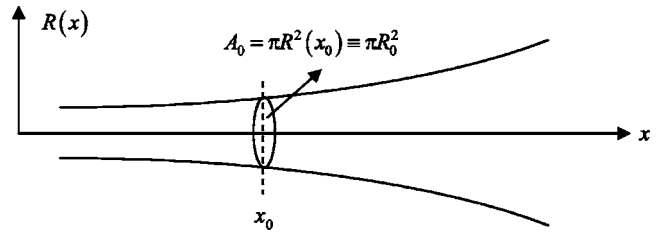


FIG. 1. Axysymmetric tube.

$$\frac{\partial(\rho_T A v_T)}{\partial x} = - \frac{\partial(\rho_T A)}{\partial t}, \quad (1)$$

$$- \frac{\partial(A p_T)}{\partial x} = \frac{\partial(\rho_T A v_T)}{\partial t} + \frac{\partial(\rho_T A v_T^2)}{\partial x}, \quad (2)$$

where (ρ_T, v_T, p_T) are the density, velocity, and pressure variables. For the case of a steady flow, they can be decomposed into a steady mean component and an acoustic one: $(\rho_T, v_T, p_T) = (\bar{\rho} + \rho, \bar{v} + v, \bar{p} + p)$. In general, the mean values depend on the longitudinal coordinate x : $\bar{\rho}(x), \bar{v}(x), \bar{p}(x)$. Thus, the Mach number $M(x) = \bar{v}(x)/c_0$, where c_0 is the sound propagation speed in the absence of mean flow, is also x dependent.

Assuming small acoustic perturbations and low Mach numbers (so that terms in M^2 can be neglected), those equations can be recombined into a single linear one containing only one acoustic variable. If the velocity potential $\phi(x, t)$ is chosen as fundamental variable, the approximate equation is⁸

$$\frac{1}{c_0^2} \frac{\partial^2 \phi}{\partial t^2} + \frac{2M_0 A_0}{c_0} \frac{1}{A} \frac{\partial^2 \phi}{\partial t \partial x} - \frac{\partial^2 \phi}{\partial x^2} - \frac{1}{A} \frac{dA}{dx} \frac{\partial \phi}{\partial x} = 0, \quad (3)$$

where (M_0, A_0) are the Mach number and the cross-sectional area of the tube at a given reference position x_0 : $(M_0, A_0) = (M, A)_{x=x_0}$.

The acoustic pressure and the acoustic velocity are related to the velocity potential through

$$p(x, t) = \bar{\rho} \frac{d\phi}{dt} = \bar{\rho} \left(\frac{\partial \phi}{\partial t} + \bar{v} \frac{\partial \phi}{\partial x} \right), \quad v(x, t) = - \frac{\partial \phi}{\partial x}. \quad (4)$$

The derivation of Eq. (3) that has been presented assumes from the very beginning a crosswise uniform acoustic field. Thus it is difficult to establish the degree of approximation contained in the final result. Rienstra²¹ proposes an interesting and more rigorous derivation starting from a 3D formulation and applying a perturbation method. As Eq. (3) is just the starting point of this paper, Rienstra's work is not reproduced here.

Equation (3) is a generalization to the Webster equation, as it considers nonzero mean flow. We now seek traveling solutions. As we know from the case of zero mean flow, it is advisable to use a new variable $\theta(x, t)$ which relates to $\phi(x, t)$ through the cross-sectional area: $\phi(x, t) = \theta(x, t) / \sqrt{A(x)}$. Equation (3) becomes

$$\frac{1}{c_0^2} \frac{\partial^2 \theta}{\partial t^2} + \frac{2M_0 A_0}{c_0} \frac{1}{A} \frac{\partial^2 \theta}{\partial t \partial x} - \frac{M_0 A_0}{c_0} \frac{1}{A^2} \frac{dA}{dx} \frac{\partial \theta}{\partial t} - \frac{\partial^2 \theta}{\partial x^2} + \frac{1}{2A} \left[\frac{d^2 A}{dx^2} - \frac{1}{2A} \left(\frac{dA}{dx} \right)^2 \right] \theta = 0. \quad (5)$$

The main difference between the case of zero mean flow and that of nonzero mean flow is that the sound propagation speed may depend on the spatial coordinate: $c(x)$. Thus, the traveling solutions will have the general form:

$$\theta(x, t) = \theta_+(\tau_+) + \theta_-(\tau_-),$$

with

$$(\tau_+, \tau_-) = \left(t + \int \frac{dx}{c_+(x)}, t - \int \frac{dx}{c_-(x)} \right) \triangleq (t + f(x), t - g(x)), \quad (6)$$

and

$$\frac{df}{dx} = f' = \frac{1}{c_+(x)}, \quad \frac{dg}{dx} = g' = \frac{1}{c_-(x)}.$$

The positive and negative indexes indicate the inward direction (pointing to decreasing values of x) and the outward one (pointing to the increasing x values), respectively. Of course, traveling solutions with arbitrary shape cannot exist whenever the duct presents cut-off frequencies. Thus, such general formulation will only lead to acoustically “transparent” ducts.

Whenever the cross section is constant (and so the mean flow \bar{v} is x independent), the values of $c_{\pm}(x)$ can be easily obtained by analyzing the propagation phenomenon from a reference frame moving with the mean flow (where the simple 1D wave equation would apply) and then transforming the solution back to the fixed frame (the duct walls).²² This leads to the well-known result: $c_{\pm} = c(1 \mp M_0)$.

However, the variable cross section (and so the x -dependence of \bar{v} and M) does not allow that simple rationale, as the moving frame would be a different one at each x value. Assuming (as usual) that the fixed frame is an inertial one, the moving frame would not be inertial, and the Webster equation for zero mean flow would not apply!

The investigation of the propagation speeds $c_{\pm}(x)$ for general axisymmetric geometries will be done directly in the time domain. There is no guarantee that traveling solutions are valid for any geometry and any oscillating regime. Our general calculation will allow us to establish in which cases they do exist.

The starting point of the calculation is the general formulation of the existence of traveling solutions. Whenever they exist, the wave equation can be written as

$$\frac{\partial^2 \theta}{\partial \tau_+ \partial \tau_-} = 0. \quad (7)$$

Taking into account that

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial \tau_+} + \frac{\partial}{\partial \tau_-}, \quad \frac{\partial}{\partial x} = f' \frac{\partial}{\partial \tau_+} - g' \frac{\partial}{\partial \tau_-}, \quad (8)$$

and performing some tedious calculations, Eq. (7) can be written as

$$\begin{aligned} \frac{\partial^2 \theta}{\partial \tau_+ \partial \tau_-} &= \frac{f' g'}{(f' + g')^2} \frac{\partial^2 \theta}{\partial t^2} + \frac{f' - g'}{(f' + g')^2} \frac{\partial^2 \theta}{\partial t \partial x} \\ &+ \frac{1}{f' + g'} \frac{d}{dx} \left(\frac{f'}{f' + g'} \right) \frac{\partial \theta}{\partial t} - \frac{1}{(f' + g')^2} \frac{\partial^2 \theta}{\partial x^2} \\ &- \frac{1}{f' + g'} \frac{d}{dx} \left(\frac{1}{f' + g'} \right) \frac{\partial \theta}{\partial x} = 0. \end{aligned} \quad (9)$$

Comparison between Eqs. (5) and (9) leads to

$$\frac{f' g'}{(f' + g')^2} = \frac{1}{c_0^2} H(x), \quad (10a)$$

$$\frac{f' - g'}{(f' + g')^2} = \frac{2M}{c_0} H(x), \quad (10b)$$

$$\frac{1}{f' + g'} \frac{d}{dx} \left(\frac{f'}{f' + g'} \right) = -\frac{1}{c_0} \frac{M}{A} \frac{dA}{dx} H(x), \quad (10c)$$

$$-\frac{1}{f' + g'} \frac{d}{dx} \left(\frac{1}{f' + g'} \right) = 0, \quad (10d)$$

$$\frac{1}{(f' + g')^2} = H(x), \quad (10e)$$

$$\frac{d^2 A}{dx^2} - \frac{1}{2A} \left(\frac{dA}{dx} \right)^2 = 0, \quad (10f)$$

where $H(x)$ is an arbitrary function. The unknowns in this differential system are only four: (H, f, g, A) . This means that its compatibility will have to be verified.

Equation (10f) defines the tube profiles for which the traveling solutions $[\theta_+(\tau_+), \theta_-(\tau_-)]$ are valid. As in general the cross-sectional area can be written as $A(x) = \pi R^2(x)$:

$$\begin{aligned} \frac{d^2 A}{dx^2} - \frac{1}{2A} \left(\frac{dA}{dx} \right)^2 = 0 &\Rightarrow R(x) = R_0 + m(x - x_0), \text{ with } R_0 \\ &= R(x_0). \end{aligned}$$

That is, only cylindrical ($m=0$) and conical profiles are transparent acoustic ducts and accept general traveling solutions, and thus will be the only ones to be considered in what follows. For the conical tube, $m = \tan \alpha$, where α is the cone aperture. The x coordinate will be taken always positive for divergent cones and negative for convergent ones [Figs. 2(a) and 2(b), respectively].

Equations (10a), (10b), and (10e) lead to

$$\begin{aligned} f'(x) &= \frac{1}{c_0} (M \pm \sqrt{1 + M^2}), \quad g'(x) = \frac{1}{c_0} (-M \pm \sqrt{1 + M^2}), \\ H(x) &= \frac{(c_0/2)^2}{1 + M^2}. \end{aligned} \quad (11)$$

With that, Eqs. (10c) and (10d) become, respectively,

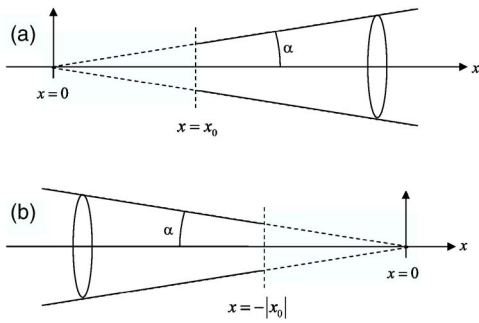


FIG. 2. (Color online) Divergent (a) and convergent (b) conical tubes.

$$\frac{1}{1+M^2} \frac{dM}{dx} = -\frac{M}{A} \frac{dA}{dx},$$

$$\frac{M}{1+M^2} \frac{dM}{dx} = 0.$$

For low values of M , the fluid behaves as incompressible, and so the Mach number and the cross-section area are inversely proportional (or, what is the same, the product MA is constant). Therefore:

$$\frac{1}{A} \frac{dA}{dx} = -\frac{1}{M} \frac{dM}{dx}.$$

This transforms Eq. (10c) into

$$\frac{M^2}{1+M^2} \frac{dM}{dx} = 0.$$

The left-hand sides of Eqs. (10c) and (10d) are of order M^3 and M^2 , respectively, and so are consistent with the zero value appearing at the right-hand side. Thus the system of Eq. (10) has no compatibility problems.

III. INWARD AND OUTWARD PROPAGATION VELOCITIES

The inward and outward propagation velocities can be obtained straightforward from Eq. (11). For low Mach numbers, it simplifies to

$$f'(x) = \frac{1}{c_+(x)} \approx \frac{1}{c_0}(M \pm 1), \quad g'(x) = \frac{1}{c_-(x)} \approx \frac{1}{c_0}(-M \pm 1);$$

$$H(x) \approx \left(\frac{c_0}{2}\right)^2. \quad (12)$$

Of course there are not two different solutions for $f(x)$ and $g(x)$ (as the \pm signs seem to suggest) but just one. Taking one sign or another gives exactly the same results, though with interchanged names. Thus only the upper sign will be kept from now on. The inward and outward propagation velocities are

$$c_+(x) = \frac{c_0}{1+M} \approx c_0(1-M), \quad c_-(x) = \frac{c_0}{1-M} \approx c_0(1+M). \quad (13)$$

As the Mach number is inversely proportional to the tube cross-section area, the outward propagation velocity

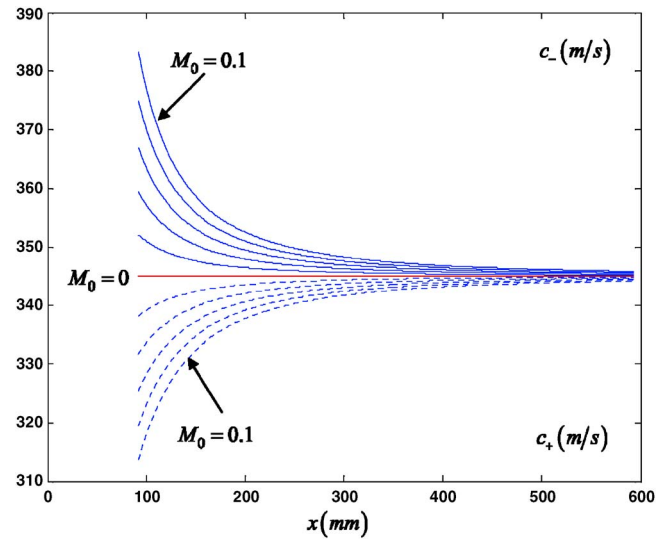


FIG. 3. (Color online) Outward (continuous) and inward (discontinuous) propagation speeds for the divergent cone with $\alpha = \pi/10$ in Fig. 2(a) and five different values of $M(x_0)$: $M_0 = \{0, 0.02, 0.04, 0.06, 0.08, 0.1\}$.

increases/decreases when the cone is convergent/divergent. The inward propagation velocity shows exactly the opposite behavior. For the cylindrical case, $M = M_0$, and so $c_+ = c_0(M_0 - 1)$, $c_- = c_0(M_0 + 1)$, as expected.

Figure 3 shows the x dependence of those propagation velocities for the divergent cone with $\alpha = \pi/10$. Six different cases are considered, according to the different maximum values of M (which happen at $x = 100$ mm). The reference value c_0 has been taken as $c_0 = 345$ m/s.

A simple integration yields the approximate coordinates (τ_+, τ_-) for the conical duct:

$$\begin{aligned} (\tau_+, \tau_-) &\approx \left(t + \frac{1}{c_0} \int (1+M) dx, t - \frac{1}{c_0} \int (1-M) dx \right) \\ &= \left(t - t_0 + \frac{(1-M)x}{c_0} - \frac{(1-M_0)x_0}{c_0}, t - t_0 \right. \\ &\quad \left. - \frac{(1+M)x}{c_0} + \frac{(1+M_0)x_0}{c_0} \right) \approx \left(t - t_0 \right. \\ &\quad \left. + \frac{x}{c_0(1+M)} - \frac{x_0}{c_0(1+M_0)}, t - t_0 - \frac{x}{c_0(1-M)} \right. \\ &\quad \left. + \frac{x_0}{c_0(1-M_0)} \right). \quad (14) \end{aligned}$$

For the particular case of a time harmonic dependence, the traveling waves would be

$$\begin{aligned} \theta(x, t) &= \theta_+(\tau_+) + \theta_-(\tau_-) = [A_+ e^{jk_+(x)x} e^{-jk_+(x_0)x_0} \\ &\quad + A_- e^{-jk_-(x)x} e^{jk_-(x_0)x_0}] e^{j\omega(t-t_0)}, \end{aligned}$$

with

$$\begin{aligned} k_+(x) &= \frac{\omega}{c_0(1+M)} \approx k_0(1-M), \quad k_-(x) = \frac{\omega}{c_0(1-M)} \\ &\approx k_0(1+M), \quad k_0 = \omega/c_0. \end{aligned}$$

The mean flow is thus responsible for a phase shift in the

acoustic variables, as pointed out by Campos.¹⁴ This result coincides with that obtained by Davies and Doak.⁸

Once the x dependence of (c_+, c_-) is known, it is possible to calculate the time-domain responses of anechoic conical ducts. In general, the acoustic pressure and the acoustic velocity contain two traveling terms:

$$p(x, t) = p_+(x, t) + p_-(x, t), \quad v(x, t) = v_+(x, t) + v_-(x, t).$$

Using Eqs. (4) and (13) and the fact that

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial \tau_+} + \frac{\partial}{\partial \tau_-}, \quad \frac{\partial}{\partial x} = \frac{1+M}{c_0} \frac{\partial}{\partial \tau_+} - \frac{1-M}{c_0} \frac{\partial}{\partial \tau_-},$$

the traveling pressure and velocity waves can be written as

$$p_{\pm}(x, t) \approx \frac{1}{\sqrt{A(x)}} \left\{ (1 \pm M) \frac{d\theta_{\pm}}{d\tau_{\pm}} - \frac{c_0}{x} M \theta_{\pm} \right\}, \quad (15)$$

$$v_{\pm}(x, t) = \frac{1}{Z_0 \sqrt{A(x)}} \left\{ \mp (1 \pm M) \frac{d\theta_{\pm}}{d\tau_{\pm}} + \frac{c_0}{x} \theta_{\pm} \right\}, \quad (16)$$

where the terms of order M^2 have been neglected.

IV. RESPONSE OF AN ANECHOIC CONE TO AN IMPULSE INPUT VELOCITY

The outward and inward pressure impulse responses at the section located at $x=x_0$ [denoted by $h_{p-}^v(t)$ and $h_{p+}^v(t)$, respectively] correspond to the pressure generated by an input velocity impulse produced at the same section. For the case of the outward response, the initial conditions can be formulated as

$$\theta_-^v(x > x_0, t \leq t_0) = 0,$$

$$v_-^v(x_0, t) = q_v \delta(t - t_0) = \frac{1}{Z_0} \frac{1}{\sqrt{A_0}} \left[(1 - M_0) \frac{d\theta_-^v}{d\tau_-} + \frac{c_0}{x_0} \theta_-^v(\tau_-) \right]_{x_0}, \quad (17)$$

where q_v is the amplitude of the velocity impulse. The fact that the tube is anechoic implies the absence of inward wave: $\theta_-^v(x > x_0, t) = 0$.

The solution of Eq. (17) is

$$\theta_-^v(x_0, t) = q_0 Z_0 \sqrt{A_0} \frac{1}{1 - M_0} \varepsilon(t - t_0) \exp\left(\frac{c_0}{x_0} \frac{1}{1 - M_0} (t - t_0)\right),$$

where $\varepsilon(t - t_0)$ is the Heaviside step distribution. As $\theta_-^v = \theta_-^v(\tau_-)$, this solution can be extended to any value of the spatial coordinate:

$$\theta_-^v(\tau_-) = q_0 Z_0 \sqrt{A_0} \frac{1}{1 - M_0} \varepsilon(\tau_-) \exp\left(\frac{c_0}{x_0} \frac{1}{1 - M_0} \tau_-\right). \quad (18)$$

TABLE I. Outward pressure and velocity time-domain responses of a conical tube to an impulsive velocity input with steady mean flow.

$$p_-^v(x, t) \approx q_v Z_0 \frac{\sqrt{A_0}}{\sqrt{A}} \frac{1}{1 - M_0} \left\{ (1 - M) \delta(\tau_-) - \left(\frac{c_0}{x_0} \frac{1 - M}{1 - M_0} + \frac{c_0}{x} M \right) \times \varepsilon(\tau_-) \exp\left(-\frac{c_0}{x_0} \frac{1}{1 - M_0} \tau_-\right) \right\}$$

$$v_-^v(x, t) = q_v \frac{\sqrt{A_0}}{\sqrt{A}} \frac{1}{1 - M_0} \left\{ (1 - M) \delta(\tau_-) - \left(\frac{c_0}{x_0} \frac{1 - M}{1 - M_0} - \frac{c_0}{x} \right) \times \varepsilon(\tau_-) \exp\left(-\frac{c_0}{x_0} \frac{1}{1 - M_0} \tau_-\right) \right\}$$

$$h_{p-}^v(t) = \frac{1}{q_v} p_-^v(x_0, t) = Z_0 \delta(t - t_0) - Z_0 \frac{1 + M_0 c_0}{1 - M_0 x_0} \times \varepsilon(t - t_0) \exp\left(-\frac{c_0}{x_0} \frac{1}{1 - M_0} (t - t_0)\right)$$

where

$$M(x) = \frac{M_0 A_0}{\pi [R_0 + m(x - x_0)]^2}$$

The pressure and velocity outward waves, $p_-^v(x, t)$ and $v_-^v(x, t)$, respectively, can be calculated from $\theta_-^v(\tau_-)$ according to Eqs. (15) and (16). The outward pressure impulse response is simply $h_{p-}^v(t) = (1/q_v) p_-^v(x_0, t)$. Table I presents the analytical expressions for these three responses for $M_0 \neq 0$. Their particularizations to $M_0 = 0$ coincide with the time-domain responses of a conical duct without mean flow obtained by Agulló *et al.*²⁰

The mean flow does not have any consequences on the qualitative behavior of these responses: The initial impulse is always followed by an exponential wake. For divergent cones, the x coordinates always take positive values, and so the exponential is a decreasing one. For convergent cones, however, $x_0 < 0$ and the time responses diverge.

Figure 4 shows $p_-^v(x, t)$ for a divergent and a convergent cone with angle aperture $\alpha = (\pi/10)$ rad and three different maximum values of M_0 (0.1, 0.06, and 0.02). The initial impulses have been suppressed (to better visualize the exponential evolution) and the amplitudes have been normalized to the maximum amplitude of the $M_0 = 0.1$ case.

Figures 5(a) and 5(b) show, respectively, the outward pressure impulse response $h_{p-}^v(t)$ for the five divergent conical ducts shown in Fig. 6 and their corresponding convergent ones, for the same $M_0 = 0.1$ value. The initial impulses have been omitted, and the amplitudes have been normalized to that of the cone with $\alpha = (\pi/10)$ rad, which will be taken as reference case. Having taken the same M_0 value, the exponential growth rate will depend only on the x_0 value. As the cone angle decreases [$\alpha < (\pi/10)$ rad], the x_0 value increases and the impulse response becomes smoother as compared to

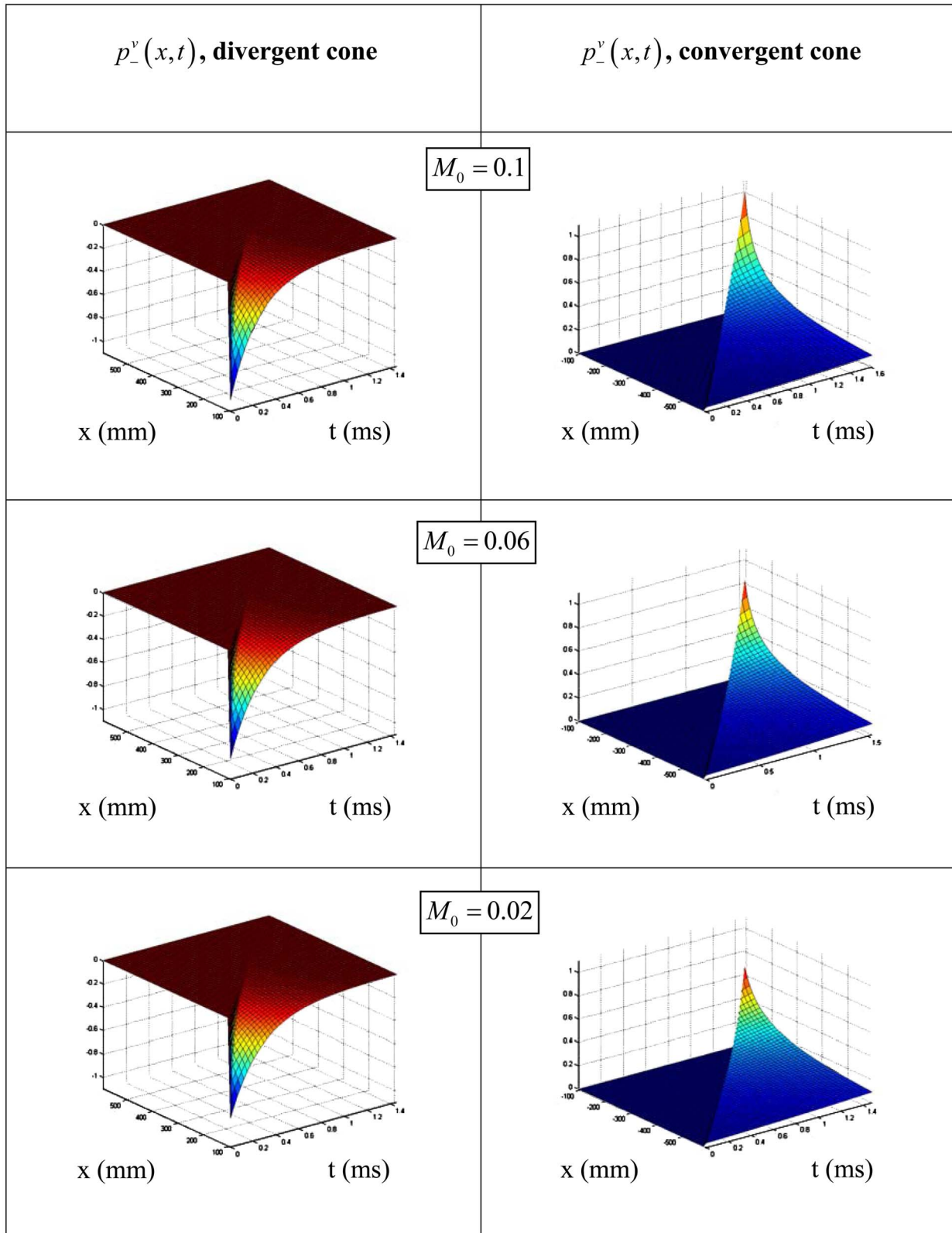


FIG. 4. (Color online) Progressive pressure $p^v(x,t)$ generated by a velocity impulse for the divergent (left) and convergent (right) cones depicted in Fig. 2 (cone angle $\alpha = \pi/10$) and different Mach numbers. The initial impulses have been suppressed and the amplitudes have been normalized to the maximum amplitude of the $M_0 = 0.1$ case.

the reference case: In Fig. 5(a) the exponential wake is weaker, and in Fig. 5(b) the exponential growth rate is lower. As the angle aperture α tends to zero (cylindrical tube), both the exponential wake in Fig. 5(a) and the exponential growth in Fig. 5(b) tend to disappear.

The same kind of calculations leads to the inward pressure impulse response. The results are given in Table II. No qualitative differences due to the mean flow can be observed. Again the zero mean flow expressions coincide with those given in Ref. 20.

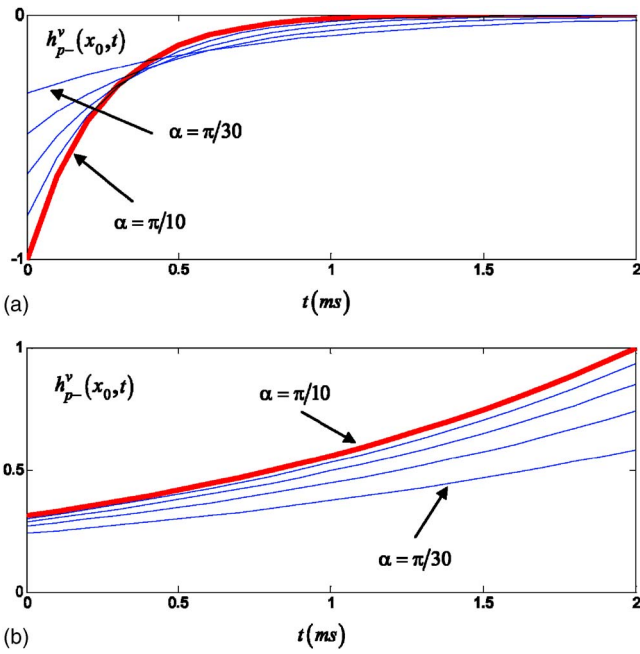


FIG. 5. (Color online) Outward pressure impulse response $h_p^v(t)$ due to an impulsive velocity for the five divergent cones (a) shown in Fig. 6 and the corresponding convergent (b) ones with $M_0=0.1$. The initial impulses have been omitted, and the amplitudes have been normalized to that of the cone with $\alpha=(\pi/10)$ rad.

Tables I and II show that the inward waves for a divergent cone ($x > 0$) are very close in shape to the outward ones in a convergent cone ($x < 0$), so no graphical representations will be shown.

V. RESPONSE OF AN ANECHOIC CONE TO AN IMPULSE INPUT PRESSURE

The outward and inward velocity impulse responses at section located at $x=x_0$ [denoted by $h_{v-}^p(t)$ and $h_{v+}^p(t)$, respectively] correspond to the velocity generated by an input pressure impulse produced at the same section. Their calculation parallels that presented for the pressure impulse responses.

For the case of the outward response, the initial conditions are

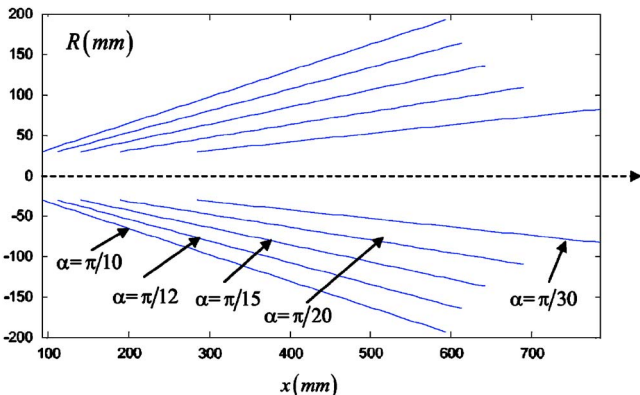


FIG. 6. (Color online) Five divergent conical tubes with the same input section but different apertures.

TABLE II. Inward pressure and velocity time-domain responses of a conical tube to an impulsive velocity input with steady mean flow.

$$p_+^v(x, t) \approx -q_v Z_0 \frac{\sqrt{A_0}}{\sqrt{A}} \frac{1}{1+M_0} \left\{ (1+M) \delta(\tau_+) + \left(\frac{c_0}{x_0} \frac{1+M}{1+M_0} - \frac{c_0}{x} M \right) \varepsilon(\tau_+) \exp\left(\frac{c_0}{x_0} \frac{1}{1+M_0} \tau_+ \right) \right\}$$

$$v_+^v(x, t) = q_v \frac{\sqrt{A_0}}{\sqrt{A}} \frac{1}{1+M_0} \left\{ (1+M) \delta(\tau_+) + \left(\frac{c_0}{x_0} \frac{1+M}{1+M_0} - \frac{c_0}{x} \right) \varepsilon(\tau_+) \exp\left(\frac{c_0}{x_0} \frac{1}{1+M_0} \tau_+ \right) \right\}$$

$$h_{p+}^v(t) = \frac{1}{q_v} p_+^v(x_0, t) = -Z_0 \delta(t-t_0) - Z_0 \frac{c_0}{x_0} \frac{1-M_0}{1+M_0} \varepsilon(t-t_0) \exp\left(\frac{c_0}{x_0} \frac{1}{1+M_0} (t-t_0) \right)$$

where

$$M(x) = \frac{M_0 A_0}{\pi [R_0 + m(x-x_0)]^2}$$

$$\theta_-^p(x > x_0, t \leq t_0) = 0,$$

$$p_-^p(x_0, t) = q_p \delta(t-t_0) = \frac{1}{\sqrt{A_0}} \left\{ (1-M_0) \left(\frac{d\theta_-^p}{d\tau_-} \right)_{x_0} - \frac{c_0}{x_0} M_0 \theta_-^p(x_0, t) \right\}, \quad (19)$$

TABLE III. Outward pressure and velocity time-domain responses of a conical tube to an impulsive pressure input with steady mean flow.

$$p_+^p(x, t) \approx q_p \frac{\sqrt{A_0}}{\sqrt{A}} \frac{1}{1-M_0} \left\{ (1-M) \delta(\tau_-) + \left(\frac{c_0}{x_0} \frac{M_0(1-M)}{1-M_0} - \frac{c_0}{x} M \right) \varepsilon(\tau_-) \exp\left(\frac{c_0}{x_0} \frac{1}{1-M_0} \tau_- \right) \right\}$$

$$v_+^p(x, t) = \frac{q_p \sqrt{A_0}}{Z_0 \sqrt{A}} \frac{1}{1-M_0} \left\{ (1-M) \delta(\tau_-) + \left(\frac{c_0}{x_0} \frac{M_0(1-M)}{1-M_0} + \frac{c_0}{x} \right) \varepsilon(\tau_-) \exp\left(\frac{c_0}{x_0} \frac{1}{1-M_0} \tau_- \right) \right\}$$

$$h_{v-}^p(t) = \frac{1}{q_p} v_+^p(x_0, t) = \frac{1}{Z_0} \delta(t-t_0) + \frac{1}{Z_0 x_0} \frac{c_0}{1-M_0} \varepsilon(t-t_0) \exp\left(\frac{c_0}{x_0} \frac{M_0}{1-M_0} (t-t_0) \right)$$

where

$$M(x) = \frac{M_0 A_0}{\pi [R_0 + m(x-x_0)]^2}$$

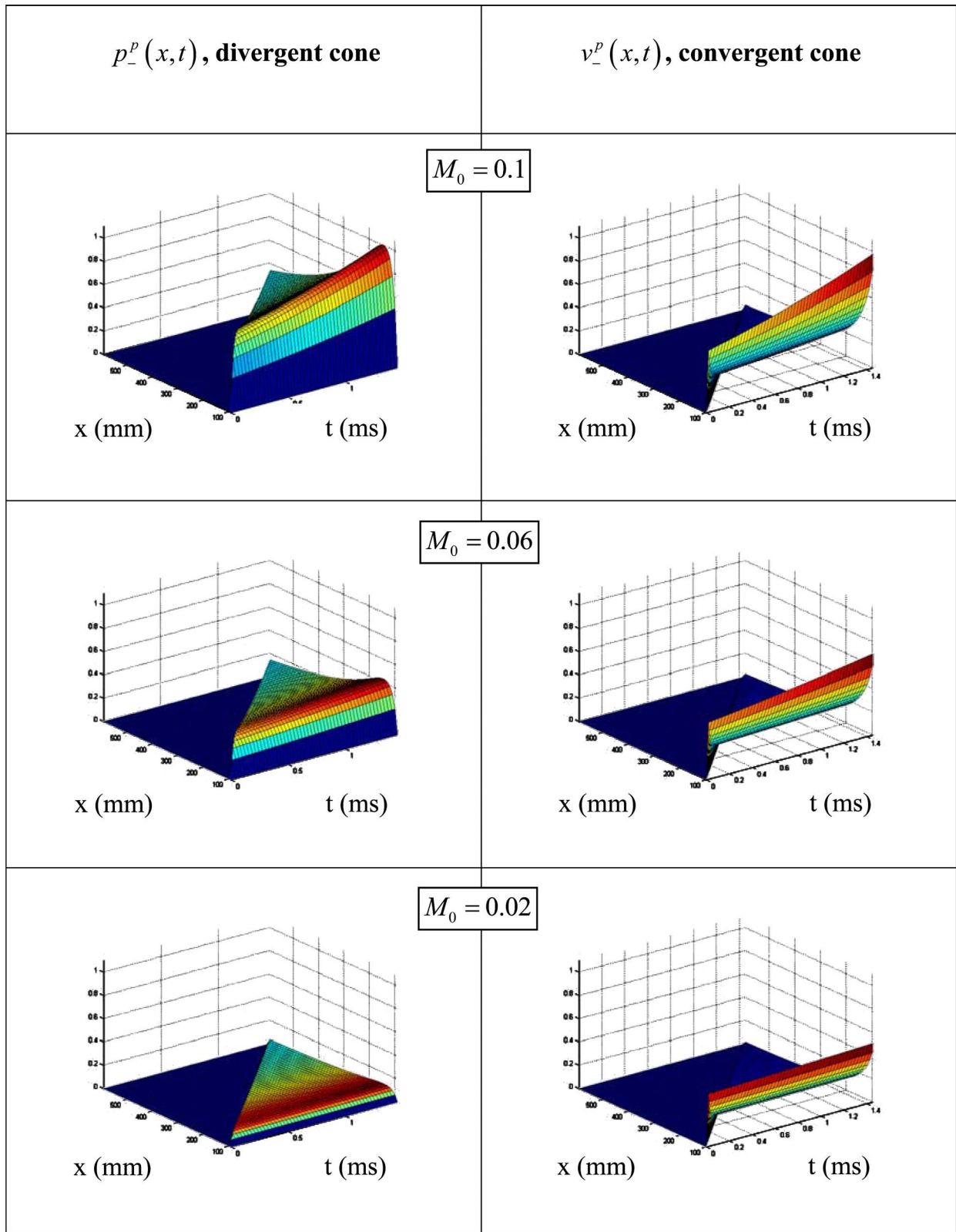


FIG. 7. (Color online) Progressive pressure $p^p(x,t)$ and velocity $v^p(x,t)$ generated by a pressure impulse for the divergent cone depicted in Fig. 2 and different Mach numbers. The initial impulses have been suppressed and the amplitudes have been normalized to the maximum amplitude of the $M_0=0.1$ case.

where q_p is the amplitude of the pressure impulse. Again $\theta_+^p(x,t)=0$.

The integration of Eq. (19) leads to the general outward potential function:

$$\theta_-^p(\tau_-) = q_p \frac{\sqrt{A_0}}{1 - M_0} \varepsilon(\tau_-) \exp\left(\frac{c_0}{x_0} \frac{1}{1 - M_0} \tau_-\right).$$

The resulting pressure and velocity outward waves and the outward velocity impulse response are presented in Table III.

Unlike the case treated in the previous section, the mean flow does change the qualitative behavior of these responses: Whereas the pressure impulse remained an impulse as it propagated in the absence of mean flow, it spreads exponentially whenever $M_0 \neq 0$! This surprising behavior can be understood from the qualitative analysis of the outward impulse responses $h_{p-}^v(t)$ and $h_{v-}^p(t)$.

Let us consider the case of a divergent cone with zero mean flow. The impulse response $[h_{p-}^v(t)]_{M=0}$ generated by an impulsive velocity (which can be obtained from the general case presented in Table I) shows a negative exponential wake [with exponent $-(c_0/x_0)$]. The compensation of that depression calls for the injection of an air flow at the same section. This is why the pressure impulse response $[h_{p-}^p(t)]_{M=0}$ (which can be obtained from Table III) contains the step distribution $(1/Z_0)(c_0/x_0)\varepsilon(t-t_0)$.

The existence of a nonzero mean flow leads to an impulse response $h_{p-}^v(t)$ whose wake has an exponent $-(c_0/x_0) \times (1-M_0)^{-1}$. As $(1-M_0)^{-1} > 1$, the depression compensation calls for a flow injection higher than just the step $(1/Z_0) \times (c_0/x_0)\varepsilon(t-t_0)$. Thus it is not surprising that the $h_{v-}^p(t)$ in the upper part of Table III contains an increasing exponential. This exponential results in an exponential growth in $p^p(x,t)$. Again, the behavior of convergent cones is opposite to that of the divergent cones due to the negative sign of the spatial coordinate.

Figure 7 shows $p_{-}^p(x,t)$ and $v_{-}^p(x,t)$ for a divergent cone with angle aperture $\alpha=(\pi/10)$ rad and three different maximum values of M_0 (0.1, 0.06, and 0.02). The initial impulses have been suppressed (to better visualize the exponential evolution) and the amplitudes have been normalized to the maximum amplitude of the $M_0=0.1$ case.

Figure 8 shows the outward and inward velocity impulse responses, $h_{v-}^p(t > t_0)$ and $h_{v+}^p(t > t_0)$, respectively, for a conical divergent tube with angle aperture $\alpha=(\pi/10)$ rad, and different values of M_0 . The initial impulses have been omitted, and the amplitudes have been normalized to that of $M_0=0.1$. As the Mach number tends to $M_0=0$, the diverging exponential tends to a step velocity, as expected.

The inward responses to a pressure impulse are given in Table IV. The same features associated with the nonzero mean flow appear, so no graphics will be shown.

VI. CONCLUSIONS

The analytical time-domain responses of anechoic conical ducts with a small steady mean flow were derived exclusively based on a time-domain calculation from the 1D linear wave equation. They reproduce the well-known cases of cylindrical tubes and those of conical tubes without mean flow.

In the absence of mean flow, the time-domain responses of conical tubes show a stable behavior whenever the acoustic wave propagates in the widening area direction, but they become exponentially divergent when they propagate in the narrowing area one. Moreover, a pressure impulse keeps its impulse shape as it propagates in either direction.

The mean flow can drastically change these behaviors. Thus, the pressure wave generated by an impulsive velocity input diverges regardless of the widening or narrowing char-

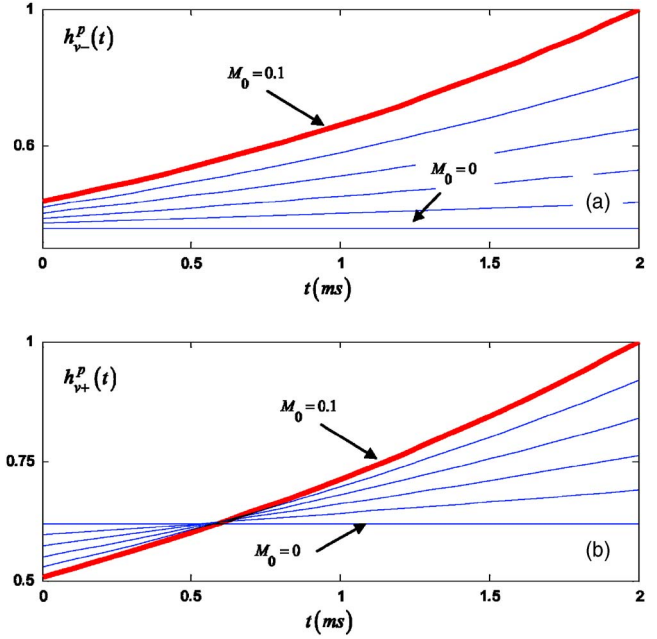


FIG. 8. (Color online) Outward (a) and inward (b) velocity impulse responses $[h_{v-}^p(t)$ and $h_{v+}^p(t)$, respectively] due to an impulsive pressure for a conical divergent tube with angle aperture $\alpha=(\pi/10)$ rad, and different M_0 : {0 0.02 0.04 0.06 0.08 0.1}. The initial impulses have been omitted, and the amplitudes have been normalized to that of $M_0=0.1$.

acter of the area in the propagation direction. Pressure impulses do not remain unchanged in shape but spread as they propagate. A qualitative physical interpretation of the main features of these results is proposed, based on the zero flow responses.

¹L. K. Schubert, "Numerical study of sound refraction by a jet flow. II. Wave acoustics," J. Acoust. Soc. Am. **51**, 447–463 (1972).

TABLE IV. Inward pressure and velocity time-domain responses of a conical tube to an impulsive pressure input with steady mean flow.

$$p_{+}^p(x,t) \approx q_p \frac{\sqrt{A_0}}{\sqrt{A}} \frac{1}{1+M_0} \left\{ (1+M) \delta(\tau_+) + \left\{ \frac{c_0 M_0 (1+M)}{x_0} \frac{1}{1+M_0} - \frac{c_0}{x} M \right\} \varepsilon(\tau_+) \exp\left(\frac{c_0}{x_0} \frac{M_0}{1+M_0} \tau_+\right) \right\}$$

$$v_{+}^p(x,t) = -\frac{q_p \sqrt{A_0}}{Z_0 \sqrt{A}} \frac{1}{1+M_0} \left\{ (1+M) \delta(\tau_+) + \left\{ \frac{c_0 M_0 (1+M)}{x_0} \frac{1}{1+M_0} - \frac{c_0}{x} \right\} \varepsilon(\tau_+) \exp\left(\frac{c_0}{x_0} \frac{M_0}{1+M_0} \tau_+\right) \right\}$$

$$h_{v+}^p(t) = \frac{1}{q_p} v_{+}^p(x_0,t) = -\frac{1}{Z_0} \delta(t-t_0) + \frac{1}{Z_0 x_0} \frac{1-M_0}{1+M_0} \varepsilon(t-t_0) \exp\left(\frac{c_0}{x_0} \frac{M_0}{1+M_0} (t-t_0)\right)$$

where

$$M(x) = \frac{M_0 A_0}{\pi [R_0 + m(x-x_0)]^2}$$

- ²A. D. Pierce, "Wave equation for sound in fluids with unsteady inhomogeneous flow," *J. Acoust. Soc. Am.* **87**, 2292–2299 (1990).
- ³A. Selamet, N. S. Dickey, and J. M. Novak, "A time-domain computational simulation of acoustical silencers," *J. Vibr. Acoust.* **117**, 323–331 (1995).
- ⁴J. M. Middelberg, T. J. Barber, S. S. Leong, K. P. Byrne, and E. Leonardi, "Computational fluid dynamics analysis of the acoustic performance of various simple expansion chamber mufflers," *Proceedings of Acoustics 2004*, Gold Coast, Australia, pp. 123–127.
- ⁵T. Y. Lung and A. G. Doige, "A time averaging transient testing method for acoustic property of piping systems and mufflers with flow," *J. Acoust. Soc. Am.* **73**, 867–876 (1983).
- ⁶M. Willatzen, "The influence of a liquid flow on sound fields confined by conical walls," *J. Sound Vib.* **248**, 847–863 (2001).
- ⁷L. M. B. C. Campos, "On the fundamental acoustic mode in variable area, low Mach number nozzles," *Prog. Aerosp. Sci.* **22**, 1–27 (1985).
- ⁸P. O. A. L. Davies and P. E. Doak, "Spherical wave propagation in a conical pipe with mean flow," *J. Sound Vib.* **137**, 343–346 (1990).
- ⁹V. Easwaran and M. L. Munjal, "Plane wave analysis of conical and exponential pipes with incompressible mean flow," *J. Sound Vib.* **152**, 73–93 (1992).
- ¹⁰N. A. Eisenberg and T. W. Kao, "Propagation of sound through a variable-area duct with a steady compressible flow," *J. Acoust. Soc. Am.* **49**, 169–175 (1971).
- ¹¹V. Easwaran and M. L. Munjal, "Transfer matrix modelling of hyperbolic and parabolic ducts with incompressible mean flow," *J. Acoust. Soc. Am.* **90**, 2163–2172 (1991).
- ¹²E. Dokumaci, "An exact transfer matrix formulation of plane sound wave transmission in inhomogeneous ducts," *J. Sound Vib.* **217**, 869–882 (1998).
- ¹³J. Zehnlin and S. Jiazheng, "Four-pole parameters of a duct with low Mach number flow," *J. Acoust. Soc. Am.* **98**, 2848–2850 (1995).
- ¹⁴L. M. B. C. Campos, "On longitudinal acoustical propagation in convergent and divergent nozzle flows," *J. Sound Vib.* **117**, 131–151 (1987).
- ¹⁵A. H. Shapiro, *The Dynamics and Thermodynamics of Compressible Fluid Flow* (Krieger, Malabar, 1954, re-issued 1985), pp. 907–911.
- ¹⁶J. Agulló, A. Barjau, and J. Martínez, "Alternatives to the impulse response $h(t)$ to describe the acoustical behavior of conical ducts," *J. Acoust. Soc. Am.* **84**, 1606–1612 (1988).
- ¹⁷D. P. Berners, "Acoustics and signal processing techniques for physical modeling of brass instruments," Ph.D. dissertation, Stanford University, Stanford, CA.
- ¹⁸T. Kar and M. L. Munjal, "Analysis and design of conical concentric tube resonators," *J. Acoust. Soc. Am.* **116**, 74–83 (2004).
- ¹⁹S. N. Y. Gerges, R. Jordan, F. A. Thieme, J. L. Bento Coelho, and J. P. Arenas, "Muffler modeling by transfer matrix method and experimental verification," *J. Braz. Soc. Mech. Sci. & Eng.* **27**, 132–140 (2005).
- ²⁰J. Agulló, A. Barjau, and J. Martínez, "On the time-domain description of conical bores," *J. Acoust. Soc. Am.* **91**, 1099–1105 (1992).
- ²¹S. W. Rienstra, "Webster's horn equation revisited," *SIAM J. Appl. Math.* **65**, 1981–2004 (2005).
- ²²M. L. Munjal, *Acoustics of Ducts and Mufflers* (Wiley, New York, 1987), pp. 17–19.

Low frequency sound scattering from spherical assemblages of bubbles using effective medium theory

Thomas R. Hahn

University of Miami, RSMAS, AMP, 4600 Rickenbacker Causeway, Miami, Florida 33149

(Received 22 March 2005; revised 29 August 2005; accepted 6 September 2007)

The determination of the acoustic field scattered by an underwater assembly of gas bubbles or similar resonant monopole scatterers is of considerable theoretical and practical interest. This problem is addressed from a theoretical point of view within the framework of the effective medium theory for the case of spherically shaped assemblages. Although being valid more generally, the effective medium theory is an ideal instrument to study multiple scattering effects such as low frequency collective resonances, acoustically coupled breathing modes of the entire assembly. Explicit expressions for the scattering amplitude and cross sections are derived, as well as closed form expressions for the resonance frequency and spectral shape of the fundamental collective mode utilizing analytical S -matrix methods. This approach allows, in principle, a simultaneous inversion for the assembly radius and void fraction directly from the scattering cross sections. To demonstrate the validity of the approach, the theory is applied to the example of idealized, spherically shaped schools of swim bladder bearing fish. The analytic results of the theory are compared to numerical first-principle benchmark computations and excellent agreement is found, even for densely packed schools and frequencies across the bladder resonance. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2793610]

PACS number(s): 43.30.Ft, 43.20.Fn, 43.30.Sf [KGF]

Pages: 3252–3267

I. INTRODUCTION

Assemblages of gas bubbles of various kinds are an important part of the active and passive acoustic environment in the ocean. Bubble plumes of several different types have been studied from an acoustic point of view and have received strong attention in the literature for many decades. Among these are active^{1,2} and passive^{3–8} acoustic studies of bubble assemblages generated by breaking surface gravity waves. Other areas that have sparked theoretical and observational interest in the acoustics of bubble clouds are the scattering of low frequency sound from schools of swim bladder bearing fish^{9–15} as well as acoustic location and monitoring of plumes of rising methane bubbles^{16,17} escaping from methane seeps associated with hydrate occurrences at the seafloor. Although the latter concerns mainly backscattering applications using conventional single-, split-, or dual-beam echo sounders mounted on the seafloor or on the survey vessel, passive observations have also been explored theoretically.¹⁸ Assemblages of bubbles are not only observed in the water column of the oceans but can also occur as bubble clouds embedded in the sediments of the seafloor, e.g., methane bubbles generated by biological processes, which cause significant alterations of the acoustic backscatter.^{19–21}

If the void fractions in the bubble assemblages are large enough, interesting low frequency physics occurs that has its origin in the effects of multiple scattering. This has first been the center of attention when it was realized that wind-driven low frequency noise^{22–25} exists in the ocean that is due to bubbles created by breaking waves that cannot be explained by the resonance of individual bubbles. It was suggested^{7,26–28} that collective modes of oscillations of the

entire bubble cloud provide the mechanism responsible for this low frequency ambient noise component. Subsequently, this hypothesis found support repeatedly in acoustic measurements from laboratory breaking waves^{3–6} as well as recordings from various bubble clouds generated by other entrainment mechanisms^{29–31} or injection³² and is now a standard component in the modeling of wind generated ambient noise.³³ Figure 1, reproduced from Hahn *et al.*,³⁰ shows the power spectral density of the noise produced by an air entraining plunging jet. The plume has no artificial boundaries and, at low frequencies, the spectra are dominated by collective modes.

Next to this focus on passive acoustic emissions, broadband backscattering returns^{1,2} from bubble assemblages have been studied in the same context. It was found that, at low-frequencies, the measured target strengths (TS) are (a) well explained by collective monopole modes, (b) that average backscattering levels do not depend much on the precise shape of the cloud but rather on the overall linear dimensions, and (c) that Born approximations of the scattering amplitudes become invalid at higher void fractions.

All these phenomena are closely related to the alteration of the phase speed and attenuation in the region occupied by the bubbles.^{34–36} This change in the bulk parameters due to the presence of bubbles is the large-scale manifestation of multiple scattering.³⁷ In this light, emission of sound and scattering of acoustic waves from assemblages of bubbles are simply the response of a resonator, formed by an impedance contrast between the bubbly region and the surrounding bubble-free medium, to different forms of excitations. These are the fundamental assertions of the effective medium theory (EMT), which traces back to the pioneering work of Foldy.³⁸ These insights, derived from the effective medium

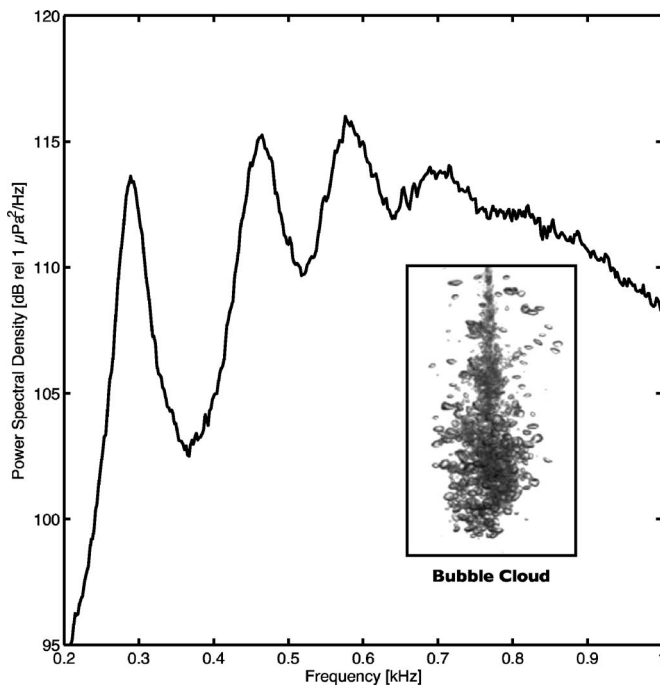


FIG. 1. Power spectral density of the acoustic signal recorded close to the bubble cloud generated by a plunging water jet exhibiting strong low frequency resonances due to the impedance contrast of the bubbly region and the surrounding essentially bubble free water. The jet impinges with a velocity of about 10 m/s. The inset shows the corresponding bubble cloud, which extends vertically to about 20 cm depth.

picture, physically explain the occurrences of low frequency emissions and TS enhancements of assemblages at frequency below the resonance frequency of the individual bubbles. In consequence, the EMT has successfully been applied to layers and plumes of bubbles formed by breaking waves.³⁹ Based on this resonator picture, several other more controlled systems have also been modeled, and compared favorably to observations.^{30,32}

From this discussion, it becomes clear that the effects of multiple scattering significantly enrich the scattering and radiation behavior of assemblages of bubbles. This encourages unique ways to acoustically explore these systems that go beyond traditional high frequency surveys. If observed, resonance frequencies of collective modes scale, for a given geometry, with the linear dimension of the assembly and the void fraction of the bubbly region:³⁹ if one is known, the other one can be obtained from the measurement of the resonance peaks. This opens up the possibility of inversions, for example, for the void fraction.³⁰ However, the involved eigenvalue problems are usually solved numerically for the (complex valued) resonance frequencies, which limits direct physical insight and makes inversions cumbersome. Additionally, spectral features other than the mere resonance *peak* have, to the knowledge of the author, not been explicitly treated.

This paper addresses these issues by giving fully analytic expressions for the entire resonance shape of isolated fundamental modes of spherically shaped bubble assemblages. This increases the power of inversions that can be made directly from the spectrum. *Both* features, linear dimension *and* void fraction, can ideally be obtained if the

scattering characteristics of the constituents are known. This, in principle, allows a counting of the total constituent numbers.

It is also a fact that the EMT not only gives a correct description of the scattering and radiation behavior of bubble assemblages in the regime of dominant collective modes, but also if these are due to high damping or small number densities not manifest in the acoustic field and also in the case of high frequencies. That is, when multiple scattering effects are not predominant or not fully developed. This makes an investigation of this transition regime within the EMT possible.

To highlight the potential of these findings, this paper treats, as an example, an idealized spherical school of swim bladder bearing fish. For this case, the applicability of the effective medium approach is investigated in detail and first-principle, direct computations are used as a benchmark. Further, the spectral modeling and its use for inversions as well as the scattering behavior in the transition regime with its associated physical effects are demonstrated in detail.

The main body of the paper is divided into two parts: in Sec. II, the general theory is developed and Sec. III discusses the exemplary application to spherical fish schools. For completeness, the theoretical development is preceded by a brief review of linear bubble dynamics and a discussion of numerical benchmark solutions that also set the stage for computing the cross sections within framework of the EMT.

II. THEORY

A. Linear bubble dynamics

For reasons of completeness and reference we begin with a brief review of those elements of linear bubble dynamics that are most relevant to the following theoretical treatment. More complete reviews can be found elsewhere, for example, in the book of Leighton.⁴⁰ The far-field scattered wave of an object exposed to an incoming plane wave with time dependence $e^{-i\omega t}$ and angular frequency ω is proportional to f_s/r . For resonant monopole radiation, the scattering amplitude f_s of a gas bubble of radius a is given by the familiar equation:^{40,41}

$$f_s = \frac{a}{\left[\left(\frac{\omega_0}{\omega} \right)^2 - 1 \right] - i\delta}, \quad (1)$$

which exhibits spherical symmetry and incorporates the scattering behavior close to the resonance frequency ω_0 due to various loss mechanisms prescribed by the dimensionless damping constant δ . The appropriate value for the resonance frequency ω_0 results from the detailed balance of stiffness and inertia:

$$\omega_0^2 = \frac{P_{b0}}{\rho a^2} \left[3\gamma_{\text{eff}} - \frac{2\sigma}{aP_{b0}} \right]. \quad (2)$$

Here, γ_{eff} is the frequency dependent effective polytropic exponent^{28,36} of the bubble gas cyclically undergoing compressions and rarefactions, $P_{b0} = P_0(z) + 2\sigma/a$ the bubble equilibrium pressure, which is the sum of the hydrostatic

pressure $P_0(z)$ at depth z and the pressure due to the surface tension σ , and ρ is the fluid density. If we ignore the effects of surface tension and the frequency dependence of the effective polytropic index, we obtain Minnaert's⁴² expression for the resonance frequency:

$$\omega_M^2 = \frac{3\gamma_{\text{eff}}P_0}{\rho a^2}. \quad (3)$$

Herein, γ_{eff} is now considered independent of the frequency or, for freely ringing bubbles, of the bubble radius.

The dimensionless damping constant δ most notably contains three frequency dependent terms,^{40,43}

$$\delta \approx \delta_{\text{rad}} + \delta_{\text{visc}} + \delta_{\text{th}} = ka + \frac{4\mu}{\rho\omega a^2} + \frac{P_{b0}}{\rho\omega^2 a^2} \text{Im } \phi, \quad (4)$$

radiation damping, δ_{rad} , responsible for the re-radiation of sound and the correct causal structure of f_s , viscous damping δ_{visc} due to viscous dissipation in the medium surrounding the bubble, and thermal³⁶ damping δ_{th} , resulting from the loss of heat to the medium as the bubble expands and contracts. In this treatment, $k=\omega/c$ is the wave number and assumed to satisfy $ka \ll 1$. The thermal effects are described by the complex valued function ϕ , whose real part gives the effective polytropic index $\gamma_{\text{eff}} = \text{Re } \phi/3$ in terms of the ratio of the specific heats, γ , and the thermal diffusivity D of the bubble gas:

$$\phi = \frac{3\gamma}{1 - 3(\gamma - 1)i\chi[(i\chi)^{1/2} \coth(i\chi)^{1/2} - 1]},$$

$$\chi = \frac{D}{\omega a^2}. \quad (5)$$

As long as the wavelength of the acoustic wave is, as assumed, much larger than the bubble size, higher order multipole (i.e., shape) oscillations can be ignored, and Eq. (1) will correctly model the ‘‘breathing’’ response of the bubble. The previous expression leads to the familiar resonance scattering cross section of individual bubbles:

$$\sigma_s^b = \frac{4\pi a^2}{\left[\left(\frac{\omega_0}{\omega} \right)^2 - 1 \right]^2 + \delta^2}. \quad (6)$$

The optical theorem allows the computation of the total cross section $\sigma_{\text{tot}}^b = \sigma_s^b + \sigma_e^b$ incorporating scattering *and* extinction from the scattering amplitude in forward direction. For the spherically symmetric scattering amplitude f_s , the optical theorem takes the simpler form $\sigma_{\text{tot}}^b = (4\pi/k)\text{Im } f_s$. From this and Eq. (6) we read off the extinction cross section that describes the removal of energy from the sound field due to thermal and viscous damping of the bubble:

$$\sigma_e^b = \frac{\delta_{\text{visc}} + \delta_{\text{th}}}{\delta_{\text{rad}}} \sigma_s^b. \quad (7)$$

The extinction cross section σ_e^b then describes the energy loss per unit time per incoming intensity due to energy dissipation for one single scatterer.

For air bubbles in sea water $\sigma=0.073 \text{ N m}^{-1}$, $\mu = 0.001 \text{ Pa s}$, the sound speed $c \approx 1500 \text{ m/s}$, $D=2.08E-5 \text{ m}^2 \text{ s}^{-1}$, $\rho \approx 1026 \text{ kg/m}^3$, and $\gamma=1.4$. For mm-sized bubbles in the low-kilohertz frequency regime, the changes of state of the bubble gas are nearly adiabatic and $\gamma_{\text{eff}} = 1.2-1.4$.

B. Numerical benchmark

For a given geometric configuration of an assembly of many scatterers, each with the acoustic scattering characteristics given by Eq. (1), the acoustic field at any point of observation can rigorously be calculated for any given source distribution; the typical scattering problem corresponding to a source at infinity. Let us assume an ensemble of N scatterers at positions \vec{r}_i . For simplicity, we shall assume *identical monopole* scatterers, all with scattering strength f_s ; both assumptions can be relaxed, albeit with increasing numerical complexity.^{44,45}

Following the self-consistent approach of Foldy,³⁸ we write the total pressure field at a point of observation \vec{r} as the sum of an incoming field $p_0(\vec{r})$ and a scattered field $p_s(\vec{r})$

$$p(\vec{r}) = p_0(\vec{r}) + p_s(\vec{r}) = p_0(\vec{r}) + \sum_{i=1}^N f_s p^i(\vec{r}_i) G(k; \vec{r} - \vec{r}_i), \quad (8)$$

in terms of the ‘‘external field on the i -th scatterer’’ $p^i(\vec{r}_i)$ evaluated at \vec{r}_i and the free-space Green’s function

$$G(k; \vec{r}_i - \vec{r}_j) = \frac{e^{ik|\vec{r}_i - \vec{r}_j|}}{|\vec{r}_i - \vec{r}_j|}. \quad (9)$$

On the other side, the external background field on the particular scatterer i is the sum of the incoming field $p_0(\vec{r})$ at its locus and the scattered field generated by *all the other* scatterers propagated to the position \vec{r}_i :

$$p^i(\vec{r}_i) = p_0(\vec{r}_i) + \sum_{\substack{j=1 \\ j \neq i}}^N f_s p^j(\vec{r}_j) G(k; \vec{r}_i - \vec{r}_j). \quad (10)$$

The distinction between the field ‘‘due to all the other scatterers’’ and the field radiated by a particular scatterer itself is meaningful, because one is regular and the other has a singularity at the considered point in space.

The fundamental equations of multiple scattering, Eqs. (8) and (10), form a set of N linear equations, with a solution for the scattered field, written in $N \times N$ -matrix form:

$$p_s(\vec{r}) = f_s [(\mathbf{1} - f_s \mathbf{G}^k)^{-1} \mathbf{p}_0]^T \cdot \mathbf{G}^k(\vec{r}), \quad (11)$$

with the matrix \mathbf{G}^k defined as

$$\mathbf{G}_{i,j}^k = \begin{cases} G(k; \vec{r}_i - \vec{r}_j), & i \neq j \\ 0, & i = j, \end{cases} \quad (12)$$

and the vectors \mathbf{p}_0 and $\mathbf{G}^k(\vec{r})$ as

$$\mathbf{p}_{0i} = p_0(\vec{r}_i),$$

$$\mathbf{G}^k(\vec{r})_i = G(k; \vec{r}_i - \vec{r}). \quad (13)$$

The computation of the solution, Eq. (11), can easily be done numerically, typically by using a Gauss elimination scheme

to avoid time-consuming matrix inversions, without further simplification. The caveat being the use of significant computational resources as the number of scatterers increases. In principle, several hundred scatterers can readily be included on a basic PC platform, at least for the computation of the field at one point and for one frequency only. More resources are necessary if the total cross section as a function of frequency is wanted.

The results of these computations can then be averaged over n_{av} random realizations of the assembly of scatterers to obtain the usual multiple scattering ensemble averages^{37,38}

$$\langle \rangle = \frac{1}{n_{av}} \sum_{i=1}^{n_{av}} . \quad (14)$$

The coherent (I_C^s) and diffusive (I_D^s) portions of the observed scattered intensity I^s and of the cross sections are given in terms of the averages $\langle p_s(\vec{r}) \rangle$ and $\langle |p_s(\vec{r})|^2 \rangle$ of the scattered acoustic field and of its square modulus:

$$\begin{aligned} I^s &= \langle |p_s(\vec{r})|^2 \rangle, \\ I_C^s &= |\langle p_s(\vec{r}) \rangle|^2, \\ I_D^s &= I^s - I_C^s = \langle |p_s(\vec{r}) - \langle p_s(\vec{r}) \rangle|^2 \rangle. \end{aligned} \quad (15)$$

For the computations presented in this work, the relative accuracy in the averaged quantity is chosen to be 2% or better. This typically requires the computation of the scattered field according to Eq. (11) for $n_{av} \approx 300-3000$ randomly generated configurations of N scatterers. The result is the field at one point in space for one particular wave number k . To compute the dependence of the quantities of interest on k , or frequency $f = kc/(2\pi)$, the outlined procedure is followed stepping through the entire range of consideration. Scattering cross sections are computed by looking at a receiver location in the far field, that is, at a distance from the assembly at which the radial part of the scattered intensities well reached their asymptotic $1/r^2$ behavior. Typically, the distance from the center of the assembly is chosen to be at least 50 times its radius.

Alternatively, Eq. (10)—which is an implicit equation for the “external field on scatterer i ”-vector $\mathbf{p}_{es}^i = p^i(\vec{r}_i)$ —can be solved by Neumann iteration

$$\begin{aligned} \mathbf{p}_{es} &= \mathbf{p}_0 + f_s \mathbf{G}^k \mathbf{p}_{es} \\ &= \mathbf{p}_0 + f_s \mathbf{G}^k \mathbf{p}_0 + f_s \mathbf{G}^k f_s \mathbf{G}^k \mathbf{p}_0 \\ &\quad + f_s \mathbf{G}^k f_s \mathbf{G}^k f_s \mathbf{G}^k \mathbf{p}_0 + \dots \\ &= \mathbf{p}_0 + \sum_{n=1}^{\infty} \underbrace{f_s^n \mathbf{G}^k \cdot \mathbf{G}^k \cdot \dots \cdot \mathbf{G}^k}_{n \text{ times}} \cdot \mathbf{p}_0 \end{aligned} \quad (16)$$

the scattered field can then be evaluated from Eq. (8)

$$p_s(\vec{r}) = f_s \mathbf{p}_{es}^T \cdot \mathbf{G}^k(\vec{r}). \quad (17)$$

Each term in the series represents the entire path from an arbitrary initial scatterer to the scatterer at the considered locus, going through $n-1$ other scatterers. In this sense, higher order terms in the Neumann series contain the effect of multiple scattering. However, there is a caveat: this Neu-

mann series does not converge unconditionally, the spectral radius of the matrix $f_s \mathbf{G}^k$ must be smaller than unity. Using Gerschgorin’s circle theorem⁴⁶ to estimate the spectral radius, we arrive at the convergence criterion

$$\frac{3}{2} \frac{|f_s| N}{R} = 2\pi n R^2 |f_s| < 1. \quad (18)$$

R indicates the size of the volume occupied by the scatterers (here assumed to be of spherical shape) and n the density of scatterers. This hints that for sufficiently dense scatterers and sufficiently high scattering strength, the series (16) might not converge. Additionally, as the volume increases, the Neumann series is stable only for smaller and smaller densities. To remedy this situation, the multiple scattering series needs to be rearranged by directly summing subclasses containing an infinite number of terms. Ultimately, this leads, in the case of an infinite medium, to a renormalization of the effective wave number, which is the foundation of the effective medium theory.³⁷

C. Cross sections from effective medium theory

In his pioneering work on multiple scattering of waves, Foldy³⁸ demonstrated that the coherent part of the scalar field $p_{coh} \equiv \langle p \rangle$ scattered by an ensemble of scatterers is given by the solution of the Helmholtz equation:

$$(\nabla^2 + k_e^2) p_{coh} = 0, \quad (19)$$

where the effective wave number k_e is, in general, a complex valued function of position and is for a random ensemble of identical isotropic point scatterers “to lowest order” given by

$$k_e^2 = k^2 + 4\pi n f_s. \quad (20)$$

Here, f_s is the scattering strengths of an individual scatterer as given in Eq. (1) and n is the volumetric density of the scatterers in the region occupied by them, which can be a function of position. In diagrammatic language, to lowest order means the neglect of backscattering between individual scatterers, that is, the involvement of scatterers more than once in a multiple scattering chain. This leads to a good approximation, if³⁷

$$\left| \frac{4\pi n f_s^2}{k} \right| \ll 1. \quad (21)$$

The generalization to the non-isotropic case and to non-identical scatterers with a given statistical distribution of defining parameters is straightforward and will not be considered here. Together with given boundary conditions at the interface between the region occupied by the scatterers and the free medium outside, and also assuming radiation condition at infinity, Eqs. (19) and (20) define a closed mathematical model for the coherent part of the scattered field, which is valid as long as Eq. (21) holds.

In this approximation, the coherent scattered pressure field due to a harmonic plane wave incident on a spherical assembly of bubbles can be obtained by considering the non-rigid sphere scattering problem.⁴⁷ Consider a sphere of radius R at the origin of a spherical coordinate system. An inbound plane wave of frequency ω traveling along the posi-

tive z -axis is scattered by a sphere with bulk properties ρ_e and c_e . The surrounding medium is void of scatterers and has the free bulk parameters ρ and c . At the boundary the pressure and the radial component of the velocity are continuous and the radiation condition holds at infinity to ensure that scattered energy is only outgoing. Outside the sphere, the spatial part of sound pressure field, omitting the time-dependence $e^{-i\omega t}$, is given by a spherical wave expansion of the scattered field

$$\begin{aligned}
 p &= p_0 + p_s \\
 &= \sum_{n=0}^{\infty} (2n+1) i^n P_n(\cos \vartheta) j_n(kr) \\
 &\quad + \sum_{n=0}^{\infty} \sum_{m=0}^n A_{nm} Y_n^m(\cos \vartheta) h_n^{(1)}(kr). \quad (22)
 \end{aligned}$$

The first part represents the spherical wave expansion of the incoming plane wave $p_0 = e^{ikr \cos \vartheta}$, $h_n^{(1)}$ and j_n are the spherical Hankel and Bessel functions of the first kind, and Y_n^m and P_n denote the spherical harmonics and the Legendre polynomials, respectively. In all these equations the point of observation \vec{r} is given in spherical polar coordinates, r denotes the radial distance from the center of the sphere and ϑ the polar angle measured from the positive z -axis.

Similarly, the field inside the sphere can be represented by an equivalent spherical wave expansion:

$$p_e = \sum_{n=0}^{\infty} \sum_{m=0}^n B_{nm} Y_n^m(\cos \vartheta) j_n(k_e r). \quad (23)$$

The coefficients A_{nm} and B_{nm} are readily found from the continuity conditions at the interface

$$\begin{aligned}
 p &= p_e \Big|_{r=R}, \\
 \frac{1}{i\omega\rho} \partial_r p &= \frac{1}{i\omega\rho_e} \partial_r p_e \Big|_{r=R}
 \end{aligned} \quad (24)$$

to yield

$$\begin{aligned}
 A_{n0} &= -i^n \sqrt{4\pi(2n+1)} \frac{j_n'(kR) + i\beta_n j_n(kR)}{h_n^{(1)'}(kR) + i\beta_n h_n^{(1)}(kR)}, \\
 B_{n0} &= i^n \sqrt{4\pi(2n+1)} \frac{1}{j_n(k_e R)} \\
 &\quad \times \frac{j_n(kR) h_n^{(1)'}(kR) - j_n'(kR) h_n^{(1)}(kR)}{h_n^{(1)'}(kR) + i\beta_n h_n^{(1)}(kR)}. \quad (25)
 \end{aligned}$$

As the problem exhibits rotational symmetry about the z -axis, all coefficients are zero for $m > 0$. The parameters $\beta_n = i(\rho c / \rho_e c_e) j_n'(k_e R) / j_n(k_e R)$ describe the impedance contrast at the surface of the sphere.

The expressions Eq. (25) give the desired result. From this we obtain the coherent part of the scattered field and, consequently, the expansion of the scattering amplitude F_s of the assembly in terms of the partial waves amplitudes $a_n(k)$

$$F_s(k, \vartheta) = \sum_n (2n+1) a_n(k) P_n(\cos \vartheta),$$

$$a_n(k) = \frac{1}{2ik} (S_n(k) - 1). \quad (26)$$

According to the above definition, the S -matrix $S_n(k)$ follows directly, the star denotes complex conjugation:

$$S_n(k) = -\frac{h_n^{(1)*'}(kR) + i\beta_n h_n^{(1)*}(kR)}{h_n^{(1)'}(kR) + i\beta_n h_n^{(1)}(kR)} = e^{2i\delta_n(k)}, \quad (27)$$

and is for real valued k and k_e (no energy dissipation) unitary, which incidentally also leads to the above definition of the scattering phase shifts $\delta_n(k)$. Following from this result, the zero-order S -matrix, $S_0(k)$, takes the particularly simple form

$$\begin{aligned}
 S_0(k) &= -e^{-2ikR} \frac{kR \sin(k_e R) - ik_e R \cos(k_e R)}{kR \sin(k_e R) + ik_e R \cos(k_e R)} = e^{2i\delta_0(k)}, \\
 \delta_0(k) &= \frac{1}{2} \arctan \frac{kk_e \cos(2k_e R)}{k_e^2 \cos^2(k_e R) - k^2 \sin^2(k_e R)} - kR. \quad (28)
 \end{aligned}$$

These expressions finally yield the total scattering cross section σ_s of the spherical assembly of suspended scatterers:

$$\begin{aligned}
 \sigma_s &= 4\pi \sum_n (2n+1) |a_n(k)|^2 \\
 &= \frac{4\pi}{k^2} \sum_n (2n+1) \left| \frac{j_n'(kR) + i\beta_n j_n(kR)}{h_n^{(1)'}(kR) + i\beta_n h_n^{(1)}(kR)} \right|^2. \quad (29)
 \end{aligned}$$

Both quantities, the scattering amplitude and the scattering cross section, can readily be evaluated numerically; at low frequencies only the first few terms of the sum contribute significantly. Generally, the sum has to be carried to $n = O(kR)$.

D. Resonance contributions

It is evident from these expressions that the scattering amplitude and, hence, the cross section show resonances at those particular values of $k = \omega/c$ for which the denominators in the quantities above become zero, or:

$$(kR) h_n^{(1)'}(kR) j_n(k_e R) - (k_e R) h_n^{(1)}(kR) j_n'(k_e R) = 0. \quad (30)$$

Here, density differences between the effective medium and the scatterer-free surrounding space have been ignored. This holds very closely for many applications such as oceanic bubble clouds and trivially for fish schools that are in general neutrally buoyant. The effective wave number k_e is a function of k and given by Eq. (20).

Generally, Eq. (30) has no solutions for real values of k . However, solutions can be found for complex values $k^p = k_r^p + ik_i^p$, which correspond to the poles of the appropriately analytically continued S -matrix. From the location of these poles, the resonance contributions to the cross section can be obtained. If the poles are close to the real axis, the physical domain of the wave number, sharp resonances will be observed. Poles that are remote from the real axis will only contribute weak and wide resonances, which in some cases might not be resolved individually at all. In this case the resonances are not explicitly manifest in the spectrum.

Often, collective modes occur at frequencies well below the resonance frequency of the individual monopole scatterers. In this regime, $\omega \ll \omega_0$, we can for the time being ignore damping terms and, from Eqs. (1) and (3), Eq. (20) can be simplified to

$$k_e = \kappa k, \quad (31)$$

$$\kappa = \sqrt{1 + \frac{\beta \rho c^2}{\gamma_{\text{eff}} P_0}},$$

where we introduced the void fraction $\beta = n4\pi a^3/3$.

This particular form of k_e in terms of a real-valued κ allows analytical treatment. Defining $z_{(e)} = k_{(e)}R$, the condition for the low frequency resonances of the lowest mode, $n=0$, can from Eq. (30) and the definition of the zero-order spherical Bessel functions be written as

$$z \sin z_e + i z_e \cos z_e = 0, \quad (32)$$

which can also be directly obtained from $S_0(k)$ in Eq. (28). If real and imaginary parts are separated, the previous condition is equivalent to

$$(z_r^2 + z_i^2) \sin[z_r \kappa] (\cosh[z_i \kappa] + \kappa \sinh[z_i \kappa]) = 0, \quad (33)$$

$$(z_r^2 + z_i^2) \cos[z_r \kappa] (\sinh[z_i \kappa] + \kappa \cosh[z_i \kappa]) = 0.$$

The real valued solutions z_r^p and z_i^p of this equation for $\kappa > 0$ are found to be

$$z_r^p = \frac{1}{2} \frac{(2m+1)\pi}{\kappa}, \quad (34)$$

$$z_i^p = -\frac{\tanh^{-1}[\kappa^{-1}]}{\kappa},$$

for any integer value m . As expected, the distance of the pole from the real axis increases with diminishing impedance contrast and vice versa (the inverse hyperbolic tangents is a monotonically increasing function that goes to infinity as the argument goes to one). This effect is shown in Fig. 2 for an example of a spherical fish school. As the populations become denser, the poles move closer to the real axis and low frequency *collective* modes appear as seen by the variation of the shades along the real axis.

From the locus (z_r^p, z_i^p) of the low frequency resonance-poles of the S -matrix in the complex z -plane, the center frequency and width of the scattering cross section can be calculated. Off the real axis the appropriate analytical continuation of $S_n(z)$ is given by:

$$S_n(z) = -\frac{h_n^{(1)*}(z^*) + i\bar{\beta}_n(z_e^*)h_n^{(1)*}(z^*)}{h_n^{(1)}(z) + i\beta_n(z_e)h_n^{(1)}(z)}, \quad (35)$$

$$\bar{\beta}_n = i\frac{c}{c_e} \frac{j_n^*(z_e^*)}{j_n^*(z_e^*)}.$$

This, for real k_e , reproduces the derived result Eq. (27), and as $h_n^{(1)*}(z^*) = (-1)^n h_n^{(1)}(-z)$ and $h_n^{(1)*}(z^*) = (-1)^{n+1} h_n^{(1)}(-z)$ (the same relations hold for $j_n^*(z_e^*)$ and $j_n^*(z_e^*)$), the S -matrix can be rewritten in terms of the Jost-function $f_n(z)$

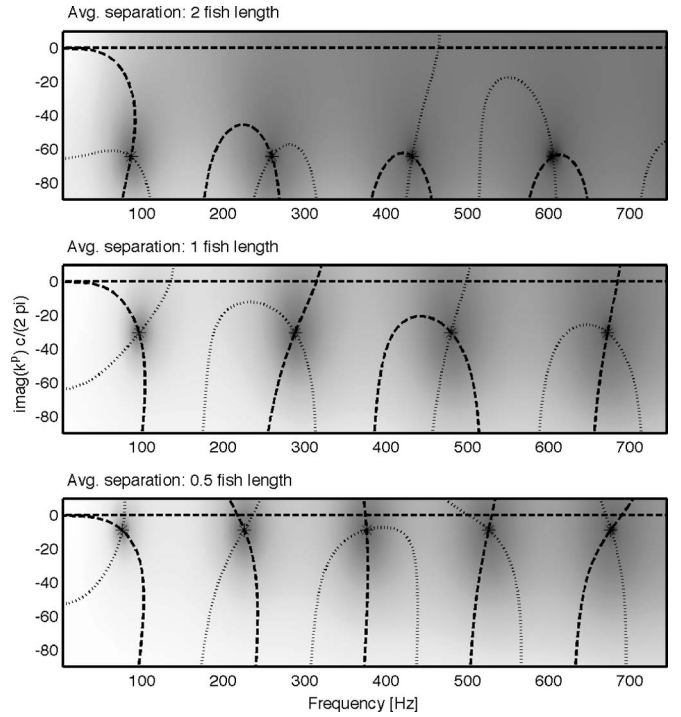


FIG. 2. The contour lines of $zh_n^{(1)}(z)j_n(z_e) - z_e h_n^{(1)}(z)j_n'(z_e) = 0$ (real part dashed line, imaginary part dotted line) to visualize the low frequency resonance poles of the S -matrix Eq. (35). The gray shades in the background show $|1/(zh_n^{(1)}(z)j_n(z_e) - z_e h_n^{(1)}(z)j_n'(z_e))|$, the dark region indicating high values around the poles. The exemplary calculation is based on a school of 3000 fish; each 20 cm long, located at a depth of 2 m, and on various average fish separations (from top: 2, 1, and 0.5 times the fish length). The star symbols denote the location of the resonance poles as calculated directly from Eq. (34).

$$S_n(z) = (-1)^n \frac{f_n(z)}{f_n(-z)},$$

$$f_n(z) = h_n^{(1)}(-z) + i\beta_n(-z_e)h_n^{(1)}(-z). \quad (36)$$

To isolate the contribution of a single resonance to the scattering cross section, we write for $z \approx z^p$, $f_n(-z) \propto (z - z^p)$. As $f_n(z) = f_n^*(-z^*)$, we obtain

$$S_n(z) = \frac{z - z^p}{z - z^p} \bar{S}_n(z). \quad (37)$$

$\bar{S}_n(z)$ describes the slowly varying background contribution to $S_n(z)$. The resonance contributions to the partial waves amplitude $a_n(k)$ naturally follow the Breit-Wigner resonance formula

$$a_{0m}(k) = \frac{1}{k} \frac{\Gamma_{0m}/2}{(\hat{k}_{0m}^2 - k^2) - i\Gamma_{0m}/2}, \quad (38)$$

$$\hat{k}_{0m}^2 R^2 = z_r^2 - z_i^2 = \frac{1}{\kappa^2} ((2m+1)^2 \pi^2 - 4 \tanh^{-1}[\kappa^{-1}]^2),$$

$$\Gamma_{0m} R^2 / 2 = -2z_r^p z_i^p = \frac{(2m+1)\pi}{\kappa^2} \tanh^{-1}[\kappa^{-1}],$$

which leads to the resonance contribution to the assembly scattering cross section

$$\begin{aligned}\sigma_{nm}^{\text{res}} &\simeq \frac{4\pi(2n+1)}{k^2} \frac{(\Gamma_{nm}/2)^2}{(\hat{k}_{nm}^2 - k^2)^2 + (\Gamma_{nm}/2)^2} \\ &= \frac{4\pi(2n+1)\hat{R}_{nm}(k)^2}{(\hat{k}_{nm}^2/k^2 - 1)^2 + (k\hat{R}_{nm}(k))^2}.\end{aligned}\quad (39)$$

The first index gives the order of the multipole; the second counts the poles in one term of the multipole expansion. The parameters $\Gamma_{0m}/2$ and \hat{k}_{0m} determine the width and the center of the (monopole) collective resonances. In the second line above, the resonance cross section is expressed in terms of the length scale $\hat{R}_{nm}(k)$, which is defined in terms of the resonance width: $\hat{R}_{nm} = (\Gamma_{nm}/2)/k^3$. At the lowest resonance, at which $k = \hat{k}_{00}$, $\hat{R} \equiv \hat{R}_{00}(\hat{k}_{00})$ is of order of the radius R . This is apparent from Eq. (38) as $\hat{k}_{00}R = O(1)$.

From now on we will focus on the lowest collective mode, $m=0$, which for practical applications is the single most important one. To simplify the notation, indices are occasionally dropped if no confusion can arise and all quantities are meant to describe the lowest mode $n=0$ and $m=0$. The collective resonance frequency, ignoring viscous and thermal damping, which would cause z_r^p and z_i^p to pick up a small imaginary part, is from Eq. (38) given by

$$\begin{aligned}\hat{\omega}_0 &= \frac{c}{2\kappa R} \sqrt{\pi^2 - 4 \tanh^{-1}[\kappa^{-1}]^2} \\ &= \frac{c}{2R} \left(1 + \frac{\beta \rho c^2}{x^2 \gamma P}\right)^{-1/2} \\ &\quad \times \sqrt{\pi^2 - 4 \tanh^{-1} \left[\left(1 + \frac{\beta \rho c^2}{x^2 \gamma P}\right)^{-1/2} \right]^2}.\end{aligned}\quad (40)$$

To model the resonance cross section correctly, the non-resonance contribution, Eq. (39), has to be included as well. From Eq. (37) we recognize that

$$\begin{aligned}\bar{S}_0(z) &= \frac{z - z^p}{z - z^{p*}} S_0(z) = e^{2i(\delta_0(k) - \delta_{r0}(k))}, \\ e^{2i\delta_{r0}(k)} &= \frac{z - z^{p*}}{z - z^p},\end{aligned}\quad (41)$$

where $\delta_{r0}(k)$ is the zero-order resonance phase-shift. The pole contribution to $S_0(z)$ is by inspection also unitary, as it must be because of energy conservation. Using the explicit expression Eq. (34) for the imaginary and real parts of the poles and Eq. (28) for the zero-order S -matrix (or phase shifts), the background part $\bar{S}_0(z)$ can be calculated directly. Including this background, Eq. (37), the scattering cross section for a single resonance becomes

$$\sigma^{\text{res}}(k) = \frac{|\bar{S}_0(\hat{k}_0) + 1|^2}{4} \frac{4\pi\hat{R}(k)^2}{(\hat{k}_0^2/k^2 - 1) + (k\hat{R}(k))^2}.\quad (42)$$

In the previous expression, the background is assumed to be slowly varying and taken constant across the resonance. As $|\bar{S}_0(\hat{k}_0 + 1)|^2 = 2(1 + \cos(2\delta_0(k) - 2\delta_{r0}(k)))$ is between 0 and 4, the background factor in Eq. (42) remains always smaller

than one. If the resonance is wide, i.e., the imaginary part of the resonance poles is large, the background cannot be assumed to be constant over the entire resonance. In this case, the Breit–Wigner resonance shape will not only be scaled but also distorted. The full result equivalent to the previous one is

$$\begin{aligned}\sigma^{\text{res}}(k) &= \frac{|\bar{S}_0(k) + 1|^2}{4} \frac{4\pi\hat{R}(k)^2}{(\hat{k}_0^2/k^2 - 1)^2 + (k\hat{R}(k))^2} + \sigma_{bg}^{\text{nonres}}(k), \\ \sigma_{bg}^{\text{nonres}}(k) &= \frac{\pi(\hat{k}_0^2/k^2 - 1)(\hat{k}_0^2/k^2 - 1)|\bar{S}_0(k) - 1|^2 + 4k\hat{R}(k)\text{Im}\bar{S}_0(k)}{k^2(\hat{k}_0^2/k^2 - 1)^2 + (k\hat{R}(k))^2}.\end{aligned}\quad (43)$$

Eq. (39) has an interesting interpretation. Its analogy to Eq. (6) suggests that at frequencies close to the collective resonance, the entire cloud of monopole scatterers behaves like a single bubble of radius \hat{R} . Using this analogy, damping, which has been ignored in Eqs. (38), can be incorporated directly into (39)

$$\begin{aligned}\sigma^{\text{res}}(k) &= \frac{|\bar{S}_0(\hat{k}_0) + 1|^2}{4} \frac{4\pi\hat{R}(k)^2}{(\hat{k}_0^2/k^2 - 1)^2 + (k\hat{R}(k) + \bar{\delta})^2} \\ &\quad + \sigma_{bg}^{\text{nonres}}(k)\end{aligned}\quad (44)$$

by means of a simple physical argument: the value of the effective damping constant $\bar{\delta}$ of the entire assembly in Eq. (44) is to be adjusted such that the total energy loss per unit time is equal to the sum of the viscous and thermal energy losses of all the N individual scatterers separately. At frequencies close to $\hat{\omega}_0$, this can from Eq. (7) be written in the form

$$P_\delta = 4\pi a^2 I_{\text{av}} \frac{N(\delta_{\text{visc}} + \delta_{\text{th}})}{\hat{k}_0 a} \left(\frac{\hat{\omega}_0}{\omega_0}\right)^4.\quad (45)$$

Equation (45) needs some explanation. First, we are considering frequencies $\hat{\omega}_0 \ll \omega_0$ allowing us to approximate the single bladder scattering cross section σ_s^b as $4\pi a^2(\hat{\omega}_0/\omega_0)^4$. Second, due to the multiple scattering between the swim bladders, the field inside the assembly is significantly enhanced at frequencies close to the resonance $\hat{\omega}_0$. This is seen from Eqs. (23) and (25). As a result, the bubbles are not exposed to the incoming sound field but, instead, to the enhanced field inside the school p_e . To simplify matters, we average the corresponding intensity over the entire region occupied by the bubbles to get

$$\begin{aligned}I_{\text{av}} &= \frac{3}{4\pi R^3} 2\pi \int_0^R \int_0^\pi \sin \vartheta d\vartheta r^2 dr |p_e(r)|^2 \\ &= \frac{3}{16\pi(Rk_e)^3} |B_{00}|^2 (2Rk_e - \sin(2Rk_e)).\end{aligned}\quad (46)$$

The absorbed power, Eq. (45), can, on the other side, also be calculated for the equivalent school sized bubble of

radius \hat{R} that is exposed directly to the incoming sound field and driven at its resonance $\hat{\omega}_0$. This idea leads to

$$P_\delta = \frac{|\hat{S}_0(\hat{k}_0) + 1|^2}{4} \frac{4\pi\hat{R}^2}{(\hat{k}_0\hat{R} + \hat{\delta})^2} \frac{\hat{\delta}}{(\hat{k}_0\hat{R})}. \quad (47)$$

To get the equivalent damping constant $\hat{\delta}$, we equate Eqs. (45) and (47) describing the same physical quantities from two points of view and obtain:

$$\hat{\delta} = \frac{4}{|\hat{S}_0(\hat{k}_0) + 1|^2} NI_{av}(\delta_{visc} + \delta_{th})(\hat{k}_0\hat{R})^2 \frac{a}{\hat{R}} \left(\frac{\hat{\omega}_0}{\omega_0}\right)^4 \eta_{corr},$$

$$\eta_{corr} \cong 1 + 2NI_{av}(\delta_{visc} + \delta_{th})(\hat{k}_0 a) \left(\frac{\hat{\omega}_0}{\omega_0}\right)^4. \quad (48)$$

For some applications, either δ_{visc} or δ_{th} is dominant at the frequency considered, allowing further simplification.

The presented closed-form analytical model, Eq. (44), of the lowest mode collective oscillations allows, at least in the simple case of a spherical assembly with no positional correlations between scatterer loci and point-like individual scatterers with known scattering amplitudes, an inversion for the total number of scatterers. In the next chapter, we will apply this technique to the specific case of an idealized spherical school of fish.

III. SAMPLE APPLICATION: AN IDEALIZED SPHERICAL FISH SCHOOL

A. Bioacoustics of individual fish

The scattering of sound waves at frequencies below 25 kHz from an individual pelagic fish is largely due to the resonant response of its swim bladder to the insonifying wave. Even at higher frequencies, the swim bladder contributes significantly to the scattering strength. The acoustic response peaks at the resonance frequency of the oscillating swim bladder which can, in close analogy to spherical air bubbles, Eq. (3), be approximated by^{43,48}

$$\omega_0 \approx \frac{x(\varepsilon)}{a} \sqrt{\frac{3\gamma_{eff}P_0}{\rho}} = x(\varepsilon)\omega_M. \quad (49)$$

As fish bladders are usually quite elongated, at least two parameters— a , the equivalent equilibrium bladder radius, which in terms of the swim bladder volume V_b is given by

$$a = (3V_b/4\pi)^{1/3}, \quad (50)$$

and the bladder eccentricity parameter ε —are needed to describe its geometry. This formula carries the same approximations as Minnaert's result for air bubbles and neglects the effects of bladder tension as well as the frequency dependence of γ_{eff} . However, for typical equivalent swim bladder radii, $a > 3$ mm, driven at frequencies greater than 300 Hz changes of state in the bladder gas will, except for a small thermal boundary layer, be adiabatic⁴⁹ and the well known value for diatomic gases $\gamma_{eff} \approx \gamma \approx 1.4$ can be chosen. With this simplifying assumption, Minnaert's result Eq. (49) deviates by only less than two percent from the exact result⁴⁰ for the considered bladder sizes.

Because of its elongated shape, the resonance frequency of the bladder will be somewhat higher than that of an equivalent spherical bubble. The resonance frequency increase for the prolate spheroid calculates, following Strassberg,⁵⁰ as

$$x(\varepsilon) = \varepsilon^{-1/3}(1 - \varepsilon^2)^{1/4} \log^{-1/2}\left(\frac{1 + \sqrt{1 - \varepsilon^2}}{\varepsilon}\right). \quad (51)$$

In Eq. (51), the eccentricity parameter ε is defined as the ratio of the minor to the major semi-axes.

In the case of physostomes, such as sardines and anchovies, gas in the swim bladder cannot be exchanged quickly. In this case, the volume of the bladder varies, for an isothermal depth profile, inversely to the hydrostatic pressure. Hence, the resonance frequency, Eq. (49), can be approximately evaluated from the depth h , the sea-level pressure P_0 and the bladder radius a_0 , which is usually the more directly obtained experimental quantity:¹³

$$\omega_0 = \frac{x(\varepsilon)}{a_0} \underbrace{\left(1 + \frac{\rho gh}{P_0}\right)^{5/6}}_{\approx 1+0.1h} \sqrt{\frac{3\gamma P_0}{\rho}}. \quad (52)$$

The bladder size and the bladder eccentricity increase with increasing fish length l_f . A comprehensive summary of data^{11,51,52} from sardines, coalfish and herring can be found in a series of papers by Diachok.^{13,14} The computations presented below are based on a linear dependence of a_0 and $1/\varepsilon$ on l_f obtained from a linear least square data analysis. Based on the data summarized by Diachok one arrives at: $a_0(\text{cm}) \approx 0.058l_f(\text{cm}) - 0.14$ for the equivalent bladder radius and $1/\varepsilon \approx 0.29l_f(\text{cm}) + 7.0$ for the bladder eccentricity parameter, valid for l_f between 5 and 20 cm.

Other than the increase in resonance frequency due to the bladder elongation, the bladder contribution to the scattering characteristics of an individual fish is assumed to be equivalent to air bubbles. Nevertheless, some numeric parameters need to be adapted. Viscous damping in fish is mostly argued to be due to the fish flesh surrounding the bladder. The associated viscosity is generally assumed to be about^{10,43} 50 N s/m², which is the value chosen here, unless explicitly stated otherwise. For fish from 0.1 to 0.5 m in length, a measures roughly 0.4 to 2 cm. In this regime $\text{Im } \phi$ remains below about 0.03 resulting in $\delta_{th} \ll \delta_{visc}$. Therefore, we will entirely ignore thermal damping in the following examples.

B. Analysis and results

Before we explore the application of the general analysis presented in Part II, we should verify that the underlying approximations are valid for the case of fish schools. For the effective medium theory to hold in the "lowest order" version exploited here, the condition (21) needs to be satisfied. It turns out that for actual fish schools, this condition is valid for a wide range of parameters, even for very dense schools. Figure 3 shows the actual values of the left hand side of Eq. (21) for the type of fish schools considered below. For fish of 20 cm length, spaced more than one fish length apart, the

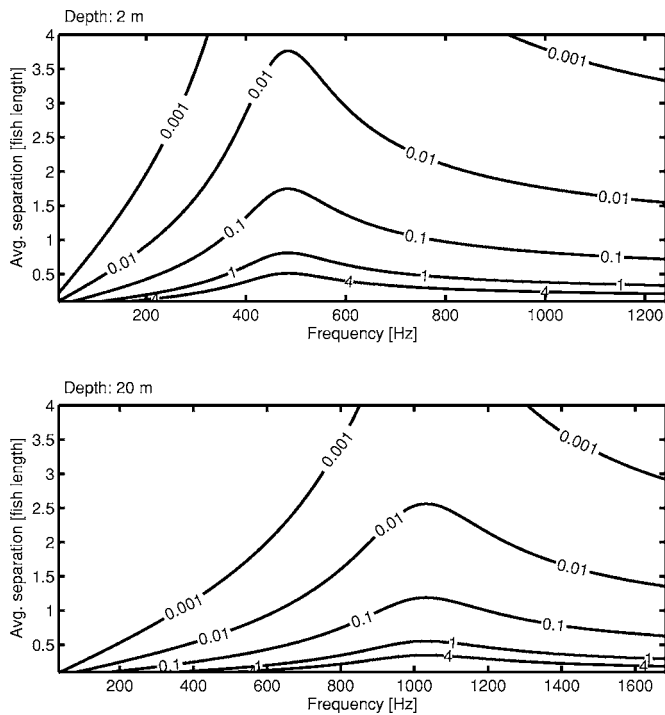


FIG. 3. The value of the left hand side of the validity condition, Eq. (21), as a function of fish separation and frequency for fish of length 0.2 cm at depths of (top) 2 m and (bottom) 20 m.

numerical value of the left hand side is well below one, *even across resonance* and close to the surface, where the scattering strength f_s is largest. This somewhat contrasts the case of mono-disperse clouds of pure air bubbles³⁴ and is the result of the larger damping in fish bladders.

Because of the validity of relation (21), we expect the effective medium approach, which we based on the Helmholtz equation (19) and the effective wave number given by Eq. (20), to compare favorably with the direct numerical benchmark calculation of the ensemble averaged pressure field from Eq. (11). All effective medium computations that are shown below are based on a straightforward evaluation of the spherical wave expansion of the total scattering cross section, Eq. (29), and the corresponding expansion for the scattering amplitude, Eqs. (26) and (27). The series is terminated once a satisfactory precision, typically an increase in the partial sums below $1E-10$, has been achieved. The number of terms to be considered depends essentially on the value of kR and can be large for bigger fish schools.

Figure 4 shows such a comparison for a relatively small and tightly spaced fish school containing 66 fish of 20 cm length at a density of about 300 fish/m³. For equally spaced fish in a cubic arrangement, this would hypothetically result in a fish separation of 0.75 times the fish length along the edge of a cubic lattice. This is what we mean by an “average spacing of 0.75 times the fish length apart.” Such a description of the density in terms of a fish separation is chosen because it is much more intuitive than stating the numeric fish density and is used in all demonstration cases given below. The fish school is assumed to reside close to the surface, nominally at a depth of 2 m. For simplicity, this depth is assumed to be the same for all individual fish. It is seen

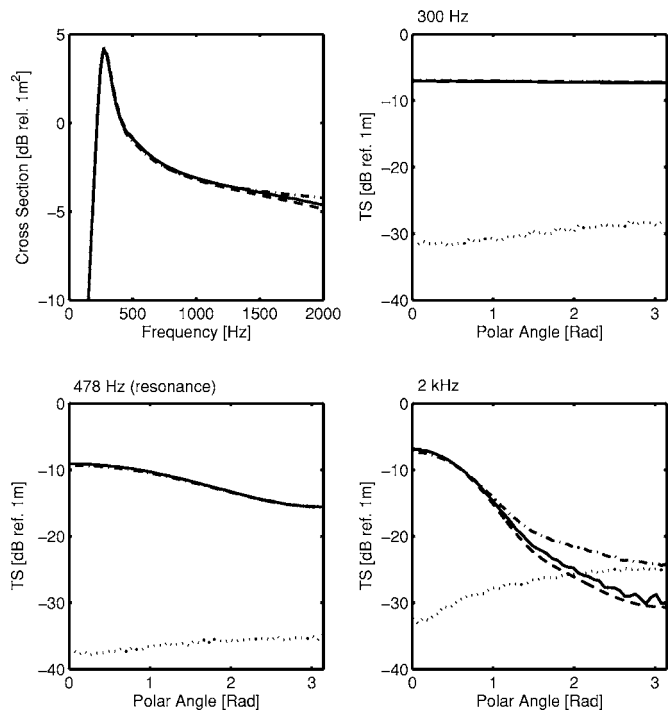


FIG. 4. Comparison between the numerical computation of the coherent field following Eq. (11), solid lines, and the effective medium calculation from Eqs. (29) and (27), dashed lines. Additionally, the dash-dotted lines indicate a numerical computation for a fish school with positional correlations based on a hard sphere model with a hard sphere radius given by one quarter of the fish length. The fish school contains 66 fish of length 0.2 m at 2 m depth at an average spacing of 0.75 times the fish length. The diffusive part of the field leads to the quantities indicated by the dotted line. The TS is indicated for frequencies of 300 Hz, 478 Hz (bladder resonance), and 2 kHz.

that, for this school, the total scattering cross section is almost entirely coherent and virtually no deviation can be found in this quantity between the effective medium model and the numerical benchmark. The TS (or the differential cross section) at low frequencies is uniform across the entire range of polar angles and peaks at higher frequencies in the forward direction, as expected. It is noteworthy that the effective medium model performs exceptionally well even across the resonance of the individual fish. In the regime where the diffusive part starts playing the dominant role, that is, at higher frequency back scattering, deviations are visible but small (1–2 dB).

If we require the distance between neighboring fish to be larger than two thirds of the average spacing, without otherwise changing the random character of the fish configurations, we get a coherent cross section that is given by the dash-dotted line in Fig. 4. Technically, this is correctly done by randomly choosing the locus of each fish and rejecting configurations that do not obey the hard sphere condition as described above. Note that for dense packing this is a computationally intense task. Because of the tight spacing and a resulting correlation length that is small compared to the wavelength, the effect is minor. Only at backscattering configurations and higher scattering angles do significant deviations from the purely random case occur. The scattering cross section, which at higher frequencies is dominated by forward scattering, is only minimally altered.

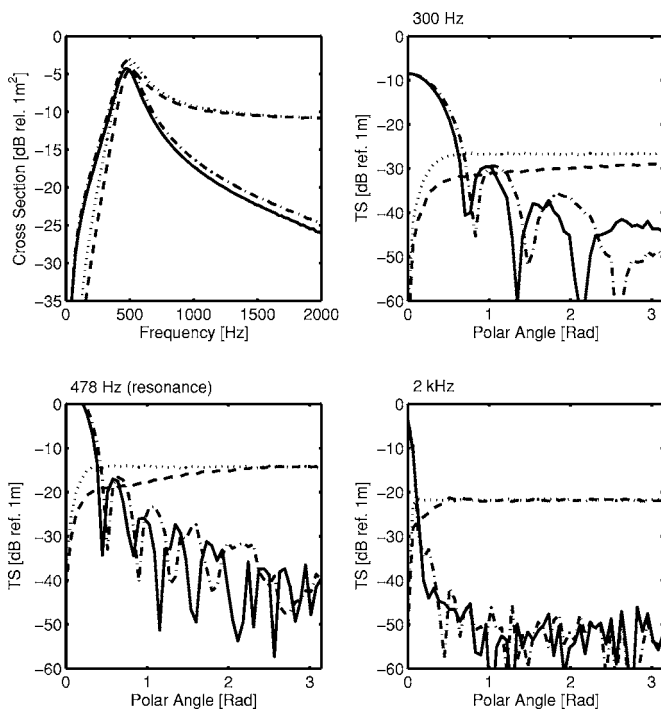


FIG. 5. Comparison of the cross section and TS between fish schools with no positional correlations between the individual fish loci and a distribution of fish exhibiting positional short-range correlations based on a hard sphere assumption. The school contains 66 fish of length 0.2 m at 2 m depth at an average spacing of 10 times the fish length. The hard sphere radius is chosen to be 3 times the fish length. The results for the uncorrelated case are indicated by the solid (coherent part) and the dotted (diffusive) lines; the computations for the hard sphere case are given by the dash-dotted (coherent) and the dashed (diffusive) lines. The TS is indicated for frequencies of 300 Hz, 478 Hz (bladder resonance), and 2 kHz.

Another scenario demonstrating the effect of short-range correlations is shown in Fig. 5. The considered school also contains 66 fish at a depth of 2 m, but at a much wider spacing. Here we assume that the fish avoid getting closer than six times the fish length to their nearest neighbors. This leads to a correlation length that is, in the frequency range of interest, of the order of half the wavelength. Again, little deviation is found in the total cross section as compared to the uncorrelated configurations. The TS, on the other side, starts to show significant deviations for configurations at which the diffusive part plays the dominant role. But in the forward direction, up to about 60° scattering angle, the deviations are still reasonably small and our simplified analytical model is applicable to these situations.

The “test” fish school is chosen to be dense and close to the surface. In this most interesting scenario, the “expansion parameter” [left-hand side of Eq. (21)] is of order one and multiple scattering contributions start playing a critical role. In consequence, this example of a relatively small fish school can serve as a powerful benchmark to demonstrate the utility of the effective medium model. There is no reason to doubt that the model performs just as remarkably for larger schools because increasing the fish number does not alter the validity of Eq. (22).

It is interesting to note that, due to the effects of multiple scattering, the resonance peak in the total scattering cross section is significantly shifted downwards relative to the in-

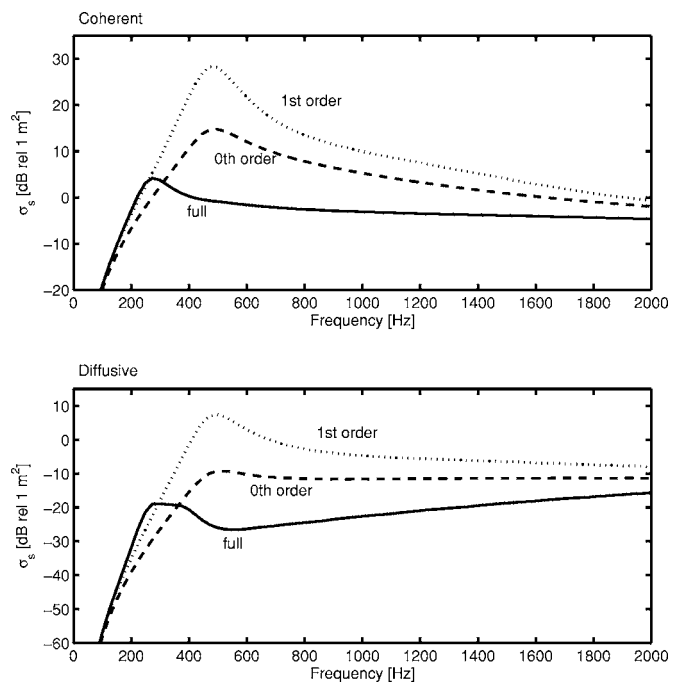


FIG. 6. Comparison between the numerical calculation of the (top) coherent and (bottom) diffusive part of the total scattering cross section for a fish school containing 66 fish of 0.2 m length at a depth of 2 m, spaced in average of 0.75 times the fish length and the Neumann series terminated at the zeroth (dashed line) and the first (dotted line) order.

dividual fish resonance. Clearly, this cannot be explained by simply adding the scattering contributions of all the individual fish (phase adjusted to the position). A comparison between the full result and the first term in the Neumann series Eq. (16) is shown in Fig. 6. Another effect of multiple scattering becomes evident: at the resonance of the individual fish the cross section is diminished. In contrast, the coherent scattering returns are increased at lower frequencies. We intuitively can give meaning to the first term in the Neumann series by noting that it simply corresponds to the sum of the individual contributions of each scatterer. However, the entire series as formulated in Eq. (16) does not converge for this choice of parameters. Our estimate for the spectral radius of the iteration matrix, as given by Eq. (18), is in this case about 7, so this should not come as a surprise. The second term in the series is also indicated and differs considerably more from the correct solution than the lowest order approximation.

Applying the effective medium model to larger fish schools, additional low frequency features appear. Figure 7 visualizes the results of an effective medium calculation for a fish school containing 3000 fish of 20 cm length. All parameters are chosen to have the same values as in the example above (an average fish separation of 0.75 times the fish length, and a nominal depth of 2 m). Here, not only is the peak in the total cross section shifted to frequencies much lower than that of the individual fish, but also an additional dominant low frequency peak appears at about 90 Hz. We will see shortly that this peak corresponds to the lowest mode of oscillation, the ground state, of the entire fish school. The acoustic coupling between the individual fish leads to a resonant behavior of the fish school as a whole,

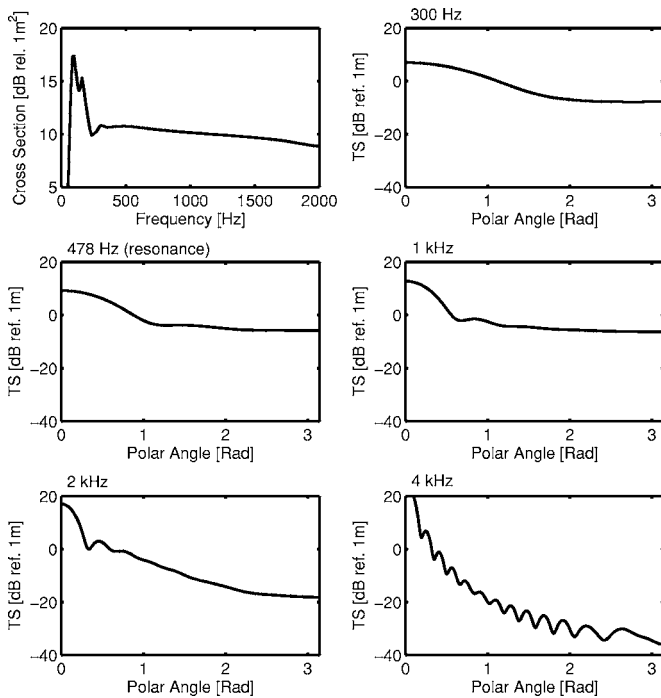


FIG. 7. Total coherent scattering cross section and TS for a school containing 3000 fish of length 0.2 m at a depth of 2 m spaced in average at 0.75 times the fish length. The TS is indicated for frequencies of 300 Hz, 478 Hz (bladder resonance), 1 kHz, 2 kHz, and 4 kHz.

which resembles that of an individual resonator. The analysis of section II, where we isolated the resonance contributions to the total scattering cross section, applies precisely to a situation like this. We will see below that by zooming in on this low frequency resonance regime, we can reconstruct the fish number without any consideration of other features of the scattering returns. The TS is still quite uniform at frequencies below the bladder resonance. At higher frequencies, more complicated interference structures become visible and the fall-off in the (coherent) TS towards the back direction becomes more evident.

As we have seen, the low frequency collective resonance modes do not occur under all circumstances. Only if the individual modes are separated enough from each other, given the widths of the resonances, can we expect to observe single noticeable peaks. Figure 8 depicts the total scattering cross section of schools like the ones above for increasing density but at fixed fish numbers. Conceptually, the fish are moving closer and closer together. For the small school (66 fish, top row), the results of increasing acoustic coupling is at first the gradual enhancement of the scattering cross section as the school gets packed more tightly. At a spacing of about 3 times the fish length, the growth in scattering strength slows and the peak frequency starts shifting to smaller frequencies. No secondary resonance features appear. We also note that the coherent scattering cross section can fall well below the incoherent sum of the individual scatterers. This is to be expected. At very low densities, the contributions of the individual scatters decouple and become diffusive. Indeed, one can easily convince oneself that, at low densities, the diffusive part of the cross section approaches the incoherent sum, which therefore does not constitute a lower bound for

the coherent part of the cross section. For the larger fish school (3000 fish, bottom row), the situation is somewhat different. Initially, the cross section also increases. At a spacing of about 4 times the fish length, the growth stops and the peak frequency starts shifting. As we continue increasing the fish density, the peak in the scattering cross section diminishes and, at a critical spacing of about 1.5 times the fish length, a second peak appears that quickly grows at further increased fish density.

Next, we turn to the modeling of the collective resonances. To start with, we need to verify the particular linearized form, Eq. (31), of the effective wave number that underlies the analytic modeling of the collective resonances of the fish school. The modification for elongated bladders in Minnaert's frequency, Eq. (49), carries through to the computation of κ , which in the case of fish becomes

$$\kappa = \sqrt{1 + \frac{\beta \rho c^2}{x(\epsilon)^2 \gamma_{\text{eff}} P_0}}. \quad (53)$$

Our treatment of collective resonances does not rely on $\kappa - 1$ being small but rather on the validity of the linearization. Indeed, Fig. 9 shows that close to the surface, the effective wave number can deviate significantly from the pure water value. Even for moderately dense schools, the deviations can easily reach 10% and can exceed 50% for very dense schools close to the surface. At greater depths, the deviations are only of the order of 1%. To investigate the quality of the linearization, Fig. 10 shows a comparison of this approximation with the full expressions Eqs. (1) and (20). For frequencies much smaller than the bladder resonance, k_e increases to a good approximation linearly. Describing collective oscillations, this is the regime we are most interested in. Closer to resonance, however, k_e increases more rapidly and the assumption of linearly increasing effective wave number starts failing.

We now apply the theory of Sec. II D to the low frequency collective resonance peak that emerged at high densities in the analysis that led to Fig. 8; the results are shown in Fig. 11. The accurate fit of the theoretical expression Eq. (44), particularly at high densities, verifies the claim that the observed peak corresponds to the lowest mode of the spherical school. Noting the wonderful match at various fish flesh viscosities, the incorporation of individual bladder damping on physical grounds and the inherent approximations employed in this scheme can now be considered verified. Widths and peaks are equally well described by the model—without fitting free parameters—from first principles. As expected, the deviations from the model become slightly larger at lower densities. In this case, the widths become larger and the peaks less pronounced. Consequently, resonance and background contributions from other modes will distort the single resonance result. This can paradoxically best be observed at the left-hand side of the resonance (there are no lower resonances). The full solution falls off more rapidly than the resonance contribution from the “ground state” due to interference from other modes. For sharp peaks, as is the case for dense bubble clouds, the resonance contribution Eq. (42) alone will yield satisfying results. Various stages of approximations are shown in Fig. 12. Although the resonance

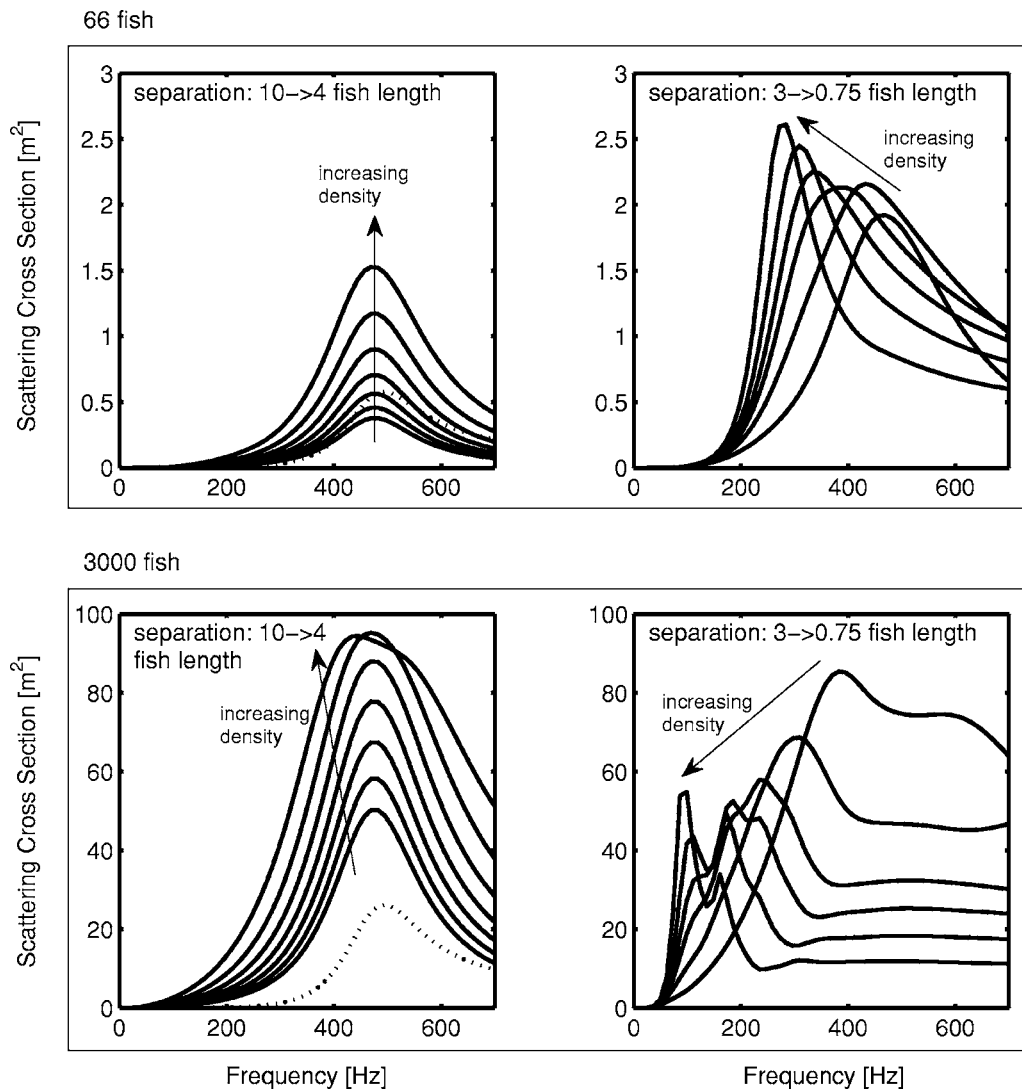


FIG. 8. Scattering cross section σ_s from Eq. (29) for fish schools containing 66 (top row) and 3000 fish (bottom row) of length 0.2 m at 2 m depth for various average separations: left images 10 to 4 times the fish length (arrows indicates decreasing separation and increasing density), and right images 3, 2, 1.5, 1.25, 1, and 0.75 times the fish length. The dashed line in the left image shows for comparison the incoherent sum of the scattering cross sections of all individual fish.

frequency and the width of the peak are approximately described by the resonance contributions alone, the background and the damping induced by the individual fish (as opposed to the radiation damping caused by the impedance contrast with the surrounding medium) must be included for moderately dense schools to achieve the quality of the matching from the previous example.

If the lowest mode resonance is clearly visible without much interference from neighboring resonances, a direct inversion for the total fish number from Eqs. (38), (41), and (44) can be attempted. For very dense schools (here we show a rather unrealistically dense school for demonstration purposes) as shown in the top panel of Fig. 13, the spacing can be reconstructed to about 1% and the school radius to about 2%, yielding a fish number estimate good to about 10%. For smaller densities, the accuracy degrades somewhat (see lower panel). Although mathematically unique, the inversion might suffer from the fact that the “hot spot” region of small

error is fairly elongated. In practice, this could lead to larger uncertainties, mistaking a small dense school for a larger one of smaller density.

As we saw above, in principle, both the fish spacing or equivalently the fish density, and the school radius can be extracted from the lowest school resonance alone. Additionally, the mere observation of the low frequency oscillations or their absence is an indicator of the fish density. Not all large schools need to show low frequency resonance peaks. Figure 14 gives the total coherent cross section of a school of 100 000 fish of 20 cm length at a depth of 50 m. Even at this large fish number, the computation is straightforward and can be done in a matter of a few minutes. Nevertheless, more than 150 partial wave terms have to be summed in Eq. (29) to obtain a stable result. At this depth, the swim bladder is significantly compressed in comparison with the surface school examples given above, which can be seen in the significant increase of the bladder resonance from 480 Hz to 1.8 kHz. This leads to a reduced individual scattering

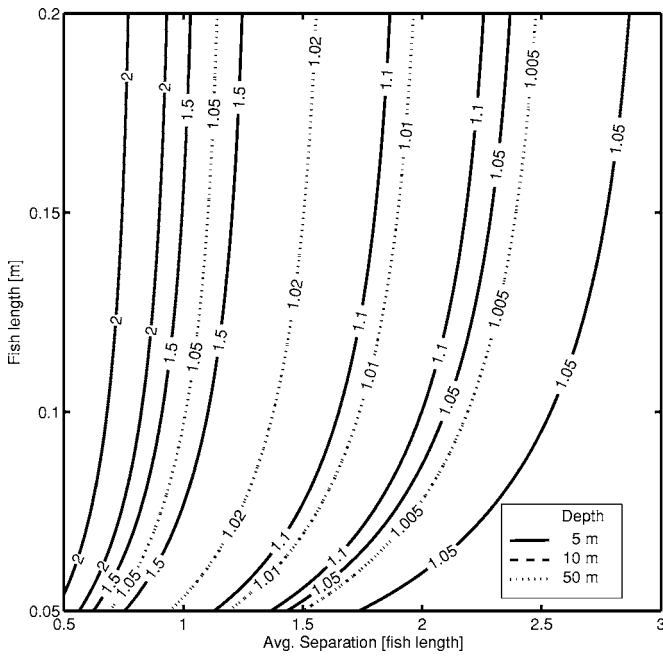


FIG. 9. κ , Eq. (31), for various fish lengths and average fish separations. The solid line corresponds to fish located at a depth of 5 m, the dashed and dotted lines show the values for fish at depths of 10 and 50 m, respectively.

strength at lower frequencies. Hence, there are no visible individual resonances at a spacing of one times the fish length. (In this case, modes would appear roughly at a fish spacing of about one half times the fish length.) Nevertheless, multiple scattering dramatically shifts the peak from the individual bladder resonance of 1.8 kHz down to about

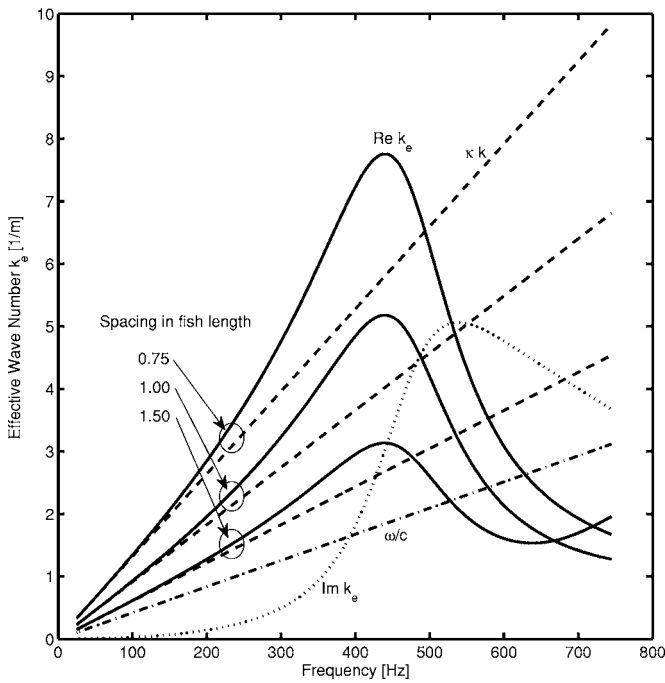


FIG. 10. The real part of the effective wave number k_e (solid lines) vs the approximation κk , Eq. (31), (dashed lines) for fish of length 0.2 cm spaced at 0.75, 1, and 1.5 times the fish length at a depth of 2 m. The natural bladder resonance frequency is about 478 Hz. The imaginary part of k_e is also shown (dotted line) for the case of fish spaced at 1 times the fish length. The dash-dotted line shows the free space wave number $k=2\pi f/c$.

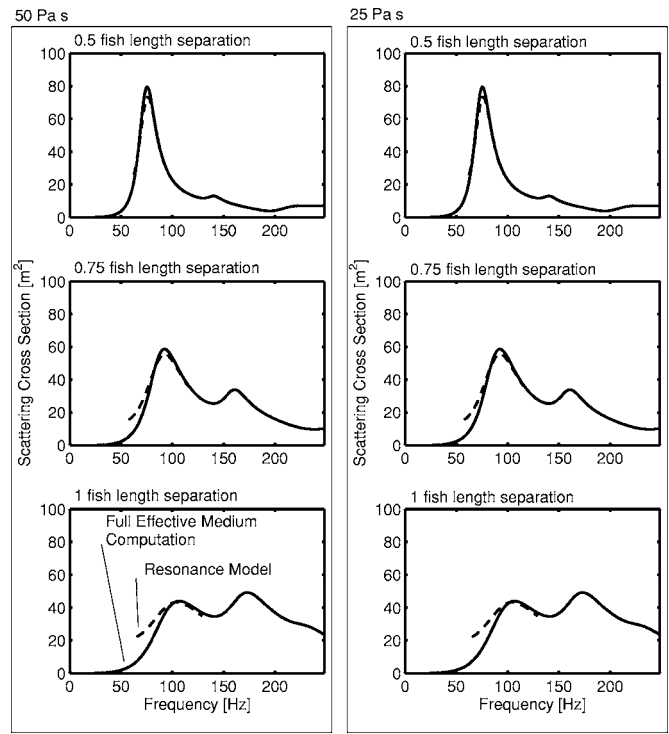


FIG. 11. Modeling the low frequency resonance scattering cross section σ_s from Eq. (29) for a fish school containing 3000 fish of length 0.2 m at 2 m depth at an average spacing of (top) 0.5, (middle) 0.75, and (bottom) 1 times the fish length. The dashed line shows Eq. (44), including all background and damping effects. The three graphs on the left are based on a fish flesh viscosity of 50 Pa s (500 P), the graphs on the right hand side are for comparison based on a smaller viscosity of 25 Pa s (250 P).

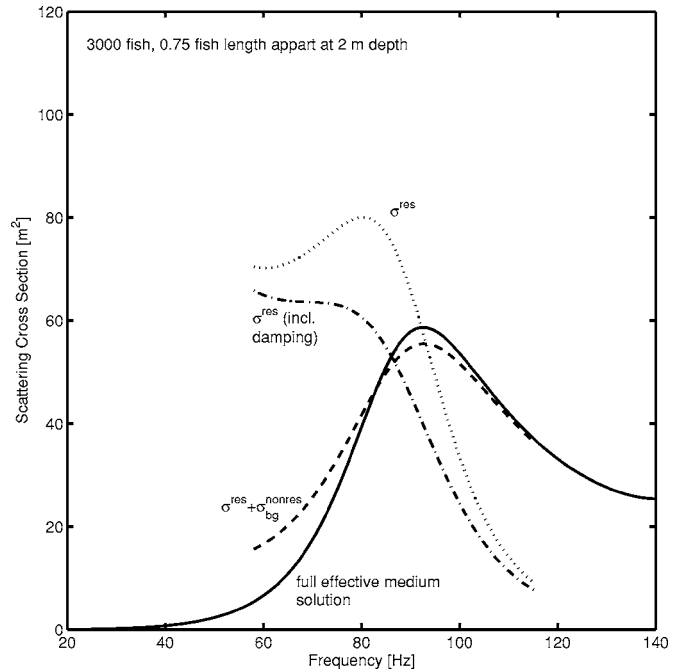


FIG. 12. Various approximations to the resonance scattering cross section σ_s from Eq. (29) for a fish school containing 3000 fish of length 0.2 m at 2 m depth at an average spacing of 0.75. (This case corresponds to the middle left-hand panel of Fig. 11.) The dotted line shows the pure resonance contribution σ_s^{res} Eq. (42), the inclusion of resonant damping [resonance part of Eq. (44), dash-dotted line] brings the peak down and the full model, Eq. (44), finally yields the excellent fit indicated by the dashed line.

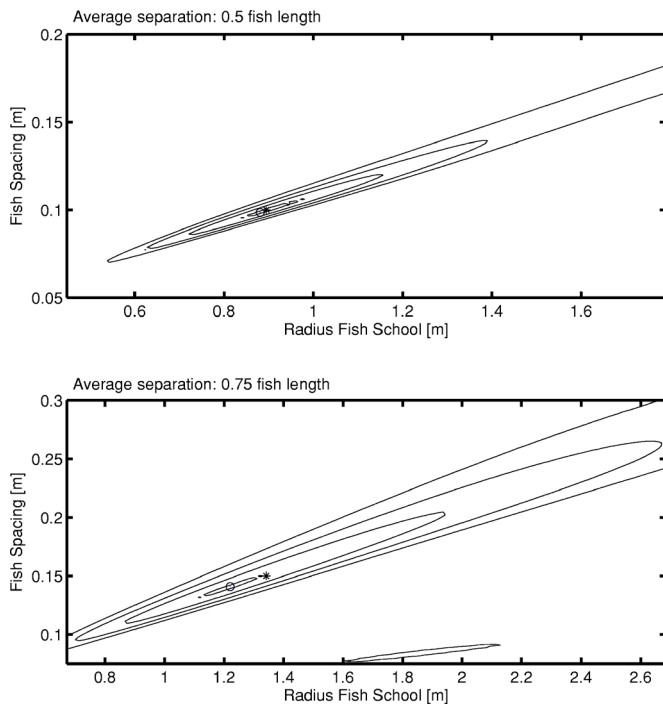


FIG. 13. (Color online) Contour lines of the norm of the difference between the effective medium scattering cross section σ_s , Eq. (29), and the resonance cross-section model $\sigma^{res}(k)$, Eq. (44), as a function of the fish spacing and the fish school radius. The fish school contains 3000 fish of length 0.2 m at 2 m depth at an average spacing of (top) 0.5 and (bottom) 0.75 times the fish length. The circle indicates the best fit and the star shows the actual values of the two parameters.

800 Hz. Again, the coherent cross section peaks in the forward direction, particularly pronounced at higher frequencies. This must be the case because coherent energy is removed from the beam. For a fish school of this size, polar interference patterns are visible in the TS below 2 kHz. Above that, the oscillations with polar angle become narrow and the coherent TS envelope falls off rapidly towards the back direction. Note that in back direction the diffusive part can be assumed to dominate at high frequencies.

C. Applicability and limitations

The exemplary application of our model to spherical fish schools of various sizes and densities mainly serves to illuminate general multiple scattering effects and their handling

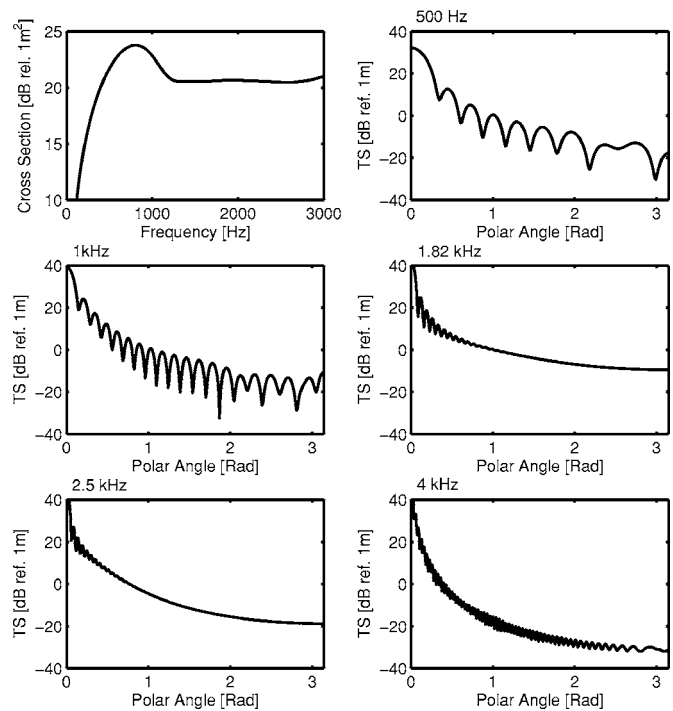


FIG. 14. Total coherent scattering cross section and TS for a school containing 100 000 fish of length 0.2 m at a depth of 50 m, spaced in average at 1.0 times the fish length. The TS is indicated for frequencies of 500 Hz, 1 kHz, 1.82 kHz (bladder resonance), 2.5 kHz, and 4 kHz.

in the framework of the effective theory approach. However, despite its idealized form, the preceding analysis of fish schools could have several practical applications. During predator-prey interactions many common forage fishes, such as herring,^{53,54} anchovy, sardines, sand lance and more, can form nearly spherical schools near the surface of extremely high densities. Figure 15 shows examples of such fish schools, referred to as a near-surface herring “bait balls.” The development of a quantitative measurement technique for these schools is key to answering several important questions regarding group foraging by marine predators. In cases like this, the presented analysis based on the more general theoretical model could be of immediate predictive use.

As a good approximation, the model in its present form could serve to estimate the low frequency behavior of ball-like schools of compact appearance, which are often

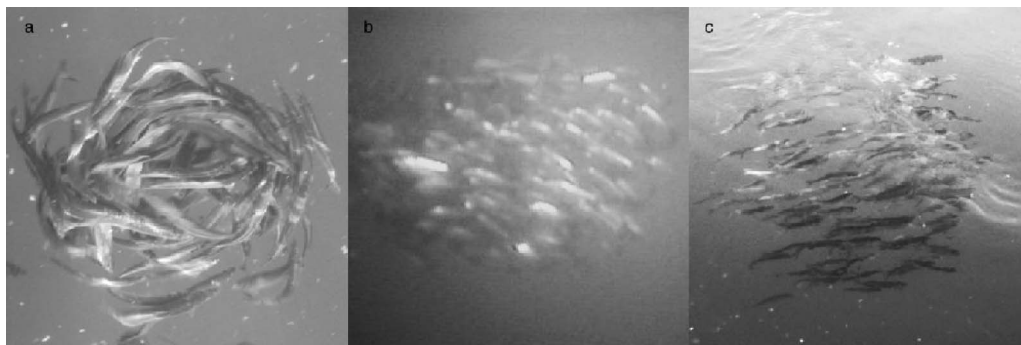


FIG. 15. Bait ball schools of Pacific Herring. (a) $O(100)$ fish of about 22 cm length spaced at 0.5 to 1 fish lengths (visual estimates), (b) $O(1000)$ fish of about 22 cm length spaced at <1 fish length (visual estimates), and (c) spherical bait ball from above (compliments of Matt Foster, Alaska Fish and Game; and Dr. G. L. Thomas, RSMAS.).

observed.⁵⁵ Although, in this case, the predictive power for the collective modes might be degraded, the model could still be used to estimate the coherent TS in a general bi-static transmitter-receiver arrangement. Together with a propagation model, such as Monterey-Miami parabolic equation model (MMPE) of K. Smith and F. Tappert, the effective medium approach could aid in the determination of the necessary source levels and in the design of the receiving elements for longer range fish school monitoring efforts. Further, presence or absence of low frequency collective resonances in the broadband spectra could, following the approach of Sec. II, yield a first indication of fish density. Supplying independent volume and shape estimates from acoustic mapping of fish concentrations (e.g., by deploying horizontal arrays of sufficient aperture) will increase the feasibility of this approach. For less compact shapes, such as layers or macro structures,⁵⁶ for geometries that entirely deviate from a ball-like shape, diffuse massive fish aggregations¹² that span hundreds of meters or kilometers, it can “as-is” not be applied, although some multiple scattering features such as frequency shifts likely might still hold qualitatively.

To validate the conjectures made in the previous two paragraphs a significant experimental effort is necessary. Such an effort would ideally combine low frequency acoustic measurements with traditional high frequency acoustics as well as optical and purse seining techniques to independently assess fish densities, school volume, degree of organization, and individual fish statistics (species, size distribution, etc.). Because much of the interesting low frequency scattering phenomena concern the coherent portion of the scattered sound field, the use of linear maximal sequences could be explored experimentally.

IV. CONCLUDING REMARKS

In this work, we presented a theoretical treatment of acoustic scattering from clouds of bubbles, or other monopole scatterers, addressing spherically shaped assemblages. This clearly has the character of a physical toy model but, nevertheless, highlights the effects of multiples scattering and their theoretical treatment within the framework of the effective medium theory. It will be straightforward to relax some of the assumptions to treat more realistic physical systems and, more importantly, will set the stage to address a series of theoretical questions that go beyond what we presented here. To mention only some: What is the effect of more complicated assembly geometries on collective resonances, frequency shifts, etc.? What is the influence of the constituent motion on low frequency acoustics? How can longer-range correlations be incorporated into the model? What is the effect of nearby surfaces? What is the effect of surface roughness? As more real data becomes available for selected applications, some or all of these questions will need to be considered. However, we believe that the model in its canonical form based on simple scattering characteristics and a clean geometry captures the essence of the most important physical properties of the system in question and serves as an anchor point for further investigations.

ACKNOWLEDGMENTS

This work is supported by Dr. Ellen Livingston, the Office of Naval Research, Grant No. N000140710896. The author would like to acknowledge Zhen Ye (deceased) for providing his own numerical benchmarks to verify his code, Orest Diachok for encouraging me to study fish school resonances at the 148th meeting of the Acoustical Society in San Diego, and Mohamed Iskandarani for valuable discussions. Most of all, the author is indebted to Tony Elfouhaily (deceased) for countless critical discussions about the model and scattering physics in general, Gary Thomas for many discussions about the biology of fish schools, and Otis Brown, the dean of the Rosenstiel School, for supporting this research. Finally, the author is very grateful for the engaged and valuable feedback from the associate editor, K. G. Foote.

- ¹R. A. Roy, W. Carey, M. Nicholas, J. Schindall, and L. A. Crum, “Low-frequency scattering from submerged bubble clouds,” *J. Acoust. Soc. Am.* **92**, 2993–2996 (1992).
- ²K. Sarkar and A. Prosperetti, “Backscattering of underwater noise by bubble clouds,” *J. Acoust. Soc. Am.* **93**, 3128–3138 (1993).
- ³A. R. Kolaini, “Sound radiation by various types of laboratory breaking waves in fresh and salt water,” *J. Acoust. Soc. Am.* **103**, 300–308 (1998).
- ⁴A. R. Kolaini and L. A. Crum, “Observations of underwater sound from laboratory breaking waves and the implications concerning ambient noise in the ocean,” *J. Acoust. Soc. Am.* **96**, 1755–1765 (1994).
- ⁵E. Lamarre and W. K. Melville, “Void-fraction measurements and sound-speed fields in bubble plumes generated by breaking waves,” *J. Acoust. Soc. Am.* **95**, 1317–1328 (1994).
- ⁶M. R. Loewen and W. K. Melville, “An experimental investigation of the collective oscillations of bubble plumes entrained by breaking waves,” *J. Acoust. Soc. Am.* **95**, 1329–1343 (1994).
- ⁷A. Prosperetti, “Bubble-related ambient noise in the ocean,” *J. Acoust. Soc. Am.* **84**, 1042–1054 (1988).
- ⁸R. Manasseh, A. V. Babanin, C. Forbes, K. Rickards, I. Bobevski, and A. Ooi, “Passive acoustic determination of wave-breaking events and their severity across the spectrum,” *J. Atmos. Ocean. Technol.* **23**, 599–618 (2006).
- ⁹D. M. Farmer, M. V. Trevorrow, and B. Pedersen, “Intermediate range fish detection with a 12-kHz sidescan sonar,” *J. Acoust. Soc. Am.* **106**, 2481–2490 (1999).
- ¹⁰C. Feuillade, R. W. Nero, and R. H. Love, “A low-frequency acoustic scattering model for small schools of fish,” *J. Acoust. Soc. Am.* **99**, 196–208 (1996).
- ¹¹V. Holliday, “Resonance structures in echoes from schooled pelagic fish,” *J. Acoust. Soc. Am.* **51**, 1098–1105 (1972).
- ¹²N. C. Makris, P. Ratilal, D. T. Symonds, S. Jagannathan, S. Lee, and R. W. Nero, “Fish population and behavior revealed by instantaneous continental shelf-scale imaging,” *Science* **311**, 660–663 (2006).
- ¹³O. Diachok, “Effects of absorptivity due to fish on transmission loss in shallow water,” *J. Acoust. Soc. Am.* **105**, 2107–2128 (1999).
- ¹⁴O. Diachok, “Absorption spectroscopy: A new approach to estimation of biomass,” *Fish. Res.* **47**, 231–244 (2000).
- ¹⁵O. Diachok, B. Liorzou, and C. Scalabrin, “Estimation of the number density of fish from resonance absorptivity and echo sounder data,” *ICES J. Mar. Sci.* **58**, 137–153 (2001).
- ¹⁶J. Greinert, Y. Artemov, V. Egorov, M. De Batist, and D. McGinnis, “1300-m-high rising bubbles from mud volcanoes at 2080 m in the Black Sea: Hydroacoustic characteristics and temporal variability,” *Earth Planet. Sci. Lett.* **244**, 1–15 (2006).
- ¹⁷R. Merewether, M. S. Olsson, and P. Lonsdale, “Acoustically detected hydrocarbon plumes rising from 2-km depths in Guaymas Basin, Gulf of California,” *J. Geophys. Res.*, [Solid Earth Planets] **90**, 3075–3085 (1985).
- ¹⁸A. O. Maksimov, “Noise spectrum of a gas plume,” *Acoust. Phys.* **51**, 435–442 (2005).
- ¹⁹A. L. Anderson and L. D. Hampton, “Acoustics of gas-bearing sediments. 1. Background,” *J. Acoust. Soc. Am.* **67**, 1865–1889 (1980).
- ²⁰B. 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).

- ²¹A. 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).
- ²²A. S. Burgess and D. J. Kewley, "Wind-generated surface noise source levels in deep-water east of Australia," *J. Acoust. Soc. Am.* **73**, 201–210 (1983).
- ²³R. D. Hollett, "Observations of underwater sound at frequencies below 1500-Hz from breaking waves at sea," *J. Acoust. Soc. Am.* **95**, 165–170 (1994).
- ²⁴B. R. Kerman, "Underwater sound generation by breaking wind-waves," *J. Acoust. Soc. Am.* **75**, 149–165 (1984).
- ²⁵P. C. Wille and D. Geyer, "Measurements on the origin of the wind-dependent ambient noise variability in shallow-water," *J. Acoust. Soc. Am.* **75**, 173–185 (1984).
- ²⁶W. M. Carey and M. P. Bradley, "Low-frequency ocean surface noise sources," *J. Acoust. Soc. Am.* **78**, S1 (1985).
- ²⁷W. M. Carey and J. W. Fitzgerald, "Low-frequency noise and bubble plume oscillations," *J. Acoust. Soc. Am.* **82**, S62 (1987).
- ²⁸A. Prosperetti, in *Sea Surface Sound*, edited by B. R. Kerman (Kluwer Academic, Dordrecht, The Netherlands, 1988), pp. 151–171.
- ²⁹W. M. Carey, J. W. Fitzgerald, E. C. Monahan, and Q. Wang, "Measurement of the sound produced by a tipping trough with fresh and salt-water," *J. Acoust. Soc. Am.* **93**, 3178–3192 (1993).
- ³⁰T. R. Hahn, T. K. Berger, and M. J. Buckingham, "Acoustic resonances in the bubble plume formed by a plunging water jet," *Proc. R. Soc. London, Ser. A* **459**, 1751–1782 (2003).
- ³¹A. R. Kolaini, R. A. Roy, L. A. Crum, and M. Yi, "Low-frequency underwater sound generation by impacting transient cylindrical water jets," *J. Acoust. Soc. Am.* **94**, 2809–2820 (1993).
- ³²S. W. Yoon, L. A. Crum, A. Prosperetti, and N. Q. Lu, "An investigation of the collective oscillations of a bubble cloud," *J. Acoust. Soc. Am.* **89**, 700–706 (1991).
- ³³P. Tkalic and E. S. Chan, "Breaking wind waves as a source of ambient noise," *J. Acoust. Soc. Am.* **112**, 456–463 (2002).
- ³⁴K. W. Commander and A. Prosperetti, "Linear pressure waves in bubbly liquids—comparison between theory and experiments," *J. Acoust. Soc. Am.* **85**, 732–746 (1989).
- ³⁵S. G. Kargl, "Effective medium approach to linear acoustics in bubbly liquids," *J. Acoust. Soc. Am.* **111**, 168–173 (2002).
- ³⁶A. Prosperetti, "Bubble phenomena in sound fields: part one," *Ultrasonics* **22**, 69–77 (1984).
- ³⁷U. Frisch, *Wave Propagation in Random Media* (Academic, New York, 1968), pp. 75–198.
- ³⁸L. L. Foldy, "The multiple scattering of waves," *Phys. Rev.* **67**, 107–109 (1945).
- ³⁹N. Q. Lu, A. Prosperetti, and S. W. Yoon, "Underwater noise emissions from bubble clouds," *IEEE J. Ocean. Eng.* **15**, 275–281 (1990).
- ⁴⁰T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
- ⁴¹C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles and Applications* (Wiley, New York, 1977), pp. xviii and 1977), 544.
- ⁴²F. Minnaert, "On musical air bubbles and the sounds of running water," *Philos. Mag.* **16**, 235–248 (1933).
- ⁴³R. H. Love, "Resonant acoustic scattering by swimladder-bearing fish," *J. Acoust. Soc. Am.* **64**, 571–580 (1978).
- ⁴⁴N. A. Gumerov and R. Duraiswami, "Computation of scattering from clusters of spheres using the fast multipole method," *J. Acoust. Soc. Am.* **117**, 1744–1761 (2005).
- ⁴⁵S. Koc and W. C. Chew, "Calculation of acoustical scattering from a cluster of scatterers," *J. Acoust. Soc. Am.* **103**, 721–734 (1998).
- ⁴⁶J. M. Ortega, *Numerical Analysis, A Second Course* (Siam, Philadelphia, 1990).
- ⁴⁷P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Mc Graw-Hill, New York, 1968).
- ⁴⁸D. E. Weston, in *Underwater Acoustics*, edited by V. M. Albers (Plenum, New York, 1967).
- ⁴⁹L. A. Crum, "The polytropic exponent of gas contained within air bubble pulsating in a liquid," *J. Acoust. Soc. Am.* **73**, 116–120 (1983).
- ⁵⁰M. Strassberg, "The pulsating frequency of nonspherical gas bubbles in liquids," *J. Acoust. Soc. Am.* **25**, 536–537 (1953).
- ⁵¹J. Blaxter and R. Batty, "Swimbladder behaviour and target strength," *Rapp. P.-V. Reun.-Cons. Int. Explor. Mer* **189**, 233–244 (1990).
- ⁵²A. Lovik and J. Hovem, "An experimental investigation of swim bladder resonances in fishes," *J. Acoust. Soc. Am.* **66**, 850–854 (1979).
- ⁵³L. Nottestad and B. E. Axelsen, "Herring schooling manoeuvres in response to killer whale attacks," *Can. J. Zool.* **77**, 1540–1546 (1999).
- ⁵⁴G. L. Thomas and R. E. Thorne, "Night-time predation by Steller sea lions," *Nature (London)* **411**, 1013–1013 (2001).
- ⁵⁵O. A. Misund, A. Aglen, and E. Fronaes, "Mapping the shape, size, and density of fish shoals by echo integration and a high-resolution sonar," *ICES J. Mar. Sci.* **52**, 11–20 (1995).
- ⁵⁶P. Petitgas and J. J. Levenez, "Spatial organization of pelagic fish: Echogram structure, spatio-temporal condition, and biomass in Senegalese waters," *ICES J. Mar. Sci.* **53**, 147–153 (1996).

Geoacoustic inversion using multipath pulse shape

Mark K. Prior,^{a)} Chris H. Harrison,^{b)} and Peter L. Nielsen^{c)}

NATO Undersea Research Centre, Viale S. Bartolomeo 400, 19138 La Spezia, Italy

(Received 20 February 2007; revised 27 June 2007; accepted 27 June 2007)

Experimental data, measured in a shallow water region of the Mediterranean Sea, are used to show that the variation of received intensity with time is well described by existing expressions [Harrison and Nielsen, *J. Acoust. Soc. Am.* **121**, 1362–1373 (2007)]. These expressions indicate that the effect of the sea-water sound speed profile can be neglected for times greater than the peak intensity arrival. Beyond this time, intensity is shown to decay at a rate determined by the seabed acoustic properties in a manner very similar to that for an isovelocity water column. It is shown that a method of determining seabed acoustic properties, previously restricted to isovelocity water columns [Prior and Harrison, *J. Acoust. Soc. Am.* **116**, 1341–1344 (2004)], can consequently be used in the presence of a sound-speed profile. The method relates the decay rate of smeared multipath arrivals to the angular derivative of seabed reflection loss. Two datasets are studied and the method is used to describe average seabed properties and to detect changes in seabed type. The seabed descriptions thus derived are used to predict total received intensity as a function of source-receiver separation. Agreement between the propagation measurements and predictions is shown to be within measurement uncertainties. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2764468]

PACS number(s): 43.30.Ma, 43.20.El [AIT]

Pages: 3268–3279

I. INTRODUCTION

Underwater acoustic propagation in shallow water environments is critically affected by the composition of the seabed (Kuperman and Jensen, 1980), but it is commonly the case that this composition is not well-enough described in databases to allow accurate prediction of acoustic propagation (Ferla and Jensen, 2002). This limits the accuracy of predictions of sonar system performance and complicates the process of the assessment of the impact of underwater sound on marine mammals (Richardson *et al.*, 1995).

While seabed properties can be assessed by coring and subsequent laboratory analysis (Hamilton, 1980), the difficulty and expense of gathering such information (in quantities sufficient to allow accurate propagation predictions) are prohibitive. An alternative approach is to make controlled measurements of acoustic propagation and to “invert” these measurements to produce estimates of the properties of the seabed in the experimental area (Collins *et al.*, 1992). Such approaches commonly use vertical arrays of sensors to measure acoustic pressure and employ computationally intensive approaches to optimize the match between measurements and the predictions of full-wave propagation models (Gerstoft, 1994).

While these inverse methods are capable of producing good descriptions of seabed acoustic properties, they are generally restricted to research applications by consequence of their use of specialist arrays and computationally intensive analysis schemes. To circumvent this restriction, methods have been proposed (Hermand, 1999; Le Gac *et al.*, 2003; Siderius and Hermand, 1999) in which inversions are per-

formed on broadband acoustic signals, received on more sparse arrays. Such approaches have the potential of being used with simple, cheap, and easily deployable sonar systems. Other attempts to develop acoustic inversion methods not reliant on large-aperture vertical arrays have used towed horizontal arrays (Siderius *et al.*, 2002; Fallat *et al.*, 2005) to gather data that are subsequently inverted using nonlinear schemes.

A more direct method for estimating seabed reflection loss properties (Harrison, 2003; Smith, 1971), based on single-hydrophone measurements of multipath pulse shape, has previously been shown to be capable of showing good agreement with matched-field inversion methods that use a vertical array of hydrophones spanning most of the water column (Prior and Harrison, 2004). The method exploits a simple relationship between seabed reflection loss and the decay rate of acoustic intensity (in dB per second) measured at a receiver located many water depths from a broadband acoustic source. The method’s use of pulse shape, rather than absolute intensity level, means that it does not require the use of calibrated sources or receivers. The method is “direct” in the sense that it does not use inverse methods such as genetic algorithms (Gerstoft, 1994) to search for an optimal environment. Despite this caveat, the method is referred to as an “inversion” because, in common with full-field approaches, it takes acoustic data and “inverts” it to produce estimates of environmental conditions.

While encouraging results have been produced by the method it has so far been based on closed-form expressions for received intensity that are themselves based on an assumption of an isovelocity water column whose validity is questionable in the majority of realistic ocean environments. Recently, however, closed-form expressions for pulse shape, including in-water refraction and lossy reflection at bound-

^{a)}Electronic mail: prior@nurc.nato.int

^{b)}Electronic mail: harrison@nurc.nato.int

^{c)}Electronic mail: nielsen@nurc.nato.int

aries, have been derived (Harrison and Nielsen, 2007) that have been shown to agree with numerical calculations.

This paper describes the use of these expressions to extend the validity of the pulse-shape-inversion method (Prior and Harrison, 2004) to include environments with significant sound-speed gradients present in the water column. It is shown that the original method can be used in an unmodified form with data gathered in the presence of a significant sound speed profile, provided that only the portion of the pulse arriving after the peak intensity is considered. The method is applied to acoustic data gathered in a shallow water region of the Mediterranean Sea to determine average seabed acoustic properties and to identify horizontal changes in seabed type along the source-receiver track.

Section II describes the essentials of the closed-form expressions for the pulse shape, on which the inversion method is based. Pulse shapes predicted by these expressions are then compared with measured data in Sec. III and seabed acoustic properties are estimated from consideration of pulse shapes in Sec. IV.

II. SUMMARY OF THEORY

The full derivation of the closed-form expressions for pulse shape is given elsewhere (Harrison and Nielsen, 2007) and only a summary of those parts relevant to the experimental conditions in which data were gathered is provided here.

The approach considers a constant uniform gradient in the water column. The reflecting surfaces that bound the water column are described as the “low-sound-speed boundary” and “high-sound-speed boundary,” separated by a water depth H and having sound speeds c_L and c_H , respectively. Reflection at the boundaries is lossy with reflection loss proportional to incident angle and with the constants of proportionality at the two surfaces being α_L and α_H . That is, at the low-sound-speed boundary, the reflection coefficient, R_L , is given by

$$R_L = \exp[-\alpha_L \theta_L], \quad (1)$$

where θ_L is the angle at which sound impinges on the boundary and α_L has units of nepers per radian.

The closed-form expressions for pulse shape are derived (Harrison and Nielsen, 2007) by calculating the distribution of intensity with angle and then transforming this into a distribution of intensity with time by consideration of the relationship between the angle of a ray path, its cycle range, and its cycle time. The expressions are general in H , c_L , c_H , α_L , α_H , and the source-receiver range, r .

The experiments considered here had source and receiver in a downward-refracting profile with a smooth sea surface. Thus the low-sound-speed boundary was the only one at which reflection loss was encountered and α_H (which in these circumstances would be a measure of sea surface reflection loss) was set to zero. Furthermore, it was assumed that the travel time of the first arrival was large in comparison to the pulse length, an assumption that is valid for propagation over ranges of more than a few hundred meters.

Under these assumptions, the closed-form expressions predict that the pulse shape is divided into two distinct por-

tions: an initial crescendo and a smoothly decaying trailing edge. The crescendo is formed by the arrival of sound that has followed paths that are reflected at the seabed but are refracted down before they interact with the sea surface. The relationships between angle, travel time, and boundary reflection loss are such that received intensity rises from its value at the time of the first arrival, t_{\min} , until reaching a peak at a later time, t_{\max} , corresponding to the arrival of rays whose path just grazes the sea surface. The width of the crescendo, T_0 , is expressed [Harrison and Nielsen, 2007, Eq. (39)] as

$$T_0 = t_{\max} - t_{\min} = T_L \frac{\delta_{c_{sr}}}{3c_L}, \quad (2)$$

where T_L is the range divided by the sound speed at the lower boundary and $\delta_{c_{sr}}$ is the sound speed difference between source and receiver. The width of the leading edge of the pulse, i.e., the part before the peak intensity arrival, thus increases with source-receiver range and with sound-speed contrast between source and receiver.

For times greater than t_{\max} , sound can only arrive at the receiver via paths that interact both with the sea surface and the seabed and the intensity associated with these paths, I , is calculated by a separate expression that indicates a decay of intensity with a form [Eq. (A19) from Harrison and Nielsen (2007)]:

$$I(\tau)d\tau = \frac{\sqrt{2}}{c_{av}Ht_0^{3/2}\sqrt{\tau}} \exp\left[\frac{-c_{av}\alpha_L\tau}{H}\left(1 + \frac{Hc'}{2c_L}\right) - \frac{c't_0c_{av}\alpha_L}{4c_L}\right]d\tau, \quad (3)$$

where τ is the reduced time, i.e., the time after the peak arrival at time t_0 , c' is the sound speed gradient, and c_{av} is an average sound speed, given by

$$c_{av} = 2 \frac{c_L c_H}{c_L + c_H}. \quad (4)$$

The expression for the pulse shape in the absence of a sound-speed profile, previously derived (Smith, 1971; Harrison, 2003) and already used to deduce seabed reflection properties (Prior and Harrison, 2004), is of the form

$$I(\tau)d\tau = (2t_0)^{-1/2} \frac{\exp\{-\alpha H\tau/c\}}{\sqrt{\tau}} d\tau. \quad (5)$$

The decay rate of logarithmic intensity, after scaling by the square root of τ , given by

$$\frac{d}{d\tau} \{\log[\sqrt{\tau}I(\tau)d\tau]\}, \quad (6)$$

is therefore modified from the isovelocity case only by the use of an average sound speed and by the presence of the multiplier $[1 + (Hc')/(2c_L)]$. The ratio in this multiplier can be estimated to be of the order of 1/100, assuming a value of 30 m/s for the numerator and 3000 m/s for the denominator. The slope of the trailing edge of the pulse is therefore quasi-independent of the sound-speed profile. This quasi-independence arises physically from the fact that the trailing

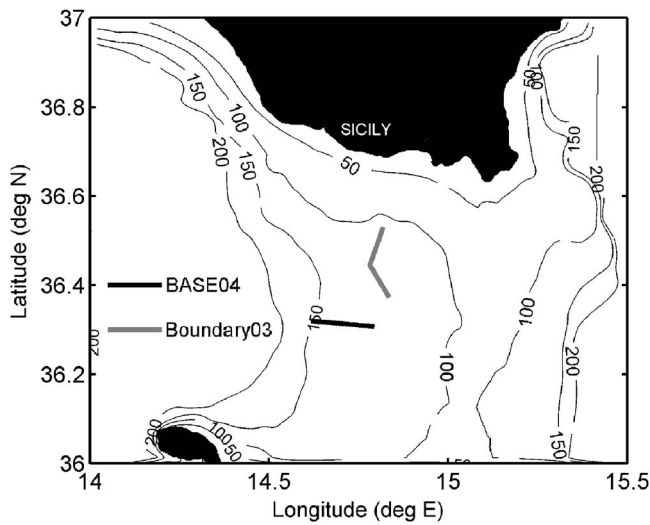


FIG. 1. Map of experimental area. Contours show water depth in meters.

edge of the pulse is made up of steep, nearly straight rays that traverse the entire water column and interact with both boundaries. Such sound is, by definition, only weakly affected by the sound-speed profile.

The pulse shape received in a downward-refracting environment is thus predicted to be made up of a leading edge whose width increases with source-receiver range and source-receiver sound-speed contrast and which ends with the arrival of the peak of the pulse. This is then predicted to be followed by a smooth decay of logarithmic intensity with time after the peak arrival. The rate of decay of intensity, after the peak intensity arrival, is predicted to be independent of source-receiver range and to be insensitive to the details of the seawater sound speed profile.

This predicted behavior is now investigated using measured data.

III. COMPARISON OF MEASURED AND PREDICTED PULSE SHAPES

A. General data description

Acoustic measurements were made during the NURC Boundary 03 and BASE 04 experiments in the Malta Plateau region, south of Sicily, following the paths shown in Fig. 1. These data were examined to investigate the shapes of pulses received after one-way transmission through a shallow water ocean environment. Transmissions were made from a towed source and received on a distant vertical line array (VLA) of receivers.

The signals used in the two experiments were linear frequency modulated and had similar center frequencies and bandwidths representing a considerable fraction of this center frequency.

The source comprised three, free-flooding rings, which, when used simultaneously, formed a transmit beam spread ± 12 deg about the horizontal axis. When only one of the three rings was used during transmission, an effectively omnidirectional beam resulted. Although the one- and three-ring configurations of the source gave different vertical directionalities on transmission, it should be remembered that after

sound has propagated to ranges equivalent to a large number of water depths, most high-angle sound will have been “stripped out” (Harrison, 2003) by repeated bottom reflections. Consequently, the angular spread of sound incident on a receiver at ranges more than 30 water depths (Harrison, 2003) is insensitive to the vertical directivity of the source. This can be illustrated by consideration of a source with a notional Gaussian beam pattern on transmission with width θ_w . The angular distribution of intensity after propagation to a range r in an environment with water depth H and a seabed reflection loss gradient of α is given by (Harrison, 2003):

$$\begin{aligned}
 I(\theta)d\theta &= \exp\left[-\frac{\theta^2}{\theta_w^2}\right] \exp\left[-\frac{\alpha r \theta^2}{2H}\right] \\
 &= \exp\left[-\theta^2\left(\frac{1}{\theta_w^2} + \frac{\alpha r}{2H}\right)\right].
 \end{aligned}
 \tag{7}$$

The bracketed term in the exponential function is effectively independent of θ_w for large r since the second term in the bracket dominates the first.

Source levels of 216 and 206 dB re $1 \mu\text{Pa}$ at 1 m were used for the three- and one-ring configurations, respectively.

Upon reception, signals were correlated with the transmitted signal (i.e., replica correlated) and, since signal bandwidth was never less than 600 Hz, temporal resolution of at least 1.7 ms was achieved. This high temporal resolution ensured that individual paths from source to receiver could be resolved, allowing investigation of the distribution of intensity with time for one-way propagation paths. No beamforming was performed on the data received on the VLA, but intensity data, depth averaged over the VLA, were used in addition to data from single hydrophones.

Although the two experiments were carried out in the same geographic area and used similar signals and sonar equipment, the two experiments covered dissimilar seabed types, had different seawater-sound-speed profiles, and had slightly different source-receiver geometries. Consequently, data from the two experiments are now considered separately.

B. Base 04

Sound speed profiles recorded by expendable bathythermograph casts during the BASE 04 experiment are shown in Fig. 2. Sound speed close to the surface was consistently higher than that at greater depths and sound-speed variations over the entire water column of up to 10 m/s were observed. The greater part of the reduction of sound speed with depth took place in a thermocline region located between 40 and 60 m below the surface.

Source and receiver were placed below this thermocline and sensor depths and pulse properties are summarized in Table I. The vertical extent of the VLA is shown by the straight, black lines in Fig. 2 and the depths of the sources and single-hydrophone receivers are shown by diamond and circular markers, respectively.

The track shown in Fig. 1 for the BASE 04 experiment covers a region in which water depth does not vary significantly from a mean value around 130 m. The seabed in the

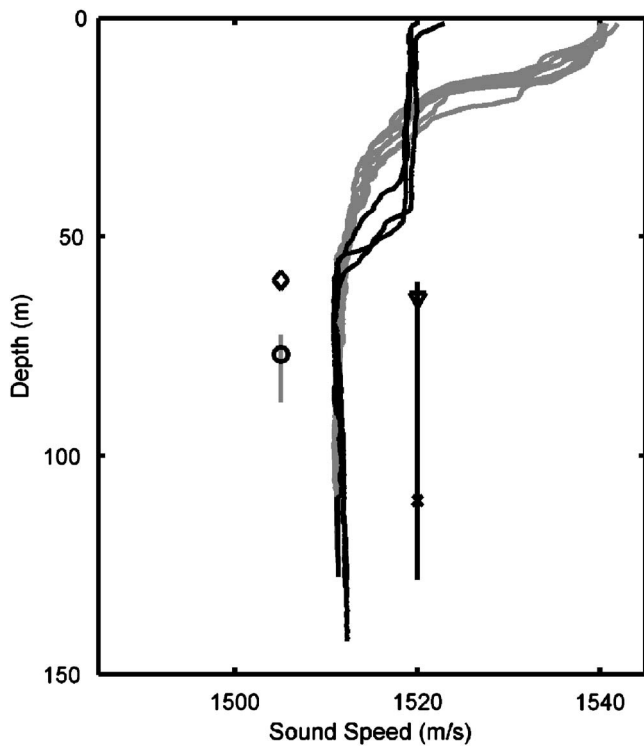


FIG. 2. Sound speed versus depth data deduced from expendable bathythermograph casts taken during the experiments. Gray lines indicate Boundary 03 data; black lines show BASE 04 data. Vertical lines show depths covered by vertical receiver arrays. Diamond shows Boundary 03 source depth, circle shows depths from which single-hydrophone data were taken for Boundary 03 data. Triangle shows BASE 04 source depth; cross shows depths from which single-hydrophone data were taken for Boundary 03 data.

area is silt (Max *et al.*, 1992) but an area of exposed sediment has been identified (Osler and Algan, 1999) and this has been tentatively associated with an area of enhanced reverberation measured in the frequency band of interest here (Prior, 2005). Figure 3 shows the area of enhanced reverberation with the BASE 04 track overlaid. The track is shown to begin in the east, inside the scattering region and to leave it about 15 km from the VLA, marked by the triangle in Fig. 3. In Sec. IVB, we will show that the edge of this region of enhanced reverberation is associated with a change in seabed type, as indicated by changes in pulse shape. The VLA was placed along the line of the track followed, about 3 km beyond the start point.

The pulse shape received on a single element of the VLA for a transmission made from a distance of 19.4 km is shown in Fig. 4 in the form of received intensity in dB *re* 1 μ Pa versus time. The pulse is shown to be made up of a

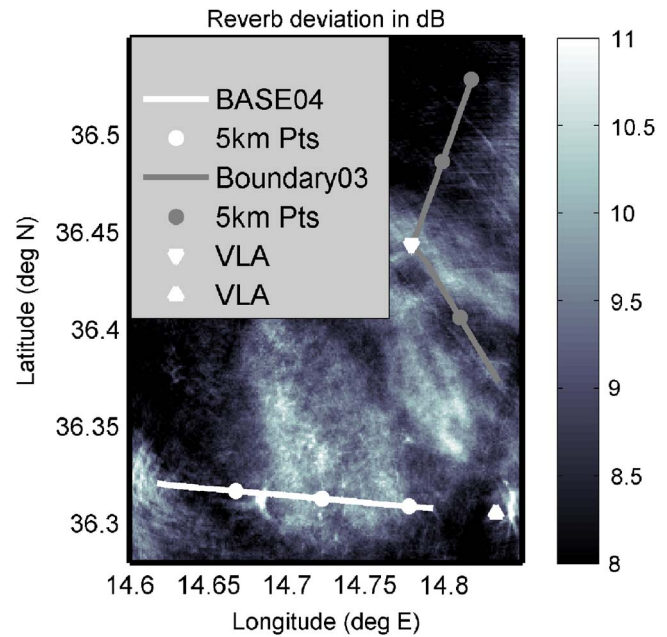


FIG. 3. (Color online) Map showing experimental tracks overlaid on shaded contours of reverberation deviation from an arbitrary reference level. Lighter areas indicate stronger local scattering.

large-amplitude leading edge, followed by a smooth fall-off in the decibel intensity level. The black line in the figure shows a least-square fit to the data of the form given in Eq. (3), i.e., intensity after the peak arrival comprises an exponentially decaying term, divided by the square root of the time after the peak. The argument of the exponential term was estimated by the curve-fitting procedure.

The expanded image of the leading edge, also shown in Fig. 4, reveals that at very short times (~ 10 ms) before the peak, there is evidence of the crescendo of refracted paths predicted by the closed-form expressions for pulse shape (Harrison and Nielsen, 2007).

Figure 5 shows the measured pulse shape for the same transmission but after depth averaging was applied over the VLA. This averaging took the form of summation of the intensity on each hydrophone as a function of time, i.e., incoherent intensity averaging was performed. The fluctuations of intensity with time are shown to be smaller than for the single-phone case and this is a consequence of the averaging process. However, the general features of a spike on the leading edge, followed by a smooth decay of decibel intensity level, are still observed. The data are also shown to be well fitted by the expression given in Eq. (3), indicated by the

TABLE I. Source and receiver characteristics for the BASE 04 and Boundary 03 experiments.

	Source depth (m)	VLA depth coverage (m)	Single receiver depth (m)	No. source rings used	Source level (dB re: 1 μ Pa @ 1 m) ^a	Lower frequency (Hz)	Upper frequency (Hz)	Pulse length (s)
BASE 04	64	60.35-128.3	110.35	3	216	800	1400	1
Boundary 03	60	72.35-87.85	76.85	1	206	700	2000	1

^aValues for source level were averaged over transmitted signal frequency band.

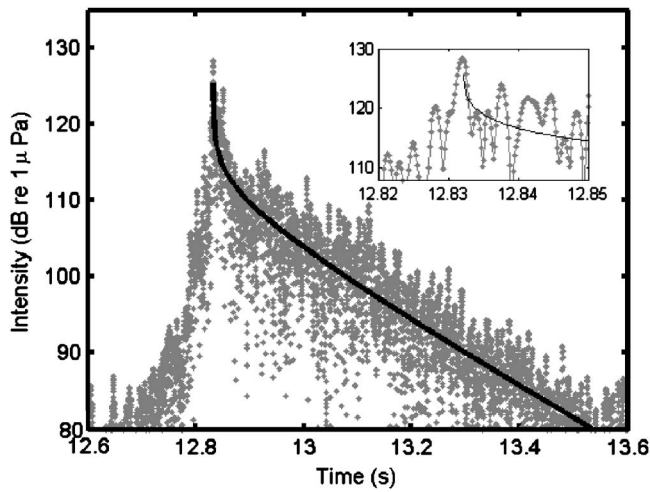


FIG. 4. BASE 04 received pulse shape in the form of intensity versus time for a single receiver located 19.4 km from the acoustic source. Thick black line in main panel shows the predicted pulse shape, fitted to the data. Inset panel shows close-up of arrivals in the vicinity of the intensity peak.

thick black line in Fig. 5. The expanded image of the leading edge of the pulse in Fig. 5 shows evidence of a crescendo before the maximum intensity arrival.

Figure 6 shows an expanded image of the leading edge of a depth-averaged pulse transmitted from a range of 3.2 km. Unlike the longer-range data in Fig. 5, there is no evidence of strong arrivals before the peak of the pulse. While this is consistent with the prediction that the time extent of the crescendo should increase with source-receiver range, the evidence contained in Figs. 5 and 6 cannot be said to provide anything more than slight confirmation of the closed-form expressions relevant to times before the peak intensity of the pulse.

C. Boundary 03

Sound speed profiles recorded by expendable bathythermograph casts during the Boundary 03 experiment are

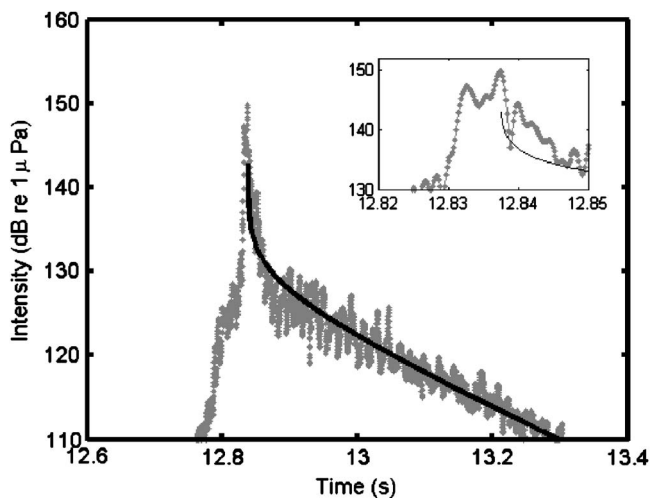


FIG. 5. BASE 04 received pulse shape in the form of depth-averaged intensity versus time. Receiver located 19.4 km from the acoustic source. Thick black line in main panel shows the predicted pulse shape, fitted to the data. Inset panel shows close-up of arrivals in the vicinity of the intensity peak.

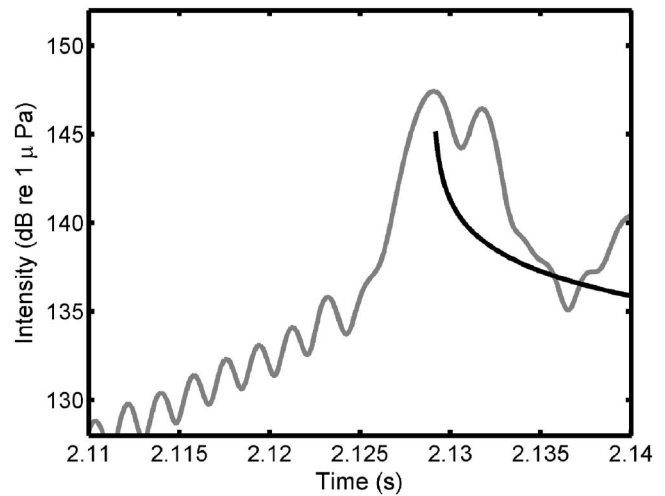


FIG. 6. BASE 04 received pulse shape in vicinity of the intensity peak in the form of depth-averaged intensity versus time. Receiver located 3.2 km from the acoustic source.

shown in Fig. 2. Maximum sound speed was observed at the surface and sound speed decreased steadily with increasing depth from this maximum value of around 1540 m/s to a value of around 1512 m/s at 50 m. Below 50 m, only a slight increase in sound speed with depth was observed.

Source and receiver were placed in the deep layer, and sensor depths and pulse properties are summarized in Table I. The vertical extent of the VLA is shown by the straight, black line in Fig. 2 and the depths of the sources and single-hydrophone receivers are shown by diamond and circular markers, respectively.

The track followed during the Boundary 03 experiment, shown in Fig. 1, was split into two legs. The northern leg followed a track that has previously been identified (Fallat *et al.*, 2005) as having a mud layer overlying a faster sediment below. The mud layer was around 10 m thick at the start of the track but this thickness reduced linearly along the track until, at the end of the track, near the VLA location, the mud layer vanished and the faster sediment was exposed. This boundary has also been associated (Prior, 2005) with the edge of the area of enhanced reverberation shown in Fig. 3.

The southern leg of the Boundary 03 track followed the path along which ambient noise measurements have been used (Harrison, 2004) to produce sub-bottom profiles. There is no evidence of a change of seabed type along this leg.

The pulse received on a single hydrophone for a transmission 9 km distant from the receiver, close to the northern end of the Boundary 03 run shown in Fig. 1, is shown in Fig. 7. The general shape of the pulse is similar to the shapes observed in the BASE 04 data, shown in Figs. 4 and 5, and the upper envelope of the pulse is well described by the form of Eq. (3), as shown by the black line in Fig. 7. The major difference between the data shown in Fig. 7 and those previously seen for the BASE 04 dataset is that the decay of intensity occurs at a much higher rate in the Boundary 03 data. This rapid fall-off of intensity was not limited to the single hydrophone whose data are shown in Fig. 7, and the same behavior was observed in depth-averaged intensity for the same pulse, as shown in Fig. 8.

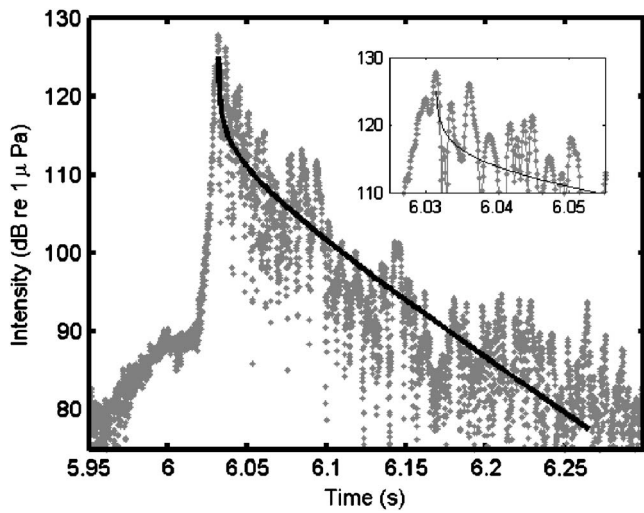


FIG. 7. Boundary 03 received pulse shape in the form of intensity versus time for a transmission made 9 km north of a single-hydrophone receiver. Thick black line in main panel shows the predicted pulse shape, fitted to the data. Inset panel shows close-up of arrivals in the vicinity of the intensity peak.

However, the rapid fall-off was not observed for the southern leg of the Boundary 03 track, as shown in Fig. 9. The figure shows the depth-averaged pulse shape received following a transmission, also 9 km distant from the VLA, and, although the shape of the pulse still follows the form predicted by Eq. (3), the rate of intensity decay is significantly less than for the pulse transmitted at a similar distance north of the VLA.

None of the data in Figs. 7 and 8, or 9 show significant evidence of any crescendo preceding the maximum intensity arrival. This is consistent with the prediction of the closed-form expressions of pulse shape (Harrison and Nielsen, 2007), that the width of the crescendo is proportional to the sound speed difference between source and receiver depths, since the source and receiver were placed at depths at which

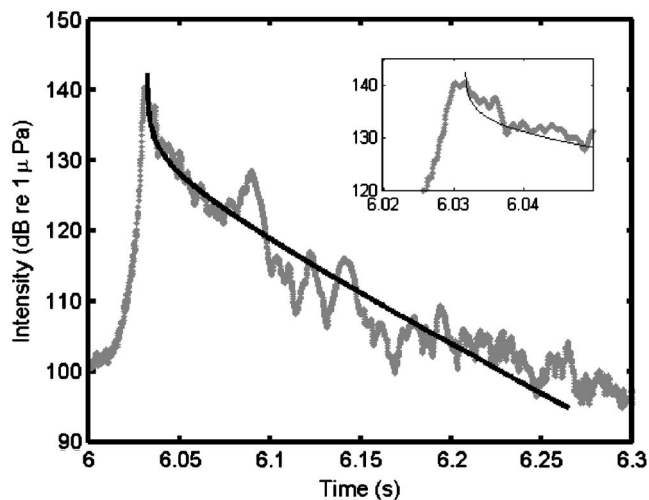


FIG. 8. Boundary 03 received pulse shape in the form of depth-averaged intensity versus time for a transmission made 9 km north of the receiver array. Thick black line in main panel shows the predicted pulse shape, fitted to the data. Inset panel shows close-up of arrivals in the vicinity of the intensity peak.

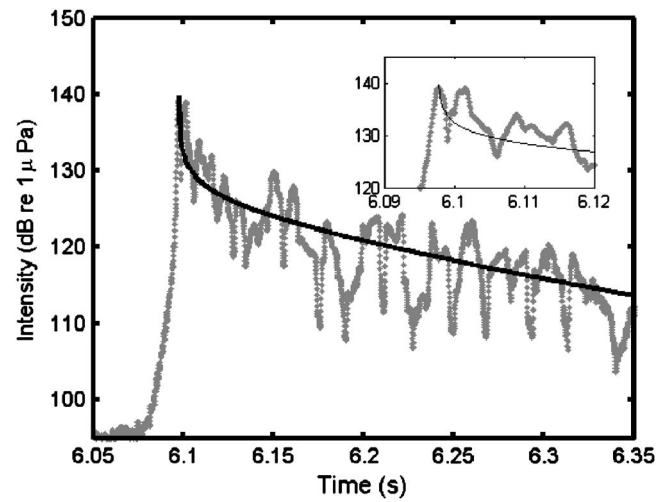


FIG. 9. Boundary 03 received pulse shape in the form of depth-averaged intensity versus time for a transmission made 9 km south-east of the receiver array. Thick black line in main panel shows the predicted pulse shape, fitted to the data. Inset panel shows close-up of arrivals in the vicinity of the intensity peak.

the sound speed differed by only 0.4 m/s. The source and receiver in the BASE 04 data had a sound speed contrast of 0.8 m/s.

D. Summary of comparison of predicted and measured pulse shapes

The acoustic data gathered during the BASE 04 and Boundary 03 experiments allowed examination of the shapes of pulses received after one-way propagation through a shallow water environment over ranges equivalent to many water depths. Good agreement was observed between the measured data and the prediction that pulses would have a trailing edge in which intensity decayed smoothly with time. This good agreement is considered to be supportive of the theoretical predictions made by Harrison and Nielsen (2007).

Only slight evidence was observed of the presence of a crescendo before the peak intensity arrival in the BASE 04 dataset. The Boundary 03 dataset showed no such evidence but this absence is consistent with the closed-form expressions, given the low sound-speed contrast observed between source and receiver depths in the Boundary 03 experiment.

The good agreement between the shape of the trailing edges of measured pulses and the predictions made by the closed-form expressions extends the possibility of using those expressions and data to estimate seabed reflection properties. This estimation process is now described.

IV. ESTIMATION OF SEABED REFLECTION PROPERTIES FROM PULSE SHAPE

A. General description of procedure

Estimates of the seabed near-grazing reflection-loss gradient were made by calculating the function given in Eq. (6). That is, the rate of change of logarithmic intensity with respect to time was calculated, after multiplication by the reduced time variable, τ . The closed-form expression for the pulse shape, Eq. (3), shows that this “rate of decay” is pro-

TABLE II. Geoacoustic parameters for typical sediment types, taken from [Hamilton \(1980\)](#)

Type	Speed (m/s)	Specific gravity	Absorption (dB/λ)	α (Nps/rad)	θ_c (deg)	$r_0/H = 2/(\alpha\theta_c^2)$
Coarse sand	1836	2.03	0.6	0.31	35.22	17.02
Very fine sand	1709	1.88	0.5	0.43	28.63	16.59
Silt	1615	1.74	0.3	0.65	21.75	21.4
Silty clay	1517	1.48	0.1	3.18	8.59	27.9

portional to the seabed near-grazing reflection-loss gradient, α_L . If the bracketed term $[1+(Hc)/(2c_L)]$ in Eq. (6) is assumed to be effectively unity, α_L can be estimated by simply multiplying the rate of decay by the water depth, H , and dividing by the seawater sound speed at the seabed, c_L .

The rate of decay was calculated from the measured data by an automatic algorithm that identified the time of peak intensity arrival. This time was then used to scale out the reduced time, τ . The natural logarithm of the resulting scaled intensity was then calculated and a least-squares-fit was made to this data to extract the rate of decay. The thick, black lines shown in Figs. 4–9 were produced using rates of decay calculated in this way and the figures show that the approach produced good descriptions of the shape of the pulses' trailing edges.

The estimates of seabed acoustic properties produced in this manner are now described for the two experimental datasets.

B. Base 04

Closed-form expressions for pulse shape predict that the slope of the trailing edge of the pulse should not change as source-receiver separation increases beyond a critical range, r_0 , this range being determined by the seabed type ([Harrison, 2003](#)) but being between 15 and 30 water depths for most seabed types. Some typical values are given in Table II. This behavior has been previously observed ([Prior and Harrison, 2004](#)) in the Malta Plateau in a situation where the seawater sound speed profile was effectively isovelocity.

Figure 10 shows the slopes of the trailing edge of pulses measured during the BASE 04 experiment, plotted as a function of source-receiver range. The slope is quantified in two ways: as intensity fall-off in dB per second and as inferred reflection-loss gradient. This latter measure corresponds to the α_L variable in the closed-form expressions and is determined by multiplying the intensity fall-off by the water depth and then dividing by the seawater sound speed at the seabed. Figure 10 shows values derived from single-hydrophone measurements and from data averaged over the entire extent of the VLA.

The data shown in Fig. 10, for ranges less than 15 km, show the behavior predicted by the closed-form expressions, i.e., the slope of the trailing edge of the pulse did not vary with range. The mean value of the intensity fall-off for the single-hydrophone data for ranges less than 15 km is 38 dB/s with a standard deviation of ± 5 dB/s about this mean. The same values for the depth-averaged data are 37 dB/s and ± 3 dB/s. These two values indicate that, while depth averaging reduced the variability of the slopes of the

trailing edge of the pulse, the mean value was effectively the same as that observed for the single-hydrophone data. The reflection-loss gradients corresponding to the measured intensity decays are 0.75 ± 0.09 nepers per radian for the single-hydrophone data and 0.74 ± 0.05 nepers per radian for the depth-averaged data. Inspection of Table II indicates that these values lie between those for silt and silty clay. The seabed in the experimental area has previously been shown to be silt ([Prior and Harrison, 2004](#)) and the values are therefore physically reasonable.

There is no evidence of the previously observed ([Prior and Harrison, 2004](#)) change of trailing-edge slope at very short ranges. However, the minimum range in the dataset (3.5 km) corresponds to 27 water depths and this is not significantly less than the critical limit of 21 water depths predicted for silt, as quoted in Table II. Consequently, the BASE 04 data did not cover ranges short enough to identify the critical range r_0 . Estimates of this range can be used ([Prior and Harrison, 2004](#)) to estimate seabed critical angle and hence sound speed, density, and attenuation, but this procedure could not be followed with the data gathered during the BASE 04 run considered here.

For ranges greater than 14 km the single-hydrophone and depth-averaged data diverge from the behavior exhibited at shorter ranges. The single-hydrophone intensity-gradient data for ranges greater than 14 km had a mean value of

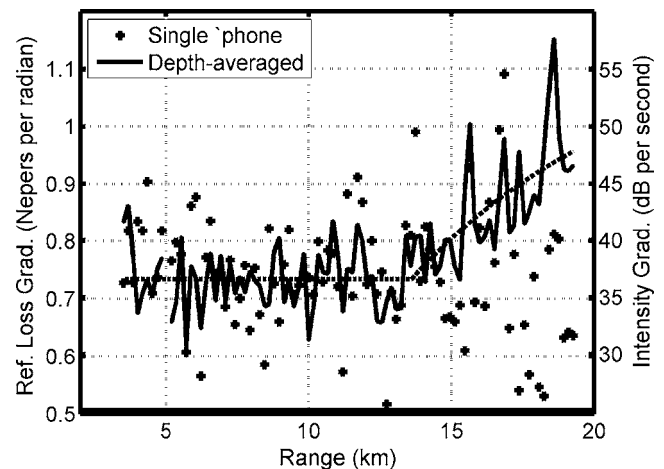


FIG. 10. Rate of decay of intensity deduced from measured pulse shape for BASE 04 dataset. Right-hand y axis gives rate in dB/s. Left-hand y-axis gives the deduced rate-of-change of seabed reflection loss versus angle at grazing incidence. Crosses show data taken from a single hydrophone, solid line shows data taken from intensity depth averaged over the whole array. Dashed line shows a parametric fit to the data calculated assuming that the measured value is an along-track-averaged measure of a range-dependent variable.

36 dB/s and standard deviation about that mean of ± 7 dB/s, indicating no significant change in mean gradient but an increase in variability. The depth-averaged data, however, showed a steady rise at longer range. Inspection of Fig. 3 shows that the 15 km point on the track (indicated by the westernmost circular marker on the BASE 04 track) is in the region where the track leaves the area of enhanced scattering that has been associated (Osler and Algan, 1999) with an area of exposed sediment. The sediment outside the exposed area has previously (Prior and Harrison, 2004) been shown to have a reflection-loss gradient higher than the mean value of 0.74 nepers per radian. It is therefore possible that the changes in behavior observed around 15 km may be explained by a change in the seabed type close to that location.

To investigate this possibility, it was hypothesized that the depth-averaged data represented an along-track-averaged measure of the seabed reflection-loss gradient, with the seabed having a reflection-loss gradient of α_1 from the start of the track to a range R_c , beyond which the seabed reflection-loss gradient changed to a value of α_2 . Under this hypothesis, the measured reflection-loss gradient would be given by

$$\alpha = \frac{1}{R} \int_0^R \alpha(r) dr = \begin{cases} \alpha_1, & R < R_c, \\ \frac{\alpha_1 R_c + \alpha_2 (R - R_c)}{R}, & R \geq R_c, \end{cases} \quad (8)$$

where α is the reflection-loss gradient inferred from the depth-averaged pulse-shape data and R is the range between source and receiver.

The depth-averaged data were fitted with curves of this type and a genetic algorithm (Houck *et al.*, 1995) used to determine values of α_1 , α_2 , and R_c that best matched the data in the sense of minimizing the root-mean-square mismatch between the data and the predictions of the expression given in Eq. (8). Search bounds for R_c were set at the minimum and maximum ranges of the experiment and the two reflection-loss gradients were searched for in bounds between 0 and 5 nepers per radian. The genetic algorithm used a population size of 80 and was run 1001 times so that the repeatability of the method's predictions could be investigated. The lowest mismatch was produced for R_c equal to 13.6 km with α_1 and α_2 equal to 0.734 and 1.49 nepers per radian, respectively.

The error bounds associated with the predicted values for the three parameters were estimated by identifying the range of values returned by the genetic algorithm for each parameter that lay within the best (i.e., lowest mismatch) 10% of the 1001 results. The possible ranges of parameter values thus estimated were $13.2 \leq R_c \leq 14.6$, $0.73 \leq \alpha_1 \leq 0.74$, and $1.4 \leq \alpha_2 \leq 1.8$. It should be noted that these parameter ranges are not unbiased estimates of the uncertainty associated with the estimation of R_c , α_1 , and α_2 (Dosso, 2003) and can only be considered as rough indications of the spread of possible values.

The dashed line in Fig. 10 shows the values of α predicted by Eq. (8), using the best-fit values. The good fit between data and Eq. (8) supports the hypothesis that the seabed type changed around 14 km along the BASE 04 track.

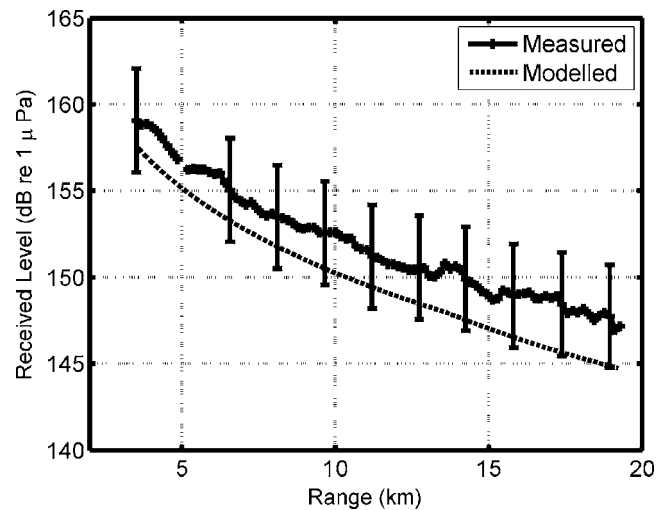


FIG. 11. Time-integrated intensity received as a function of source-receiver range for the BASE 04 dataset. Solid line shows depth-averaged measured data with error bars. Dashed line shows intensity predicted using a simple propagation expression with seabed description data derived from consideration of pulse shape.

Comparison of the genetic algorithm estimate for α_2 with the data in Table II suggests that the seabed inside the area of exposed sediment was more coarse-grained than that outside. This is consistent with the exposed sediment area causing enhanced reverberation, as shown in Fig. 3.

The values of reflection-loss gradient derived from pulse shape can be validated by comparison between total, i.e., time-summed, received intensity predicted using those values and the measured total intensity. Because reflection-loss gradient is determined only from the shape of the pulse, it is independent of the absolute received level, and the comparison between measured and predicted received levels is therefore an independent check, not a circular argument.

The pulse shapes shown in the preceding figures indicated that the majority of the paths between source and receiver were associated with the trailing edge of the pulse and were consequently only weakly affected by the sound speed profile. This allowed the use of a simple expression for intensity, I , measured by a receiver in the mode-stripping region, i.e., at ranges greater than the critical values given in Table II (Harrison, 2003):

$$I = S_l \sqrt{\frac{2\pi}{H\alpha r^3}} \exp[-2\beta r], \quad (9)$$

where S_l is the level of the source, expressed in linear, rather than logarithmic, units, β is the amplitude absorption coefficient (Thorp, 1967), and α is the range-averaged value of reflection-loss gradient, defined in Eq. (8). The absorption coefficient, β , is a function of frequency but a single value, calculated at the center frequency of the transmissions, was used. The intensity predicted by Eq. (9) and the depth-averaged intensity measured on the VLA are shown in Fig. 11. The error bars shown on the measured data are plotted only at ten ranges for clarity but are applicable to all measured values. The ± 3 dB uncertainty shown was calculated by combining an estimated ± 2 dB uncertainty in source level with an independent ± 1 dB uncertainty in receiver calibration.

tion, the two uncertainties being squared, added, and square-rooted before rounding up to an integer decibel value.

The simple formula for the received intensity, Eq. (9), is shown in Fig. 11 to predict levels within the uncertainty associated with the measurements over the entire dataset. The mean difference between measured and predicted levels was 2.1 dB, but the difference between predictions and measurements shows a slight tendency to increase with range. This agreement can be considered to indicate that the estimated values of reflection-loss gradient are reasonable descriptions of the seabed acoustic properties in the region of the BASE 04 experimental track.

It should be noted that the modeled intensity in Fig. 11 includes no dependence on seabed critical angle since the pulse shape data allowed no estimate of this angle to be made. However, noting that the average value of reflection-loss gradient at short range, 0.73, is close to the value associated (Hamilton, 1980; Prior and Harrison, 2004) with silt, the seabed critical angle might be estimated to be around 20 deg, based on a sound speed of 1615 m/s. The correction that must be made to the predicted intensity to allow for this critical angle takes the form (Harrison, 2003) of an error function by which the intensity is multiplied and whose importance decreases with range. For the BASE 04 environmental conditions, this correction is only greater than 1 dB for ranges less than 1.5 km. Any critical angle effects are therefore likely to be small.

In summary, the results from the BASE 04 experiment considered here support the predictions of pulse shape made by the closed-form expressions in that range-invariant slopes of pulses' trailing edges were observed in the BASE 04 data for ranges less than 14 km for both single-hydrophone and depth-averaged data. The divergence from this behavior at ranges greater than 14 km, most clearly observed in the depth-averaged data, is explicable in terms of a change in seabed type around 14 km along the track. The depth-averaged data allow the location and nature of this change to be quantified and the location of the hypothesized change in seabed type matches well with independent evidence. The seabed properties, summarized by a reflection-loss gradient, deduced from the pulse shape were shown to allow a simple expression for received intensity level to produce predictions that agreed with measurements, to within experimental uncertainties.

C. Boundary 03

The difference between the northern and southern legs of the Boundary 03 runs is illustrated in Fig. 12 where the rate of decay of intensity [after the square root term from Eq. (3) has been removed] is plotted as a function of source-VLA distance. The data in Fig. 12 split into two groups with mean values around 25 and 100 dB/s, respectively. The lower values correspond to measurements made along the southern leg of the track while the northern leg of the track yielded the higher values of intensity gradient shown. This is consistent with the presence of a change in seabed type along the two tracks, and the values of the inferred seabed reflection-loss gradient (shown on the left-hand y axis in Fig. 12), when

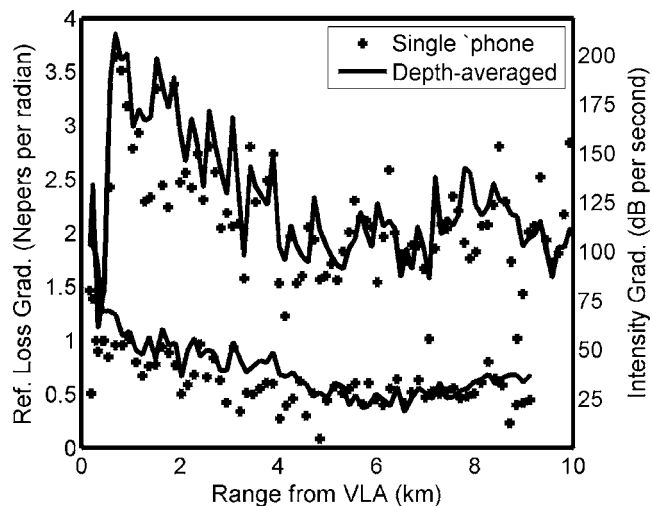


FIG. 12. Rate of decay of intensity deduced from measured pulse shape for Boundary 03 dataset. Right-hand y axis gives rate in dB/s. Left-hand y axis gives the deduced rate-of-change of seabed reflection loss versus angle at grazing incidence. Crosses show data taken from a single hydrophone for transmissions made to the north of the receiver. Circles show data taken from a single hydrophone for transmissions made to the south of the receiver. Solid line shows data taken from intensity depth averaged over the whole array for all transmissions.

compared with the data in Table II, suggest that the seabed to the north was more finely grained than that present along the southern leg. This is consistent with the previous observation (Siderius *et al.*, 2002) of a mud layer along the northern track whose thickness decreased as the VLA position was approached. Within 3 km of the VLA position, the mud layer ended, revealing a layer that has previously (Siderius *et al.*, 2002) been identified as having a higher sound speed than the mud. This change from finer-grained sediments to the north with coarser-grained sediments along the southern leg is consistent with the increase in reverberation observed at the VLA location, shown in Fig. 3.

The short-range ($r < 3$ km) data in Fig. 12 show divergence from the values to which the long-range data converge. This behavior has previously been observed (Prior and Harrison, 2004) for short-range measurements of this kind and is associated with the presence of sound at angles greater than the seabed critical angle. If the range at which the divergence occurs is estimated, then the critical angle of the seabed can also be estimated, (Prior and Harrison, 2004).

This estimation was performed by treating the northern and southern legs of the data separately. For each leg, a 1001-point range variable was produced with values linearly increasing from zero to the maximum range at which a measurement was made within the leg. For each value, r_i , of this range variable, a "score" was calculated to be the number of data points at ranges less than r_i whose values were more than one standard deviation away from the mean value, calculated for measurements made at ranges greater than r_i . The divergence range, r_0 , was then taken to be the range beyond which this score achieved a steady value. The results of this process are shown for the two legs in Figs. 13 and 14.

Figure 13 shows, as a function of source-receiver range, the inferred reflection-loss-gradient values for the northern leg of the Boundary 03 dataset and, in gray, the relative value

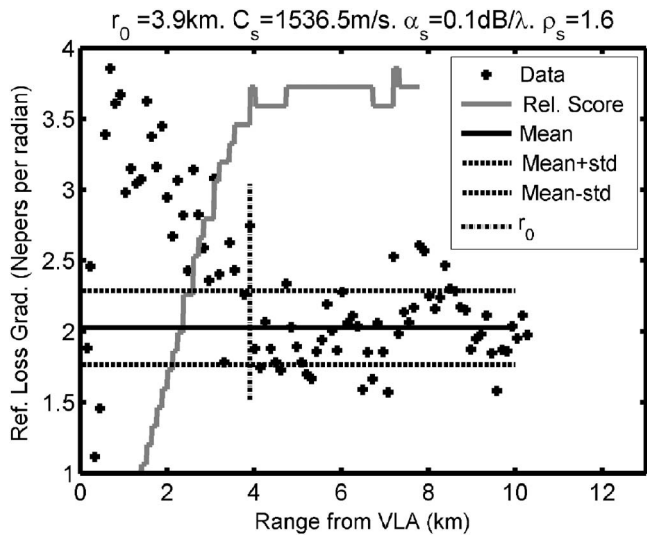


FIG. 13. Reflection-loss gradient data deduced from depth-averaged pulse shape for the northern leg of Boundary 03 dataset, plotted as a function of source-receiver range (crosses). Critical range r_0 deduced as being the shortest range at which the relative score variable (gray line) stopped changing with increasing range. Solid horizontal line shows the mean value of data from ranges greater than r_0 . Dashed lines show mean value \pm one standard deviation about the mean.

of the score used to estimate the location of the range below which the reflection-loss-gradient data diverged from their long-range behavior. A leveling-out of the score is observed at 3.9 km and this value was used to estimate the mean, long-range reflection-loss gradient and the deviation about this mean. These were found to be 2.0 nepers per radian for the mean and 0.3 nepers per radian for the standard deviation and are shown in Fig. 13 by the solid and dashed black lines.

The mean value and the estimate for r_0 were used to estimate the critical angle of the seabed using the expression (Harrison, 2003)

$$\theta_c = \sqrt{\frac{2H}{\alpha r_0}}, \quad (10)$$

where θ_c is the critical angle. The critical angle was estimated in this way to be 10.2 deg, corresponding to a seabed sound speed of 1536.5 m/s. This sound-speed value was used to estimate the seabed specific gravity via Hamilton's regression curves (Hamilton, 1978) relating density to sound speed for silt sediments. The resulting value, 1.6, was then combined with the estimated values of sound speed and reflection-loss gradient and [using Weston's expression (Weston, 1971) relating these properties for a fluid, semi-infinite seabed] the seabed attenuation coefficient was estimated to be 0.1 dB per wavelength.

While these values for sound speed, density, and attenuation lie within physical limits for silt sediments, previous surveys (Siderius *et al.*, 2002; Fallat *et al.*, 2005) have suggested that the surficial sediment along the track is more fine-grained with consequently lower sound speed. However, those surveys have also shown that there is significant layering in the seabed along the track and the fine-grained upper sediment lies on top of coarser-grained sediment with higher sound speed. The sediment properties estimated here using consideration of pulse shape therefore probably represent an "effective" sediment whose properties gave near-grazing reflection-loss gradient similar to that obtained from the actual, layered seabed.

Figure 14 shows the inferred reflection-loss-gradient data for the southern leg of the Boundary 03 run, plotted as a function of range, along with the score used to calculate the range at which the data converged to their stable value. The reflection-loss-gradient was found to have a mean of 0.54 nepers per radian with a standard deviation about this mean of 0.1 nepers per radian. The divergence range, r_0 , was estimated to be 4.3 km and this gave estimates for sediment properties of 1598.4 m/s, 0.17 dB per wavelength, and a specific gravity of 1.6. These values are within the range quoted in Table II for silty sediments and are reasonably close to values estimated using matched-field inversion (Siderius *et al.*, 2002) in the vicinity of the experiment (1554.3 m/s, 0.1 dB per wavelength, and 1.8).

While the agreement with previous, matched-field inversion results is encouraging, it is more important that the estimates of seabed properties should allow a good prediction of in-water propagation in the area. To investigate this, the values for reflection-loss gradient were used in conjunction with Eq. (9) to estimate the received level at the VLA. These estimates were compared with the measured data in the same way as for the BASE 04 dataset, previously described. For the northern leg, the water depth varied along the track from 105 m at the start of the track to 128 m at the VLA location. This range-dependent bathymetry was included in the estimates using the concept of an effective depth (Harrison, 2003; Weston and Tindle, 1979) that allows propagation between two points in an environment with a range-dependent bathymetry, $H(r)$, to be predicted using a single, effective water depth, H_{eff} , given by

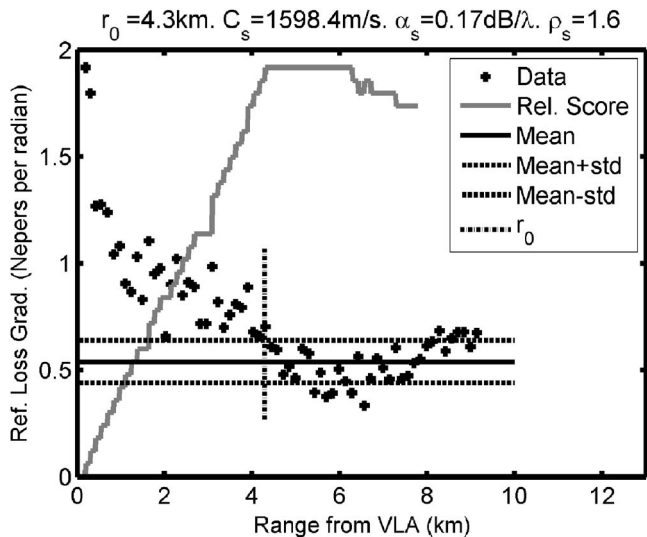


FIG. 14. Reflection-loss gradient data deduced from depth-averaged pulse shape for the southern leg of Boundary 03 dataset, plotted as a function of source-receiver range (crosses). Critical range r_0 deduced as being the shortest range at which the relative score variable (gray line) stopped changing with increasing range. Solid horizontal line shows the mean value of data from ranges greater than r_0 . Dashed lines show mean value \pm one standard deviation about the mean.

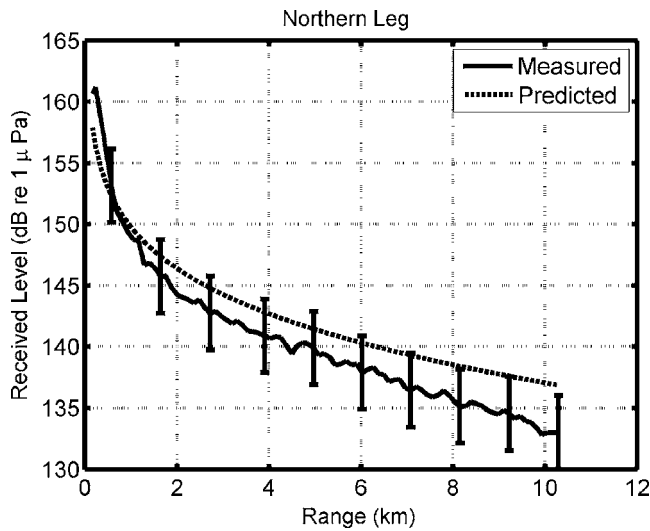


FIG. 15. Time-integrated intensity received as a function of source-receiver range for the northern leg of Boundary 03 dataset. Solid line shows depth-averaged measured data with error bars. Dashed line shows intensity predicted using a simple propagation expression with seabed description data derived from consideration of pulse shape.

$$H_{eff} = \frac{H_r^2 H_s^2}{r} \int_0^r \frac{dr'}{H^3(r')}, \quad (11)$$

where the water depth at the source is H_s and that at the receiver is H_r .

For the southern leg of the run, the water depth varied by less than 3 m over the 10 km extent of the track and the environment was considered range independent.

The received levels predicted and measured for the northern leg are shown in Fig. 15 with error bars on the measured data indicating the uncertainty associated with the measurements. Measured data were averaged over the depth of the array. The predicted levels agreed with the measurements to within the error bars for the majority of the ranges covered, but the difference between the two curves increased with range. Agreement between the measurements and predictions can still be said to be satisfactory since the mean difference between measured and modeled data was 1.9 dB.

Figure 16 shows measured and predicted levels for the southern leg of the Boundary 03 track and agreement between the two curves is similar to the northern leg, with the predictions lying within the error bars on the measurements over the entire 9 km extent of the track. The mean difference between measured and predicted data was found to be 1.7 dB. Also shown in Fig. 16 are some points of the data measured in the northern leg, shifted by $5 \log_{10}(2.0/0.54)$, i.e., shifted by the amount predicted by Eq. (9) to be the consequence of the decrease in reflection-loss gradient between the two legs from 2.0 to 0.54 nepers per radian. The shifted, northern data points are shown in Fig. 16 to agree closely with the southern data points measured at the same ranges. This indicates that the impact on received levels of the change in seabed type was well predicted by Eq. (9) and this is a further indication that the reflection-loss gradient, as deduced from the shape of the received pulse, is a useful predictor of the absolute received level.

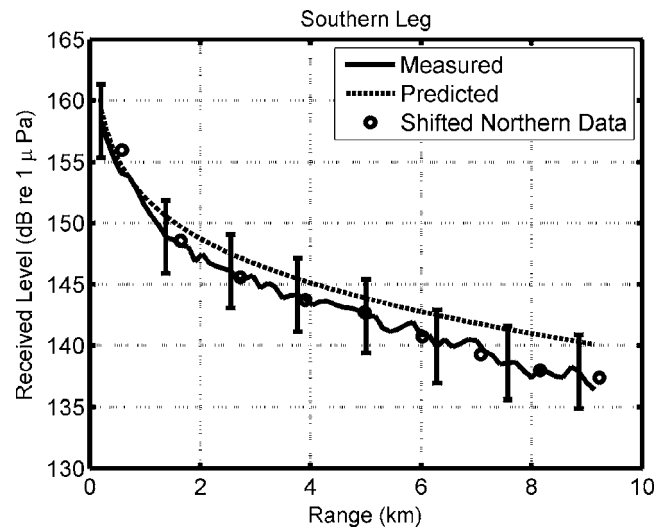


FIG. 16. Time-integrated intensity received as a function of source-receiver range for the southern leg of Boundary 03 dataset. Solid line shows depth-averaged measured data with error bars. Dashed line shows intensity predicted using a simple propagation expression with seabed description data derived from consideration of pulse shape. Circles show data points from the northern leg shifted by an amount (correctly) predicted by closed-form expressions to account for the change in seabed type.

In summary, the data from the Boundary 03 experiment showed that the previously observed change in seabed type at the VLA location was associated with a noticeable change in received pulse shape, as quantified by the slope of the pulses' trailing edge. Values for this slope were used to estimate seabed acoustic properties and these were shown to be comparable to values previously deduced using full, matched-field inversion in the same area. The seabed properties derived from consideration of pulse shape in the Boundary 03 dataset were also shown to allow the prediction of total received intensities at the VLA that agreed with independent measurements, to within experimental uncertainties.

V. CONCLUSIONS

It was shown that the pulse shape measured on a receiver many water depths from an acoustic source was well described by closed-form expressions proposed by Harrison and Nielsen (2007) for times after the arrival of the peak intensity. This shape, as was predicted by the closed-form expressions, was shown to be insensitive to the seawater sound-speed profile.

The rate at which the pulse intensity decayed, in accordance with the theoretical prediction, was used to provide an estimate of the angular gradient of near-grazing reflection loss at the seabed by fitting curves to measured pulse shapes. This method, previously demonstrated in the absence of a significant sound-speed profile, was shown to work even in the presence of vertical variations in sound speed over the water column of around 30 m/s.

The resulting seabed descriptions were shown to be consistent with existing descriptions of the seabed in the experimental area, taken from the literature. Furthermore, it was shown that changes in seabed type along the experimental tracks, previously observed in the study area, were also iden-

tifiable from changes in measured pulse shapes. Thus, not only were mean seabed properties identifiable from the measurements of pulse shape, but the locations of horizontal changes in seabed type were identified and the sizes of these changes were quantified.

The validity of the seabed descriptions derived from inspection of the pulse shape was further demonstrated by comparison between total, time-integrated received levels and the equivalent levels predicted by closed-form expressions. Agreement was found between predictions and measurements to within experimental uncertainties associated with the measurements.

Consideration of the slope of the trailing edges of pulses measured in shallow water therefore provides a possible method for rapidly estimating seabed reflection properties close to grazing incidence, i.e., in the angular regime most important for long-range propagation. The estimated properties are inherently averaged over the entire range from source to receiver and frequency averaged over the entire band of the transmissions. Such averaging is an inevitable consequence of the use of frequency-spread signals in real ocean environments. The averaged values should, however, be useful in all but rare cases which show extremely rapid variation of seabed properties with either acoustic frequency or along-track range. The method can be implemented with a single-hydrophone receiver, such as a standard sonobuoy, and has no need for numerical inversion of the acoustic data. An uncalibrated receiver could be used because of the method's sensitivity to the shape of the pulse, rather than its absolute amplitude. The method is therefore complementary to standard geophysical inversion techniques and could be realized in practical sonar systems without the need for specialist arrays or intensive processing.

ACKNOWLEDGMENTS

We thank the Captain and crew of the *RV Alliance* and the technicians of NURC's engineering department for data collection. The efforts of Georgios Haralabus, the scientist-in-charge of the BASE 04 cruise, are also gratefully acknowledged.

Collins, M. D., Kuperman, W. A., and Schmidt, H. (1992). "Nonlinear inversion for ocean bottom properties," *J. Acoust. Soc. Am.* **92**, 2770–2783.
 Dosso, S. E. (2003). "Quantifying uncertainty in geoaoustic inversion. I. A fast Gibbs sampler approach," *J. Acoust. Soc. Am.* **111**, 129–142.
 Fallat, M. R., Nielsen, P. L., Dosso, S. E., and Siderius, M. (2005). "Geoacoustic characterization of a range-dependent ocean environment using towed array data," *IEEE J. Ocean. Eng.* **30**, 198–206.
 Ferla, C., and Jensen, F. B. (2002). "Are current environmental databases

adequate for sonar predictions in shallow water?" in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. Pace and F. Jensen (Lerici, Italy) (Kluwer Academic Publishers, Dordrecht, The Netherlands).

Gerstoft, P. (1994). "Inversion of seismo-acoustic data using genetic algorithms and *a posteriori* probability distributions," *J. Acoust. Soc. Am.* **95**, 770–782.
 Hamilton, E. L. (1978). "Sound velocity-density relations in seafloor sediments and rocks," *J. Acoust. Soc. Am.* **63**, 366–377.
 Hamilton, E. L. (1980). "Geoacoustic modeling of the sea floor," *J. Acoust. Soc. Am.* **68**, 1313–1340.
 Harrison, C. H. (2003). "Closed-form expressions for ocean reverberation and signal excess with mode stripping and Lambert's law," *J. Acoust. Soc. Am.* **114**, 2744–2756.
 Harrison, C. H. (2004). "Sub-bottom profiling using ocean ambient noise," *J. Acoust. Soc. Am.* **115**, 1505–1515.
 Harrison, C. H., and Nielsen, P. L. (2007). "Multipath pulse shapes in shallow water: Theory and simulation," *J. Acoust. Soc. Am.* **121**, 1362–1373.
 Hermand, J.-P. (1999). "Broad-band geoaoustic inversion in shallow water from waveguide impulse response measurements on a single hydrophone: Theory and experimental results," *IEEE J. Ocean. Eng.* **21**, 324–346.
 Houck, C., Joines, J., and Kay, M. (1995). "A Genetic Algorithm for Function Optimization: A Matlab Implementation," North Carolina State University, Raleigh, NC, Rep. No. NCSU-IE TR 95-09 (available from site location <http://www.ise.ncsu.edu/mirage/GAToolBox/gaot>).
 Kuperman, W. A., and Jensen, F. B. (1980). *Bottom-Interacting Ocean Acoustics* (Plenum, New York).
 Le Gac, J.-C., Asch, M., Stephan, Y., and Demoulin, X. (2003). "Geoacoustic inversion of broad-band acoustic data in shallow water on a single hydrophone," *IEEE J. Ocean. Eng.* **28**, 479–493.
 Max, M. D., Kristensen, A., and Michelozzi, E. (1992). "Small-scale plio-quaternary sequence stratigraphy and shallow geology of the west-central Malta plateau," in *Geological Development of the Sicilian-Tunisian Platform*, Unesco Reports in Marine Science, Vol. **58**, pp. 117–112.
 Osler, J., and Algan, O. (1999). "A high resolution seismic sequence analysis of the Malta Plateau," SACLANT Undersea Research Centre, La Spezia, Italy. Rep. No. SR-311 (obtainable through national distribution centers, see <http://www.nurc.nato.int/pubs/index.htm>).
 Prior, M. K. (2005). "A scatterer map for the Malta Plateau," *IEEE J. Ocean. Eng.* **30**, 676–690.
 Prior, M. K., and Harrison, C. H. (2004). "Estimation of seabed reflection loss properties from direct blast pulse shape (L)," *J. Acoust. Soc. Am.* **116**, 1341–1344.
 Richardson, W. J., Greene, C. R., Malme, C. I., and Thomson, D. H. (1995). *Marine Mammals And Noise* (Academic, San Diego).
 Siderius, M., Nielsen, P. L., and Gerstoft, P. (2002). "Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array," *J. Acoust. Soc. Am.* **112**, 1523–1535.
 Siderius, M., and Hermand, J.-P. (1999). "Yellow Shark Spring 1995: Inversion results from sparse broadband acoustic measurements over a highly range-dependent soft clay layer," *J. Acoust. Soc. Am.* **106**, 637–651.
 Smith, P. J. (1971). "The averaged impulse response of a shallow-water channel," *J. Acoust. Soc. Am.* **50**, 332–336.
 Thorp, W. H. (1967). "Analytic description of the low-frequency attenuation coefficient," *J. Acoust. Soc. Am.* **42**, 270.
 Weston, D. E. (1971). "Intensity-range relations in oceanographic acoustics," *J. Sound Vib.* **18**, 271–287.
 Weston, D. E., and Tindle, C. T. (1979). "Reflection loss and mode attenuation in a Pekeris model," *J. Acoust. Soc. Am.* **66**, 873–879.

Near-specular acoustic scattering from a buried submarine mud volcano^{a)}

Anthony L. Gerig^{b)} and Charles W. Holland

Applied Research Laboratory, Penn State University, State College, Pennsylvania 16804-0030

(Received 23 March 2007; revised 8 September 2007; accepted 10 September 2007)

Submarine mud volcanoes are objects that form on the seafloor due to the emission of gas and fluidized sediment from the Earth's interior. They vary widely in size, can be exposed or buried, and are of interest to the underwater acoustics community as potential sources of active sonar clutter. Coincident seismic reflection data and low frequency bistatic scattering data were gathered from one such buried mud volcano located in the Straits of Sicily. The bistatic data were generated using a pulsed piston source and a 64-element horizontal array, both towed over the top of the volcano. The purpose of this work was to appropriately model low frequency scattering from the volcano using the bistatic returns, seismic bathymetry, and knowledge of the general geoacoustic properties of the area's seabed to guide understanding and model development. Ray theory, with some approximations, was used to model acoustic propagation through overlying layers. Due to the volcano's size, scattering was modeled using geometric acoustics and a simple representation of volcano shape. Modeled bistatic data compared relatively well with experimental data, although some features remain unexplained. Results of an inversion for the volcano's reflection coefficient indicate that it may be acoustically softer than expected. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2793703]

PACS number(s): 43.30.Gv, 43.30.Zk [RCG]

Pages: 3280–3285

I. INTRODUCTION

The geology, geophysics, geochemistry, and ocean acoustics communities all study the composition, behavior, and development of mud volcanoes.^{1,2} They are of interest to the latter primarily as potential sources of active sonar clutter. In 2002, several mud volcanoes, both buried and proud, were discovered offshore Sicily on the Malta Plateau.³ Reverberation data indicate strong scattering associated with some of the buried volcanoes at ranges of 10 km.⁴

Accurate modeling of mud volcano scattering would make predicting and simulating scattering signatures for individual volcanoes possible, thus providing the information necessary to mitigate mud volcano clutter through improved sonar design and classification schemes. In addition, it would enhance the remote sensing of mud volcano geoacoustic properties, a valuable prospect for the geoscience communities, especially in regard to buried mud volcanoes.

Mud volcanoes form due to the rise of fluidized sediments and/or gas along a fault or on top of a seafloor-piercing shale diapir. They may occur in sedimentary areas with hydrocarbon generation at depth, originate from thick clay beds, and often bubble gas (mostly methane), and sometimes oil. They are known to occur in a variety of geologic settings including the abyssal parts of inland seas, active margins, continental slopes of passive margins and continental shelves.² Scattering from mud volcanoes in general may

arise from scattering from: (a) gas bubbles within the sediment, (b) gas bubbles and or particulates ejected into the water column, and/or (c) the mound itself, (see Ref. 5). Previous modeling from long-range reverberation at the buried mud volcano that will be the object of study here, indicates that the scattering was most likely due to the latter.⁴

The purpose of this work was to develop a more refined scattering model specifically for the buried mud volcanoes observed on the Malta Plateau through the use of seismic reflection, bistatic scattering, and geoacoustic measurements taken during different sea trials from a single volcano and its immediate environment at close range. It takes a bottom-up approach that allows the features seen in measured data to guide understanding and model development. In doing so, it identifies and incorporates the primary physical features and acoustic mechanisms significant to mud volcano scattering at low frequencies (less than 2 kHz).

Figure 1 indicates the precise position of the buried mud volcano used in this work, while Fig. 2 displays seismic reflection data recorded during the Boundary 2004 sea trial for that location. The ship track for Fig. 2 passes directly over the top of the volcano. It is evident that the volcano is buried approximately 10 m below the water–seafloor interface. The intervening layer appears to be divided into two separate layers by a thin layer about 6 m below the water–sediment interface. The mud volcano itself sits on a basement that exhibits complex layering, and appears to be on the order of 5 m high, 200 m wide and, based upon parallel tracks that are not shown, roughly circular in horizontal cross section.

The following section describes the bistatic scattering data taken from the mud volcano and how they were col-

^{a)}Portions of this work were presented as “Low-frequency scattering from a mud volcano located offshore Sicily” at the 151st Meeting of the Acoustical Society of America, Providence, RI, 2006.

^{b)}Presently at Viterbo University, 900 Viterbo Drive, La Crosse, Wisconsin 54601. Electronic mail: algerig@viterbo.edu

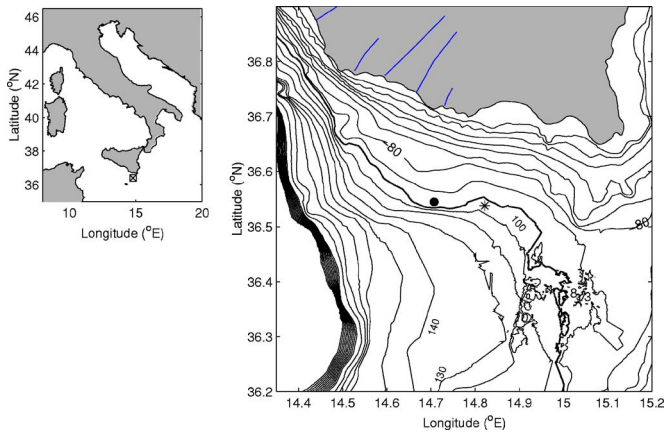


FIG. 1. (Color online) Map of the Malta Plateau, highlighting the locations of the buried mud volcano (solid circle) and the nearby geoacoustic inversion site (asterisk).

lected. The third section of the paper outlines a scattering model suggested by the data and justifies its selection. An explanation of the processing applied to the bistatic data to generate bottom loss values for the top of the mud volcano and the surrounding seabed is presented in the next section, followed by a discussion of the results and how they were used to determine input parameters to the scattering model that describe mud volcano and sediment layer composition. The final major section of the paper evaluates the scattering model by comparing results predicted by the scattering model in the time domain against the collected bistatic acoustic returns.

II. DATA COLLECTION

Figure 3 is a sketch of the data acquisition system that was used to collect bistatic scattering data from the buried mud volcano. The source was an EG&G Uniboomer that emitted a 1 kHz Ricker-like pulse every 10 s. A direct path measurement of the pulse and its corresponding spectrum are plotted in Fig. 4. Scattering was recorded by a 64-element horizontal linear array (HLA) at 4 m spacing. The data were digitized at 6000 Hz sampling rate and low-pass filtered at 1780 Hz with a seven-pole six-zero elliptic (70 dB per octave roll-off) antialias filter.⁶ Both were towed over the top of

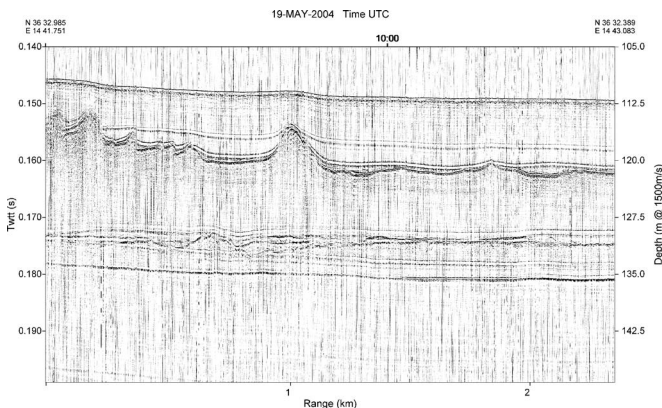


FIG. 2. Seismic reflection data for a track that passes directly over the top of the buried mud volcano.

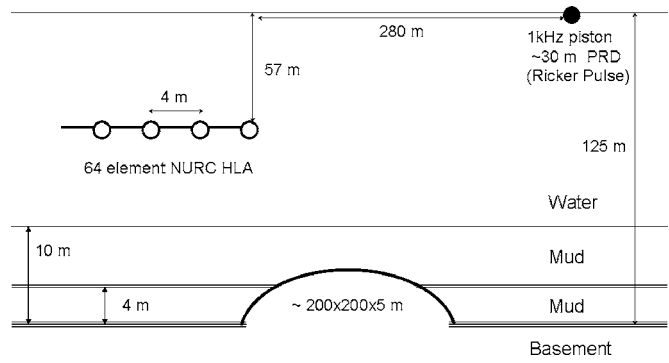


FIG. 3. Data collection geometry (PRD=pulse repetition distance).

the buried volcano at a speed of 3 m/s (6 knots), resulting in a displacement of approximately 30 m between pings. The source operated at the surface (towed at a nominal depth of 0.35 m), whereas the HLA was towed at a depth of 57 m. The separation of the source and the first element of the HLA was roughly 280 m, and the tilt of the array was generally less than 0.5° .

Based upon seismic reflection data (Fig. 2), the depth of the water column was determined to be approximately 115 m. Expendable bathythermograph data were collected every 50 min. Representative measurements are plotted in Fig. 5, and indicate that sound speeds were stable with time/

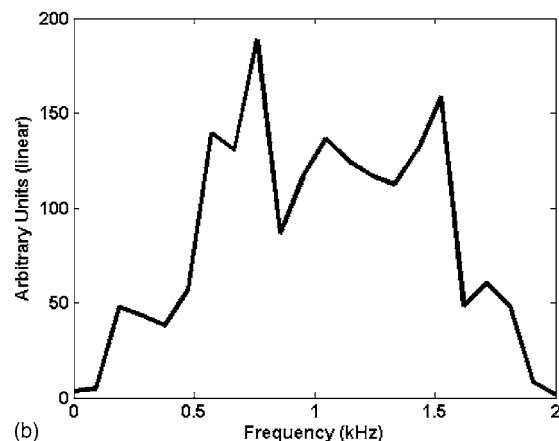
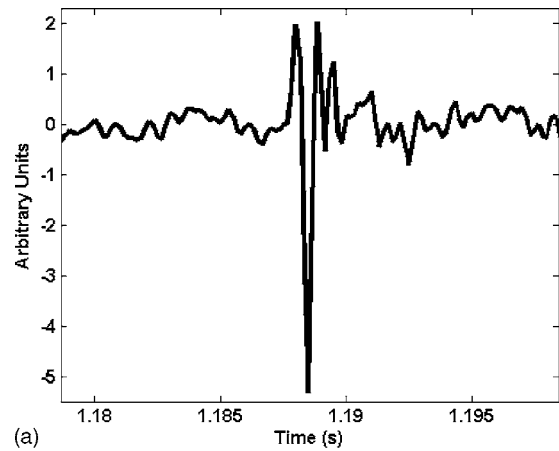


FIG. 4. Direct path measurement of (a) pulsed source and (b) corresponding spectrum.

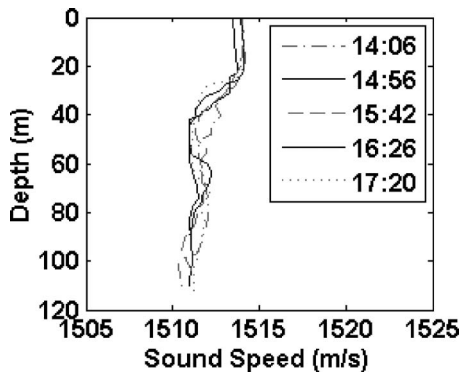


FIG. 5. XBT speed of sound data for the water column.

space and nearly isovelocity. No data on biologics (fish) were collected. The wind speed was roughly 12 m/s.

Figure 6 displays multiple time series for the first element of the HLA, positioned along the y axis according to the relative location of the hydrophone at the time of the source ping. This series is representative of the remaining elements in that, although signal levels vary from receiver to receiver, the overall structure remains unchanged. The first return indicated by the time series (at about 1.217 s) is for the water–sediment interface. The second (at about 1.223 s) was generated by the thin layer within the uppermost sediment packet. The following, stronger return (at about 1.224 s) was produced by the either the basement or the mud volcano, depending upon source/receiver location. Returns from the top of the mud volcano are on the order of those produced by the basement, while those from the sides are muted in comparison. The last return in the time series (1.229 s) is due to layering within the basement and is inconsequential to this study.

III. MODELING OF THE SCATTERED FIELD

The large size of the mud volcano in comparison to the acoustic wavelengths of the experiment suggested a model based upon a high frequency approximation. As a result, the approach was to start with a geometric acoustics approximation, and increase the complexity of the model as necessary to incorporate higher-order effects such as multiple scatter-

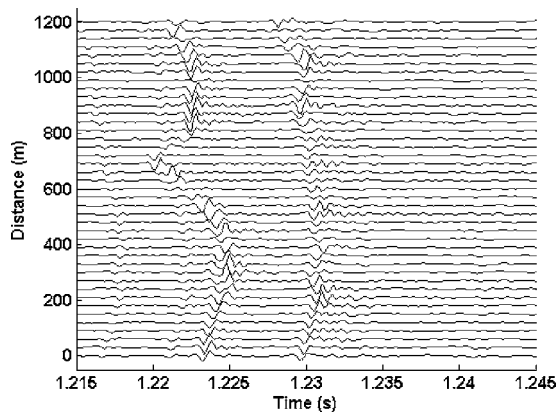


FIG. 6. Multiple time series for the first element of the HLA, positioned along the y axis according to the relative location of the hydrophone at the time of the source ping.

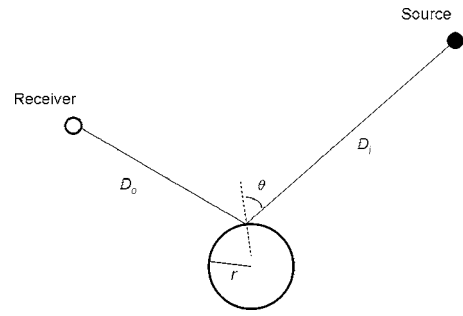


FIG. 7. Scattering geometry for Eq. (1).

ing, shadowing, diffraction, and surface roughness scattering. The mud volcano dimensions also suggested that modeled scattering would be insensitive to changes in shape, so the simple representation of a protruding sphere was chosen. The radius of the sphere and its degree of protrusion from the basement were determined by the mud volcano's approximate height and width as indicated by the seismic reflection data.

Mud volcano scattering was modeled based upon the following equation for the scattered field generated by a point source and sphere under the geometric acoustics approximation.⁷

$$P(k) = \pm \frac{\exp[ik(D_i + D_o)]}{kD_i} \left[\left(1 + \frac{D_o}{D_i} + \frac{2D_o}{r \cos \theta} \right) \times \left(1 + \frac{D_o}{D_i} + \frac{2D_o \cos \theta}{r} \right) \right]^{-1/2} \quad (1)$$

where k is the acoustic wave number of the incident field, D_i is the portion of the source/receiver eigenray between the source and sphere, D_o is the portion between the sphere and receiver, r is the radius of the sphere, and θ is the angle of incidence (see Fig. 7). The upper (lower) sign holds for a perfectly hard (soft) sphere.

Several modifications were made to this equation for the specific case under consideration. An approximate ray-based correction scheme was employed to account for the effects of the sediment layer over the mud volcano. This same scheme was used to correct for transmission through the covering layer and the curvature of the volcano in calculating bottom loss as described in the following section. In essence, Eq. (1) was multiplied by plane-wave transmission coefficients⁸ for the water/layer interface and thin layer, and a different eigenray was calculated to account for refraction at the water/layer interface. The plus/minus term was also replaced with a plane-wave reflection coefficient for the layer/volcano interface. The final result is shown in the following:

$$P(k) = T_{wl} T_{tl} R_v T_{tlo} T_{lw} \times \frac{\exp[ik(D_i + D_o)]}{kD_i} \left[\left(1 + \frac{D_o}{D_i} + \frac{2D_o}{r \cos \theta} \right) \times \left(1 + \frac{D_o}{D_i} + \frac{2D_o \cos \theta}{r} \right) \right]^{-1/2} \quad (2)$$

where T_{wl}/T_{lw} is the incoming/outgoing transmission coefficient for the water/layer interface, T_{tl}/T_{tlo} is the incoming/

TABLE I. Acoustic properties for the seabed strata, taken from Ref. 9.

Thickness (m)	Sound speed (m/s)	Density (g/cm ³)
Water column	1511	1.029
6	1490	1.32
0.2	1550	1.7
4	1490	1.6
0.8	1660	1.8
...	1615	1.7

outgoing transmission coefficient for the thin layer, R_v is the reflection coefficient for the mud volcano, D_i/D_o is the total length of the refracted incoming/outgoing portion of the eigenray, and θ is the angle of incidence for the refracted eigenray.

Several additional changes to the geometric acoustics model were considered but ultimately rejected. The effects of surface roughness were ignored for two reasons. First, the bistatic data indicated that scattering was primarily coherent at the wavelengths of interest, i.e., no significant distortion of the source waveform was evident with the exception of that due to layering. Second, surface roughness would need to be on the order of the mud volcano's height to significantly affect acoustic scattering at 1 kHz. In addition, neither multiple scattering nor shadowing was included in the model primarily because the mud volcano exhibited a low height to width ratio, but also because evidence of neither was apparent in the experimental time series. The effects of attenuation, although easily included for other cases, were neglected because geoacoustic measurements suggested that they should be minimal. Finally, no diffractive corrections were made to the geometric approximation since the representation of mud volcano shape possessed no sharp corners or edges.

IV. GEOACOUSTIC PROPERTY ESTIMATES VIA LAYER REFLECTIVITY

Knowledge of the geoacoustic properties of the seabed at the mud volcano site was required to properly account for transmission through the upper sediment layer in the scattering model [see Eq. (2)] and aid in determining the properties of the mud volcano itself. Although the geoacoustic properties of the seabed at this particular location were not measured, inversions were performed⁹ at a site 9 km away (see Fig. 1). Based upon examination of seismic reflection data from both sites, the stratigraphy for the two appeared to be nearly identical. It was therefore assumed that the inversion results from the nearby site could be applied at the mud volcano location, though this assumption was checked by comparing predicted bottom loss values ($-20 \log_{10}|R|$) based upon these results against measurements for both the water-sediment interface and the basement. Table I shows the transferred parameters from the inversion at the nearby site. Thicknesses for the covering layer were estimated based upon the seismic reflection data displayed in Fig. 2. The inversion assumed that a fluid model for all media was adequate.

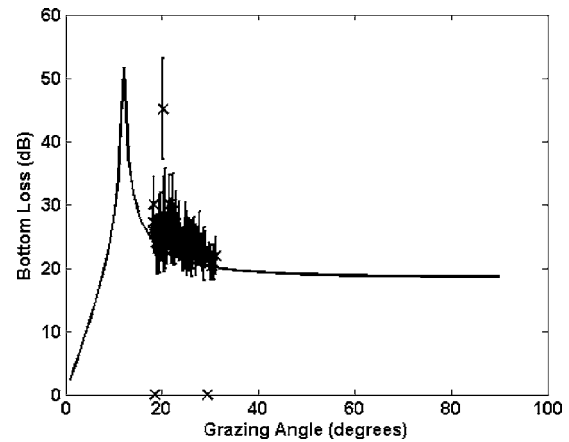


FIG. 8. Bottom loss measurements for the water-sediment interface at 1 kHz. The solid line represents theoretical values. Outliers are plotted along the abscissa.

Bottom loss measurements were also made for the top of the mud volcano, where returns from the volcano were the strongest and most consistent (see Fig. 6). These values placed limits on the unknown geoacoustic parameters for the mud volcano, which aided in determining both the possible composition of the mud volcano and a proper range of mud volcano reflection coefficient inputs to the scattering model [see Eq. (2)].

Bottom loss values were calculated following standard techniques, with a few, minor modifications to account for the curvature of the mud volcano and the effects of the layer covering both the mud volcano and basement. The same ray-based correction scheme used in modeling was employed for these calculations as well. Eq. (2) was essentially inverted for the reflection coefficient, where $r \rightarrow \infty$ for flat surfaces such as the water-sediment and basement interfaces. In addition, averaging was performed across one-third octave bands and refraction in the water column was neglected.

Figure 8 displays bottom loss measurements for the water-sediment interface at 1 kHz (the center frequency of the source). Results at each angle are for a particular HLA receiver. Markers (\times) and error bars represent mean and standard deviation values respectively for data taken over multiple pings as the source and HLA were towed. Mean values that were more than 10 dB down from those of the other receivers were considered outliers and are indicated by markers along the abscissa. The solid curve represents the results of a plane-wave theoretical calculation for the covering layer. Because returns from the top of the layer were used exclusively to measure bottom loss, the layer was treated in the calculation as a half space possessing the geoacoustic properties listed in Table I for the top of the layer. Although the measured results are consistently higher than their theoretical counterparts, the agreement between the two is reasonable. This suggests that the geoacoustic parameters at the nearby site are properly transferable to the mud volcano location.

Figure 9 shows an identical plot for the basement. The theoretical curve is for a single thin layer between two half spaces, a model that appears to work well despite the fact that the seismic reflection data indicate structural complexity

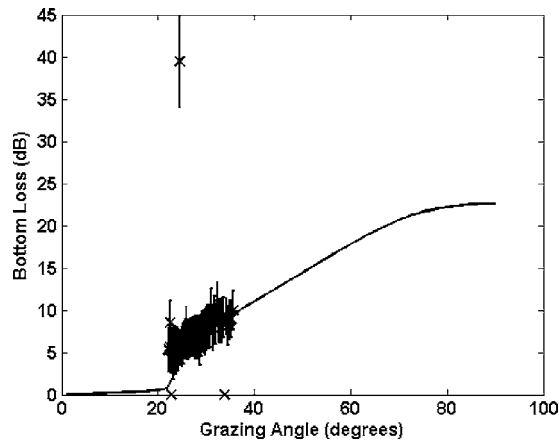


FIG. 9. Bottom loss measurements for the basement at 1 kHz. The solid line represents theoretical values. Outliers are plotted along the abscissa.

in the basement layering. Geoacoustic properties from Table I were again used in the calculation, where the properties listed for the bottom of the covering layer were employed for the upper half space. With the exception of values at lower grazing angles, the agreement between theory and experiment is quite good. This conformity not only strengthens the argument that the geoacoustic parameters for the nearby site are valid at the volcano site, but also suggests that the ray theory correction used to account for transmission through the sediment layer is adequate.

Figure 10 is a plot of the bottom loss measurements for the volcano. As data produced by the top of the volcano were used exclusively, only one measurement was available for each angle/receiver. All measurements and calculations assume that only the portion of the sediment layer above the internal thin layer covers the top of the volcano. Two theoretical curves are also displayed. Both represent possible models for the composition of the volcano, and are based upon observed similarities in the seismic reflection data of Fig. 2 between the volcano and surrounding basement. The top curve assumes that the volcano is entirely homogeneous and comprised of material identical to that of the basement under the region demonstrating complex layering (bottom

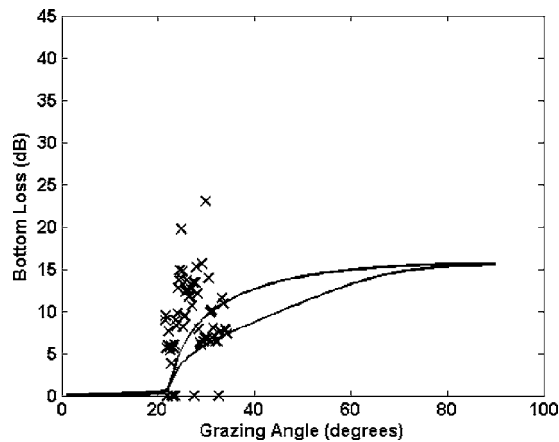


FIG. 10. Bottom loss measurements for the mud volcano at 1 kHz. The solid curves represent theoretical values, where the bottom curve assumes that the mud volcano is covered by a thin layer and the top curve does not. Outliers are plotted along the abscissa.

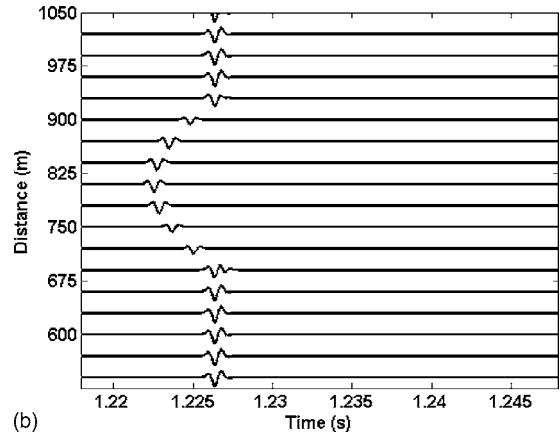
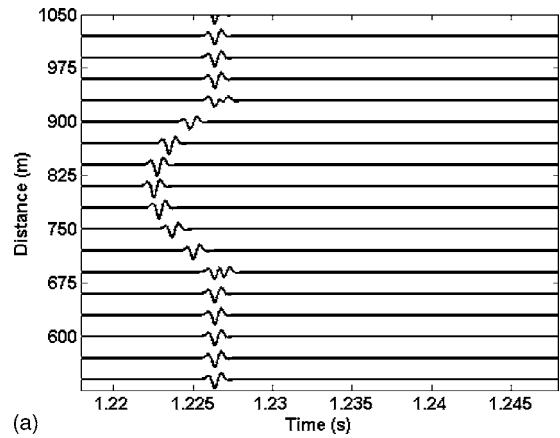


FIG. 11. Modeled time series for the first element of the HLA assuming (a) a hard covering layer and (b) no layer over the top of the mud volcano. Each of the multiple waveforms is positioned along the y axis according to the relative location of the hydrophone at the time of the source ping.

line of Table I). The bottom curve assumes that the complex layering of the basement (second to last line of Table I) extends to the volcano.

Although the comparison between theory and measurement is inconclusive, the results do support the idea that the mud volcano is materially similar to the surrounding basement. If this is true, the mud volcano must lack the hard, carbonate outer layer exhibited visually by nearby proud volcanoes [Ref. 5, Figs. 6(c) and 6(d)]. Although this is the most likely explanation for the data, it is also possible that the volcano possesses a layer that is either too thin to be observed acoustically at low frequencies, or that supports shear waves with a propagation speed that happens to be near that of compressional waves in the basement.

V. TIME SERIES MODEL-TO-DATA COMPARISONS

The scattering model outlined in Eq. (2) was evaluated by comparing predicted results in the time domain for an incoming Ricker pulse directly against the collected bistatic acoustic returns of Fig. 6. Fig. 11 displays two different modeled time series for the first receiver of the HLA. Each plot corresponds to one of the likely models of volcano composition suggested by the seismic reflection data. Although both

are designed for direct comparison with Fig. 6, they account for reflection from the volcano and the surrounding basement exclusively.

The modeled results conform to the actual data quite well, although there are a few noticeable differences. The modeled waveforms for the basement exhibit the same distortion of the incoming Ricker pulse evident in the data due to the presence of the complex top layer. However, identical distortion in the data for the top of the volcano is reflected in the second model alone because it exclusively includes layering. The amplitudes of the returns from the top of the volcano relative to those from the basement also compare better with the data for the second model. However, the relative amplitudes produced by the sides of the volcano appear to be too large, despite the fact that layering within the overlying sediment causes some muting in comparison to returns from the top. The first model better mimics the data in this regard, although it too does not precisely reproduce the virtually nonexistent side returns evident in the data. Adding a carbonate layer to the model only exacerbates the problem. It may be that the sides of the mud volcano are acoustically softer than the seismic reflection data would seem to indicate. Neither model reproduces the asymmetric returns from the sides of the volcano exhibited by the data, suggesting that a more complex representation of either volcano shape or composition is necessary if greater precision is desired. As amplitudes are relatively insensitive to shape when constrained to the dimensions of the volcano, adaptations to the latter would more likely improve model predictions.

VI. SUMMARY AND CONCLUSIONS

Bottom loss values were estimated for the top of the buried submarine mud volcano using a geometric acoustics model to correct for curvature. Although the results lack the precision necessary to invert for geoacoustic parameters, taken together with the available seismic reflection data they do suggest that the properties of the volcano are comparable to those of the surrounding seafloor. As a result, this particular mud volcano may lack the hard, carbonate top-layer possessed by the proud mud volcanoes that have been observed.

Although the scattering calculations generated by the model developed in this work approach an accurate representation of the bistatic acoustic data for the buried mud volcano, a few, minor features exhibited by the data remain unexplained. In particular, the acoustic returns generated by the sides of the mud volcano are not correctly reproduced by the calculations. However, the disagreement between theory and data is likely due to the use of imprecise reflection coefficients in the calculations, and should diminish with improved knowledge of volcano composition and structure. As a result, the general scattering model represented by Eq. (2) should adequately depict low-frequency scattering for mud

volcanoes like the one considered given the appropriate input parameters, and ought to be of use in predicting and mitigating clutter from such volcanoes.

Although the attempt to model acoustic scattering produced by this particular mud volcano has proven relatively successful, volcanoes can vary widely in size, shape and structure. As a result, continued work is necessary to examine the robustness of the geometric acoustics model. Smaller volcanoes and higher acoustic frequencies especially may require significant adaptations to the model, and could possibly demand a new approach altogether.

In addition, this work is primarily concerned with accurately modeling scattering at close range. However, mud volcanoes are also demonstrated sources of clutter at longer ranges. The geometric acoustics scattering model needs to be combined with established propagation models,¹⁰ and the results compared against long-range data in order to predict or simulate clutter at longer ranges with any confidence.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Office of Naval Research Ocean Acoustics program who supported the analysis and the NATO Undersea Research Centre (NURC) under whose auspices the experiments were conducted. The acoustic data were collected during the SCARAB98 experiment and the seismic reflection data were collected during the Boundary04 experiment—a joint program including NURC, the Naval Research Laboratory, Defence Research and Development Canada, and the Applied Research Laboratory—Penn State.

¹G. Etiope and A. V. Milkov, "A new estimate of global methane flux from onshore and shallow submarine mud volcanoes to the atmosphere," *Environ. Geol.*, **46**, 997–1002 (2004).

²A. V. Milkov, "Worldwide distribution of submarine mud volcanoes and associated gas hydrates," *Mar. Geol.*, **167**, 29–42 (2000).

³C. W. Holland, G. Etiope, A. V. Milkov, E. Michelozzi, and P. Favali, "Mud volcanoes discovered offshore sicily," *Mar. Geol.*, **199**, 1–6 (2003).

⁴C. W. Holland, A. L. Gerig, and P. Boni, "Acoustic clutter from buried submarine mud volcanoes," in *2nd Workshop on Experimental Acoustic Inversion Methods for Assessment of the Shallow Water Environment*, edited by A. Caiti, N. R. Chapman, J.-P. Hermand, and S. Jesus, Ischia, Italy June 2004 (Springer, New York, 2006), pp. 109–124.

⁵C. W. Holland, T. Weber, and G. Etiope, "Acoustic scattering from mud volcanoes and carbonate mounds," *J. Acoust. Soc. Am.*, **120**, 3553–3565 (2006).

⁶L. Troiano, P. Guerrini, and A. Barbagelata, "SACLANTCEN towed and vertical array system characteristics," NATO Undersea Research Centre, La Spezia M-117, Oct 1995.

⁷J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes*, 2nd ed. (Hemisphere, New York, 1987), p. 24.

⁸H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography*, 1st ed. (Academic, Boston, 1998), p. 41ff.

⁹C. W. Holland, "Coupled scattering and reflection measurements in shallow water," *IEEE J. Ocean. Eng.*, **27**, 454–469 (2002).

¹⁰F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, 1st ed. (Springer, New York, 2000).

Modeling scattering from azimuthally symmetric bathymetric features using wavefield superposition

John A. Fawcett

DRDC - Atlantic, P.O. Box 1012, Dartmouth, NS, Canada B2Y 3Z7

(Received 6 February 2007; revised 19 July 2007; accepted 27 August 2007)

In this paper, an approach for modeling the scattering from azimuthally symmetric bathymetric features is described. These features are useful models for small mounds and indentations on the seafloor at high frequencies and seamounts, shoals, and basins at low frequencies. A bathymetric feature can be considered as a compact closed region, with the same sound speed and density as one of the surrounding media. Using this approach, a number of numerical methods appropriate for a partially buried target or facet problem can be applied. This paper considers the use of wavefield superposition and because of the azimuthal symmetry, the three-dimensional solution to the scattering problem can be expressed as a Fourier sum of solutions to a set of two-dimensional scattering problems. In the case where the surrounding two half spaces have only a density contrast, a semianalytic coupled mode solution is derived. This provides a benchmark solution to scattering from a class of penetrable hemispherical bosses or indentations. The details and problems of the numerical implementation of the wavefield superposition method are described. Example computations using the method for a simple scattering feature on a seabed are presented for a wide band of frequencies. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2785812]

PACS number(s): 43.30.Hw, 43.30.Dr [WLS]

Pages: 3286–3295

I. INTRODUCTION

The modeling of scattering from range-dependent bathymetry or surfaces is a subject of much interest in underwater acoustics. For example, bathymetric features can cause significant sonar returns or can cast an acoustic shadow behind the feature. At higher sonar frequencies, these effects can be caused by relatively small-sized features.

There have been many methods proposed for modeling scattering from targets and/or surfaces in two- and three-dimensional, range-dependent media. These include parabolic equation methods,^{1–5} coupled mode methods,^{6–8} boundary element methods,^{9–11} multipole expansion or wavefield superposition methods,^{12–18} T-matrix methods,¹⁹ and finite difference methods.^{20,21} These various methods all have their advantages and disadvantages in terms of accuracy, computational efficiency, and generality. The wavefield superposition method of this paper is straightforward to implement and the computations can be efficiently implemented. It uses a finite computational domain for the point sources representing the fields interior and exterior to the scattering region, even for distant source and receivers. The method naturally represents the fields in the domain exterior to the scattering region by using the appropriate wave number representation of the Green's function. The azimuthal symmetry of the features considered is exploited to reduce a full three-dimensional solution to solving a sequence of two-dimensional scattering problems. There are no approximations made in terms of the Green's function or the boundary conditions. A disadvantage of the method is that, even with the azimuthal assumption, it is often not feasible to model extended scattering regions. In this case, a parabolic equation method, for example, would be more appropriate. Second,

because of the restriction to azimuthally symmetric objects, the method of this paper is not applicable to general three-dimensional scattering regions. However, the features which can be considered are still important approximations or benchmarks for small mounds, scours, seamounts, shoals, basins, etc. The same axially symmetric approximation or assumption used here has also been used by other models for bathymetric scattering.^{6–8,11} In other models, the bathymetry is taken to be invariant to one of the horizontal coordinates, once again allowing the three-dimensional solution to be constructed from a set of two-dimensional problems.²

In this paper we describe our new approach to bathymetric scattering problems using the method of wavefield superposition. Second, in this paper, a new semianalytic coupled mode approach is derived for the case of hemispherical bosses or scours in the case that the seabed has only a density jump with respect to the water column. Hemispherical bosses have been previously studied in acoustics^{22,23} and electromagnetics.²⁴ The approach of this paper provides a set of solutions to a class of penetrable bosses or scours. The results from this method will provide a benchmark example for the more general wavefield superposition method of this paper. The theory of the wavefield superposition method is described. The theoretical problem of point source placement within the scattering region and irregular frequencies is also discussed. The numerical implementation details are described. A benchmark case using the semianalytic coupled mode and wavefield superposition approaches is presented. Results from a wideband-frequency scattering computation for a simple feature on the seafloor are given both in the frequency domain and as a pulse computation using Fourier synthesis.

II. THEORY

For a scattering feature with azimuthal symmetry about the z axis, the full three-dimensional solution can be constructed from the Fourier sum of solutions to a set of two-dimensional problems. This type of approach has been used by many authors for boundary element,^{11,25} finite-element methods,²⁶ wavefield superposition methods,²⁷ thin-shell finite-element/wavefield superposition,²⁸ and coupled modes.⁵⁻⁷ The discretization of a bathymetric surface might require N_A discrete points in the azimuthal coordinate and N_R in the radial coordinate. The resulting size of the matrix for the numerical computation of the unknown points sources is of the order $(N_A \times N_R) \times (N_A \times N_R)$. On the other hand, with the azimuthal transform approach there will be N_M systems of order $N_R \times N_R$, where N_M is the required number of azimuthal Fourier components. In many cases, the computational cost of setting up and solving the N_M sets of the $N_R \times N_R$ linear systems is significantly less than that of solving the full three-dimensional system of equations.

For the pressure field, p , and the Green's function G in an azimuthally symmetric waveguide we can write

$$\begin{aligned} p(r, z, \theta) &= \sum_{m=0}^{\infty} p_m(r, z) \cos(m\theta) \\ p^{\text{inc}}(r, z, \theta) &= \sum_{m=0}^{\infty} p_m^{\text{inc}}(r, z) \cos(m\theta) \\ G(r, z, \theta; r', z', \theta') &= \sum_{m=0}^{\infty} g_m(r, z, ; r', z') \cos(m(\theta - \theta')), \end{aligned} \quad (1)$$

where (r, z) is a two-dimensional cylindrical coordinate system and θ is the azimuthal angle about the axis of symmetry, the z axis. The terms $g_m(r, z, ; r', z')$ are computed from the azimuthal transform of the Green's function. In this paper we will be considering both the Green's function for three-dimensional free space and the Green's function for a half space and their azimuthal transforms. For example, the half-space Green's function for the m th azimuthal order can be written for a source point (r', z') in the upper half space

$$\begin{aligned} g_m(r, z, ; r', z') &= - \int_0^{\infty} J_m(pr) J_m(pr') \\ &\quad \times \frac{(\exp(i\gamma_1|z - z'|) + R(p)\exp(i\gamma_1(z + z')))}{i\gamma_1} p dp, \\ &\quad z > 0 \\ &= - \int_0^{\infty} J_m(pr) J_m(pr') \\ &\quad \times \frac{T(p)\exp(-i(\gamma_2 z + \gamma_1 z'))}{i\gamma_1} p dp, \quad z < 0 \end{aligned} \quad (2)$$

where $\gamma_1 \equiv \sqrt{\omega^2/c_1^2 - p^2}$ and $R(p)$ and $T(p)$ are the reflection and transmission coefficients for the incident field in the upper half space. A similar expression exists for the case where

the source points are in the lower half space with γ , $R(p)$ and $T(p)$ replaced by the appropriate expressions.

In this paper, the expression for the pressure field from a free-field point source is given by

$$G(r, z, \theta; r', z', \theta') = \frac{1}{R} \exp(ikR) \quad (3)$$

where $k \equiv \omega/c$ and

$$R \equiv \sqrt{r^2 + r'^2 - 2rr' \cos(\theta - \theta') + (z - z')^2}. \quad (4)$$

The azimuthal transform of this function can be expressed as

$$g_m(r, z, ; r', z') = \int_0^{\pi} \frac{1}{R} \exp(ikR) \cos(m\psi) d\psi \quad (5)$$

where $\psi \equiv \theta - \theta'$, recalling from Eq. (4) that R depends on θ and θ' through their difference. There are a variety of analytical expressions for Eq. (5) either as a series of Legendre polynomials and spherical Hankel or Bessel functions or in terms of elliptic integrals.²⁵ In this paper, we will simply discretize the integral and simultaneously evaluate the integral for several values of the azimuthal order m .

For the half space Green's function, we can rewrite it in the form, for $z' \geq 0$

$$\begin{aligned} g_m(r, z, ; r', z') &= - \int_0^{\infty} J_m(pr) J_m(pr') \\ &\quad \times \frac{(R(p) - R_{\text{as}})\exp(i\gamma_1(z + z'))}{i\gamma_1} p dp \\ &\quad + t_{\text{dir}} + t_{\text{ref}}, \quad z > 0 \\ &= - \int_0^{\infty} J_m(pr) J_m(pr') \\ &\quad \times \left[\frac{T(p)\exp(-i(\gamma_2 z - \gamma_1 z'))}{i\gamma_1} \right. \\ &\quad \left. - \frac{T_{\text{as}}\exp(-i\gamma_2(z - z'))}{i\gamma_2} \right] p dp \\ &\quad + t_{\text{trans}}, \quad z < 0 \end{aligned} \quad (6)$$

where

$$\begin{aligned} t_{\text{dir}} &= \int_0^{\pi} \frac{\exp(ik_1 \sqrt{r^2 + r'^2 - 2rr' \cos\psi + (z - z')^2})}{\sqrt{r^2 + r'^2 - 2rr' \cos\psi + (z - z')^2}} \\ &\quad \times \cos(m\psi) d\psi \end{aligned} \quad (7)$$

$$\begin{aligned} t_{\text{ref}} &= R_{\text{as}} \int_0^{\pi} \frac{\exp(ik_1 \sqrt{r^2 + r'^2 - 2rr' \cos\psi + (z + z')^2})}{\sqrt{r^2 + r'^2 - 2rr' \cos\psi + (z + z')^2}} \\ &\quad \times \cos(m\psi) d\psi \end{aligned} \quad (8)$$

$$\begin{aligned} t_{\text{trans}} &= T_{\text{as}} \int_0^{\pi} \frac{\exp(ik_2 \sqrt{r^2 + r'^2 - 2rr' \cos\psi + (z - z')^2})}{\sqrt{r^2 + r'^2 - 2rr' \cos\psi + (z - z')^2}} \\ &\quad \times \cos(m\psi) d\psi \end{aligned} \quad (9)$$

and

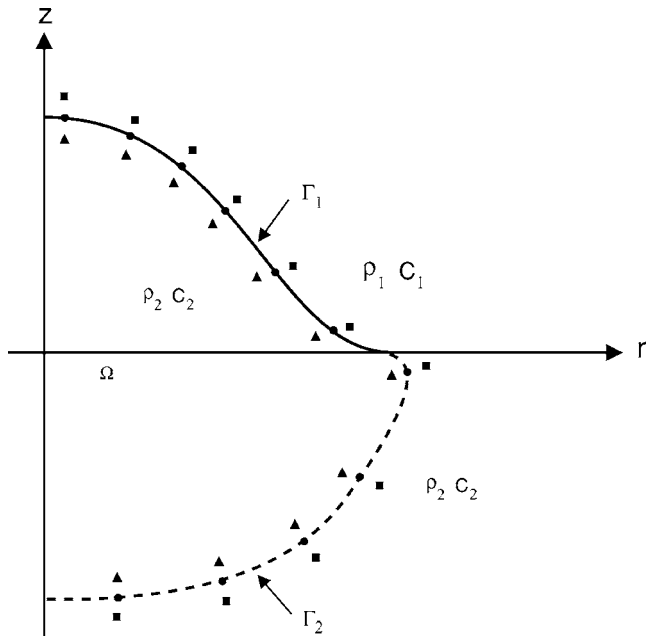


FIG. 1. A schematic of a positive bathymetric feature. The scattering chamber region Ω is surrounded by the two curves: Γ_1 which is the surface of the feature and Γ_2 which is an additional curve introduced into the seabed to enclose Ω .

$$R_{\text{as}} = \frac{\rho_2 - \rho_1}{\rho_2 + \rho_1}, \quad (10)$$

$$T_{\text{as}} = \frac{2\rho_2}{\rho_2 + \rho_1}.$$

The terms of Eqs. (7)–(9) represent singular terms which occur for $z' \rightarrow z$, $r' \rightarrow r$, and for $z=0$ (for t_{ref} and t_{trans}). In the case of a boundary integral equation method, the singular behavior of t_{dir} , t_{ref} , and t_{trans} would need to be considered analytically; in particular, the latter two terms in the case as $z, z' \rightarrow 0$. For the method of images considered in this paper, there will always be a nonzero distance between the source and the surface where the boundary conditions are applied. However, these distances can still be small for some of the discrete source/receiver positions and the analytic extraction of these terms [Eqs. (7)–(9)] is used to improve the convergence of the wave number integrals as a function of the upper limit of integration. Similar expressions to those of Eqs. (7)–(10) are obtained for $z' < 0$.

In Fig. 1 we show a schematic cross section of a “bump” on the seafloor. In this case the feature is pointing into the upper medium. The solid line is the actual bathymetry and the dashed line represents an arbitrary numerical boundary placed into the bottom. The squares represent sources placed a specified distance exterior to the region Ω and are used to represent the field within the region Ω . These are free space sources (suitably transformed into the azimuthal domain) with the sound speed c_2 . The density within the regions is that of the lower half space ρ_2 . The triangles represent sources shifted slightly to the interior of the bounding curve C and are associated with the half-space Green’s functions (in the azimuthally transformed domain) for these particular source positions. The resultant exterior field satisfies the

boundary conditions along the water/seabed interface. In the case where we are considering a basin (i.e., a negative deformation of the seabed) the situation is almost exactly the same, except that now the interior values for the sound speed and density for Ω are those of the upper half space. If a deformation is both positive and negative with respect to the flat surface $z=0$ m, then there will be an area of the exterior regions which is not the simple half space (for example, an intrusion of the upper medium into the bottom half space). In this case, one must use a larger bounding surface and introduce additional unknown point sources along the actual surface deformation. We will not consider this case in this paper.

Taking N sources with unknown amplitudes $\{a_i\}$ in the exterior region (representing the interior field) and N sources in the interior region b_i (representing the exterior field), the continuity equations across the bounding curve become, first, for the continuity of the pressure field

$$\sum_{i=1}^N a_i G_m^f(\vec{X}_k; \vec{X}_i + \vec{\delta}_i) + \sum_{i=1}^N b_i G_m^h(\vec{X}_k; \vec{X}_i - \vec{\delta}_i) + p_m^{\text{inc}}(\vec{X}_k), \quad k = 1, \dots, N \quad (11)$$

and the normal derivative across the boundary

$$\frac{1}{\rho_{\text{in}}} \sum_{i=1}^N a_i \frac{\partial G_m^f(\vec{X}_k; \vec{X}_i + \vec{\delta}_i)}{\partial n} = \frac{1}{\rho_{\text{ext}}(\vec{X}_k)} \left(\sum_{i=1}^N b_i \frac{\partial G_m^h(\vec{X}_k; \vec{X}_i - \vec{\delta}_i)}{\partial n} + \frac{\partial p_m^{\text{inc}}(\vec{X}_k)}{\partial n} \right), \quad k = 1, \dots, N. \quad (12)$$

In these equations we have indicated the discrete points along the curve C as $\vec{X}_i \equiv (r_i, z_i)$ and $\vec{X}_k \equiv (r_k, z_k)$. The source points are displaced slightly from these points, along the normal, in the positive direction $+\delta_i$ for the exterior sources and $-\delta_i$ for the interior sources. The superscripts f and h with the Green’s function refer to the freespace and the water/seabed Green’s functions, respectively. In the case where the bathymetric feature is a positive deformation into the water column, then $\rho_{\text{in}} = \rho_2$; for a negative deformation $\rho_{\text{in}} = \rho_1$. The density term $\rho_{\text{ext}}(\vec{X}_k)$ is equal to ρ_1 for \vec{X}_k such that $z_k \geq 0$ and is equal to ρ_2 otherwise. Thus the normal derivative (the radial derivative) of the incident field is discontinuous across $z=0$. It should also be noted that although we are stressing the modeling of scattering by three-dimensional bathymetric features, the region Ω can also model an embedded region of different sound velocity and density. In this case the interior values of c and ρ would be different from those of the two surrounding half spaces.

III. A DERIVATION OF A COUPLED-MODE SOLUTION FOR A CLASS OF PENETRABLE HEMISPHERICAL BOSSES OR BASINS

In previous references,^{29,30} modal solutions were derived for a half-space problem in the case where the two half

spaces had the same sound speeds but different densities. In Ref. 30 this allowed for the solution of scattering from an embedded cylinder in terms of a coupled mode solution. In this paper, we will derive a corresponding three-dimensional solution. In particular, we will describe the coupled mode problem to solve for each azimuthal order m . Let us consider a sphere which is exactly half buried between two half spaces, each with sound speed c_{ext} and densities ρ_1 and ρ_2 . The sphere has an internal sound speed c_{in} and density ρ_{in} . In the case where $c_{\text{in}}=c_{\text{ext}}$ and $\rho_{\text{in}}=\rho_1$ we have a basin problem and for $\rho_{\text{in}}=\rho_2$ we have a boss problem. The general solution in the interior of the sphere can be expanded in terms of the angular functions

$$\psi_n^m(\phi, R) = N_n^m P_n^m(\cos \phi), \quad (13)$$

where P_n^m denotes a Legendre function, and N_n^m is the normalization constant such that

$$\int_0^\pi \frac{1}{\rho_{\text{in}}} (\psi_n^m(\cos \phi))^2 \sin \phi d\phi = 1. \quad (14)$$

This angular function has the associated radial function, $j_n(k_{\text{int}}R)$ where j_n denotes a spherical Bessel function.

In the exterior domain, the angular functions have the form for $n=0, 2, 4, \dots$

$$\tau_n^m(\cos \phi) = P_{n+m}^m(\cos \phi) \quad 0 \leq \phi \leq \pi \quad (15)$$

and for $n=1, 3, 5, \dots$

$$\tau_n^m(\cos \phi) = P_{n+m}^m(\cos \phi) \quad 0 \leq \phi \leq \pi/2 \quad (16)$$

$$\tau_n^m(\cos \phi) = \frac{\rho_2}{\rho_1} P_{n+m}^m(\cos \phi) \quad \pi/2 \leq \phi \leq \pi.$$

These modal functions are then normalized such that

$$\int_0^\pi \frac{1}{\rho(\phi)} (\tau_n^m)^2(\cos \phi) \sin \phi d\phi = 1. \quad (17)$$

The associated radial functions for these angular functions are $h_n(k_{\text{ext}}R)$. The density term $\rho(\phi)$ in Eq. (17) is simply ρ_1 for $0 \leq \phi \leq \pi/2$ and ρ_2 for $\pi/2 \leq \phi \leq \pi$. The specified form of the angular modal functions is used because for $n+m = m, m+2, m+4, \dots$ the associated Legendre functions are even (and have zero derivative with respect to ϕ) and thus are continuous across $z=0$, $\phi=\pi/2$, and trivially satisfy

$$\frac{1}{\rho_1} \frac{\partial \tau}{\partial \phi} = \frac{1}{\rho_2} \frac{\partial \tau}{\partial \phi} \quad \text{at } \phi = \pi/2. \quad (18)$$

The Legendre functions $P_{n+m}^m(\cos \phi)$ are odd with respect to ϕ at $\phi=\pi/2$ for $n=1, 3, 5, \dots$ and thus, in fact, the definition of τ as in Eq. (16) is continuous (zero) at $\phi=\pi/2$ and satisfies the boundary condition of Eq. (18). We can now set up a system of equations to be satisfied along the surface (in this coordinate system, $R=a$ and $0 \leq \phi \leq \pi$) of the sphere

$$\begin{aligned} & \sum_{i=0}^N c_i \psi_i(\cos \phi) j_{i+m}(k_{\text{in}}R) \\ &= \sum_{i=0}^N d_i \tau_i(\cos \phi) h_{i+m}(k_{\text{ext}}R) \\ &+ ik_{\text{ext}} \sum_{i=0}^N \tau_i(\cos \phi) j_{i+m}(k_{\text{ext}}R) h_{i+m}(k_{\text{ext}}R_s) \tau_i(\cos \phi_s) \end{aligned} \quad (19)$$

$$\begin{aligned} & \frac{1}{\rho_{\text{in}}} \sum_{i=0}^N c_i \psi_i(\cos \phi) k_{\text{in}} j'_{i+m}(k_{\text{in}}R) \\ &= \frac{1}{\rho_{\text{ext}}(\phi)} \left(\sum_{i=0}^N d_i \tau_i(\cos \phi) k_{\text{ext}} h'_{i+m}(k_{\text{ext}}R) \right. \\ & \left. + ik_{\text{ext}} \sum_{i=0}^N \tau_i(\cos \phi) k_{\text{ext}} j'_{i+m}(k_{\text{ext}}R) h_{i+m}(k_{\text{ext}}R_s) \tau_i(\cos \phi_s) \right), \end{aligned} \quad (20)$$

where the last terms of Eqs. (19) and (20) are the expansions of the incident field for a point source at (R_s, ϕ_s) .

For the system of equations in Eq. (19), we project $1/\rho_{\text{in}} \psi_j(\cos \phi) \sin \phi$, $j=0, \dots, N$ on both sides and integrate. For the systems of equations in Eq. (20) we project $\tau_j(\cos \phi) \sin \phi$ on both sides and integrate. The last terms in Eqs. (19) and (20) are the expansion of the incident field from a point source located (in spherical coordinates) at $(R_s, \phi_s, \theta=0)$. The resulting system of equations has the matrix/vector form

$$\begin{pmatrix} D_1 & R_1 \\ R_2 & D_2 \end{pmatrix} \begin{pmatrix} \vec{c} \\ \vec{d} \end{pmatrix} = \begin{pmatrix} \vec{r}_1 \\ \vec{r}_2 \end{pmatrix}. \quad (21)$$

Here D_1 and D_2 are diagonal matrices with entries $j_{i+m}(k_{\text{in}}R)$ and $h_{i+m}(k_{\text{ext}}R)$, respectively. The matrices R_1 and R_2 have the form

$$R_1(p, q) = h_{q+m}(k_{\text{ext}}R) \frac{1}{\rho_{\text{in}}} \int_0^\pi \psi_p(\cos \phi) \tau_q(\cos \phi) \sin \phi d\phi, \quad (22)$$

$$R_2(p, q) = j_{q+m}(k_{\text{in}}R) \frac{1}{\rho_{\text{in}}} \int_0^\pi \psi_q(\cos \phi) \tau_p(\cos \phi) \sin \phi d\phi, \quad (23)$$

and \vec{r}_1 and \vec{r}_2 represent the known incident field terms. It is interesting to note that the solutions to this scattering problem provide a generalization to the case where the interface boundary condition is pressure release ($\rho_2 \rightarrow 0$) or rigid ($\rho_2 \rightarrow \infty$). We have verified numerically that in the limit of pressure release and rigid bosses (positive into the water column) that the correct analytic solutions are obtained.

The solution of the system of equations of Eq. (21) yields the interior coefficients \vec{c} and the exterior coefficients \vec{d} which can then be used to compute the scattered $p_{\text{sc}}(r, z; m)$ and total pressure fields $p(r, z; m)$. The scattered

pressure field is defined as the difference between the total field and the incident field which would exist in the absence of the boundary feature. These two-dimensional fields are computed for M values of the azimuthal order m and the three-dimensional fields are then computed (for example, the scattered field) from

$$p_{sc}(r, z, \theta) = \sum_{m=0}^M \epsilon_m p_{sc}(r, z; m) \cos(m\theta) \quad (24)$$

where $\epsilon_m=1$ for $m=0$ and $\epsilon_m=2$ otherwise.

IV. NUMERICAL COMPUTATIONS

In this section we describe some of the details of the implementation of the coupled mode benchmark solution and of the wavefield superposition approach. This is followed by a comparison of the solutions from the two methods in the case of the seabed having only a density change. Some issues with the superposition approach are described and discussed. The method is then applied in the case of a deformation of a seabed having a different sound speed, density, and attenuation.

A. Numerical implementation

The numerical implementation of the coupled mode benchmark solution of Sec. III is straightforward. The modal projection integrals of Eqs. (22) and (23) are evaluated numerically. The columns of the matrix of Eq. (21) are normalized by the appropriate spherical Bessel or Hankel function, $j_n(kR)$ or $h_n(kR)$ in the case where the order is greater than 1.5 the value of the argument. This is done in order to prevent a column of values from being excessively small or large for large orders of the Bessel and Hankel functions.

In the case of a general seabed, the solution of the system of equations for the wavefield superposition source amplitudes, Eqs. (11) and (12), yields the coefficients $\{a_i\}$ and $\{b_j\}$ of the displaced point sources. The matrix corresponding to the system of linear equations involves either the free-space or two half-space Green's function (or their normal derivatives) evaluated at a large set of source/receiver positions. The expressions for the Green's function in the region exterior to the scattering region involve evaluations of wave number integrals [see, for example, Eq. (6)] and the azimuthal transform of the extracted singular terms [e.g., Eqs. (7) and (8)]. The extracted terms are not singular in the case of the wavefield superposition method as there is always a nonzero distance between the source and boundary points and thus these terms could be included in the wave number integral. However, by extracting analytic terms [as in Eq. (6)], it is hoped that the convergence of the wave number integrals, in terms of the required upper limit of the integral, is improved. In order to evaluate the azimuthal transforms of the analytically extracted terms, 201 discretization points were used and the integrals were evaluated for multiple values of $\cos(m\psi)$ without recomputing the portion of the integrand which is independent of the azimuthal order m . In the density-jump only case, the wave number integrals are zero and only the extracted terms contribute. In the numerical

evaluation of the wave number integrals for the general seabed case, there are various parameters to consider: the number of discretization points, the contour offset into the complex- p plane, and the upper limit of integration. The suitable choice of these types of parameters arises in all implementations of wave number integrals. For example, a discussion of these issues for fast field programs can be found in Ref. 31. In this paper, we consider integrals of the form

$$\int_0^{U\omega/c} J_m(p(t)r) J_m(p(t)r_s) K(p(t), z, z_s) p(t) \frac{dp}{dt} dt \quad (25)$$

for

$$p(t) = t - i\epsilon \sin\left(\frac{\pi t}{t_{up}}\right) \quad (26)$$

and $t_{up}=U\omega/c$ (c is taken in our examples as 1500 m/s). This choice of the contour of p deforms the integration contour away from the real line except at the origin and at the upper limit of integration where it comes back to the real axis.

In all our computations we used 501 uniformly spaced values of t in Eqs. (25) and (26). This is a sufficient number to avoid any spatial aliasing issues which arise when the wave number sampling size is too large. In the case of the computation of the Green's function and its normal derivative for the system of equations for the amplitude coefficients, the upper limit of integration factor, U [Eq. (25)] varies with frequency. In the multiple frequency computation which is considered in the numerical examples, this upper limit factor U varied from 16 at $f=500$ Hz to 6 at $f=2500$ Hz. For this example, the contour offset factor ϵ in Eq. (26) was set to 2 for all frequencies. In the numerical example with this frequency range, the maximum horizontal ranges of interest for the matrix Green's function were less than 2.3 m. We will also consider a problem where the dimensions of the scattering feature are increased by a factor of 100 and the frequency is 25 Hz. In this case, the same parameters as the 2500 Hz/small feature case are used with the exception of ϵ which is decreased to 0.02.

Once the coefficients a_i and b_j of Eqs. (11) and (12) have been computed, they can be used in conjunction with the appropriate Green's function to compute the scattered and total fields. In the computations of the pressure field values for a grid of receiver positions, we simply used wave number integral representations with no subtraction of analytic expressions. In this case, the upper limit factor U for these wave number integrals and the contour offset ϵ are increased by a factor of two.

The computations of the field values are done for each azimuthal order m . The final solution is computed from the azimuthal sums of the form of Eq. (24). In cases where the incident pressure field is added to the solution (for example, to compute the total field in the exterior region), the incident field is computed only once in the form

$$p^{\text{inc}}(r, z; r_s, z_s) = \int_0^\infty J_0(p|r - r_s|)K(p; z_s)p dp \quad (27)$$

where $K(p; z_s)$ represents the appropriate kernel for the upper or lower half space. In other words, we compute the incident field with a three-dimensional expression with the origin at the source position.

The numerical code was implemented in MATLAB. The integrals were performed in a discretized form using matrix/vector operations where possible. For example, the wave number integrals of Eq. (6) (and also the set for $z' < 0$) can be written in the separable form

$$\int_0^\infty K_1(r, z, ; p)K_2(r', z'; p)F(p)dp. \quad (28)$$

For example, the second expression of Eq. (6) can be written as

$$\begin{aligned} & \int_0^\infty J_m(pr)J_m(pr') \left[\frac{T(p)\exp(-i(\gamma_2 z - \gamma_1 z'))}{i\gamma_1} \right. \\ & \quad \left. - \frac{T_{\text{as}} \exp(-i\gamma_2(z - z'))}{i\gamma_2} \right] p dp \\ & = \int_0^\infty [J_m(pr)\exp(-i\gamma_2 z)] \\ & \quad \times \left[J_m(pr') \left(T(p) \frac{\exp(i\gamma_1 z')}{i\gamma_1} - T_{\text{as}} \frac{\exp(i\gamma_2 z')}{i\gamma_2} \right) \right] p dp. \end{aligned} \quad (29)$$

This form allows us to efficiently evaluate the discretized version of the integral for the various combinations of source and receiver positions without reevaluating the integrand for each possible source/receiver pair. In addition, any integrand quantities which do not involve the azimuthal order m are only evaluated once during the computations for all the required azimuthal orders.

It was found that at high azimuthal orders the system of equations for the point sources sometimes became ill conditioned. This is a result of the fact that the coefficients of the point sources for small values of the cylindrical radius r were poorly determined because their net contribution to the pressure field is very small. Because of this problem, a value of 10^{-8} was used at lower bound on the singular values in the solution of the system of equations.

As will be illustrated in the numerical examples, there is another source of ill conditioning, due to the curve of the interior point sources and the corresponding interior Dirichlet problem.³² The solution for the interior point sources becomes ill conditioned (and this will happen at each azimuthal order) near frequencies corresponding to the eigenvalues of the interior Dirichlet problem as defined by the interior curve used for the source locations. In Ref. 32 it was suggested that one could use a combination of monopoles and dipoles to alleviate this problem. The approach we will use is to consider two curves of point sources each slightly offset from each other by a small separation Δ (along the normal direction). Then, for example, we can consider one source to have

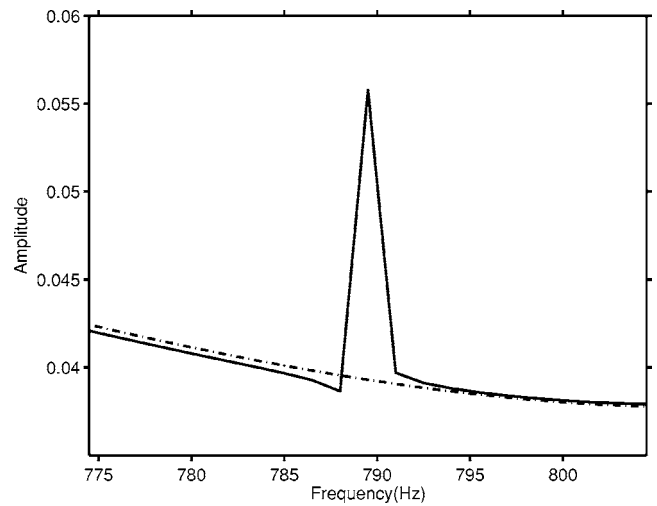


FIG. 2. The backscattered amplitude as a function of frequency at a point along the z axis and at a height 3.33 m above the seabed as computed using a single curve of monopoles (solid) and using two curves of sources approximating a combination of monopoles and dipoles (dashed).

the weighting $(0.5 + i/(k\Delta))$ and the corresponding source from the other curve to have the weight $(0.5 - i/(k\Delta))$ where $k = 2\pi f/c$. These two sources in conjunction with each other approximate a monopole and dipole along the midway curve. In the construction of the system of equations, the contributions from the two sets of point source locations are evaluated and then combined with the weightings discussed above. The system size and the number of unknowns is the same as simply using a set of monopoles.

B. Density-jump benchmark solution

As a first numerical benchmark we will use both the benchmark modal approach and the superposition approach. We will consider a spherical region (radius = 1.0 m) with an interior sound speed $c_p = 1500$ m/s and $\rho = 1$ g/cm³. The upper surrounding half space has a sound speed of 1500 m/s and a density of 1 g/cm³ and the lower half space a sound speed of 1500 m/s and a density of 5 g/cm³. This scenario represents a half-spherical basin intruding into the lower, high-density half space. A point source is located at 4 m above the seabed/water interface and along the z axis. In this case only the order $m=0$ need be considered. The bounding chamber is simply a half circle in the (r, z) space in this case. There are 62 unknown coefficients for both the interior and exterior sources resulting in a 124×124 system of equations. The interior point sources are positioned at 0.05 m from the boundary (i.e., at a radius of 0.95 m) and the exterior sources are located at a radius of 1.05 m. In Fig. 2 we show the amplitude of the scattered field (solid line) as a function of frequency for a point approximately along the z axis and at a height of 3.33 m above the seabed. As can be seen, the amplitude suffers a large jump at approximately 789.5 Hz which corresponds to $(2\pi f/c)0.95$ being equal to a zero of j_0 . The significance of this term can be seen from the expansion of the free space Green's function

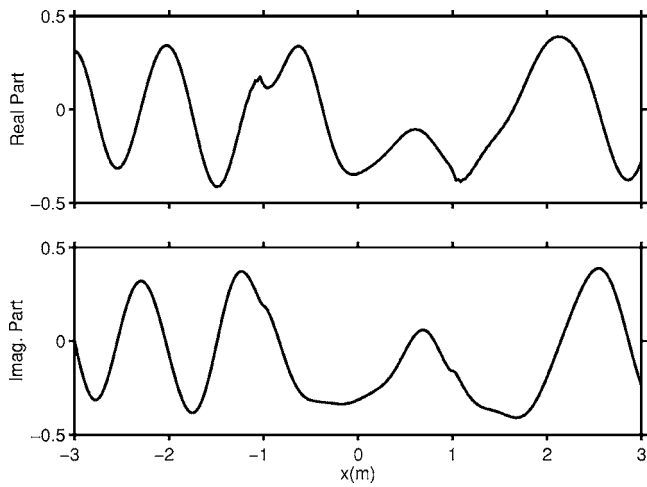


FIG. 3. A comparison of the total pressure field (real and imaginary parts) along the line $z=0$ for the basin scattering feature -2500 Hz, modal solution (solid line) and superposition method (dashed).

$$G^m(R, \phi; R_S, \phi_S) = ik \sum_{n=m}^{\infty} P_n^m(\cos \phi) P_n^m(\cos \phi_S) j_n(kR_{<}) h_n(kR_{>}). \quad (30)$$

In this equation one can see that for an interior source representing an exterior field (in the case of a homogeneous exterior) that the coefficient of the n th Legendre function will be zero in the case where $j_n(kR_{<})=0$. Next, a second set of interior sources is added at $R=0.93$ m along with the sources at 0.95 m to approximate a set of monopoles and dipoles at $R=0.94$ m (with a single set of unknown coefficients). As can be seen by the dashed curve, the new solution is now smooth as a function of frequency. [From our experience, it does not seem necessary to introduce a second set of exterior sources. This seems reasonable if one considers the expansion of the free space Green's function, Eq. (30), where we do not need to be concerned with zeros of $h_n(kR_S)$].

A point source of 2500 Hz is now located at a range of 4 m and an angle of 10° from the z axis. In this case, the positioning of the point sources is the same as for the previous example, with an interior set of two circles of slightly displaced sources and an exterior set. However, in this case we use 122 exterior point sources representing the interior field and 122 double point sources in the interior, representing the exterior field. The size of the resulting system of equations is 244×244 . The number of azimuthal terms used was given by the expression $n_m = 2\pi f / 1500 \times 1.5 + 9$. As a benchmark solution we use the modal solution, using the same number of azimuthal terms as for the scattering-chamber approach. The number of modes considered for each azimuthal order is given by the expression, $n_{\text{mode}} = 2\pi f / 1500 \times 3 + 16$. In Fig. 3 we show a comparison of the computed total fields along the line $z=0$ (solid line is modal solution, dashed is superposition approach). As can be seen, the agreement is excellent. The points $x = \pm 1$ m are the boundaries of the basin. In Fig. 4 we show the total pressure field computed using the scattering chamber approach for a grid of positions. As can be seen, this "basin" has a significant

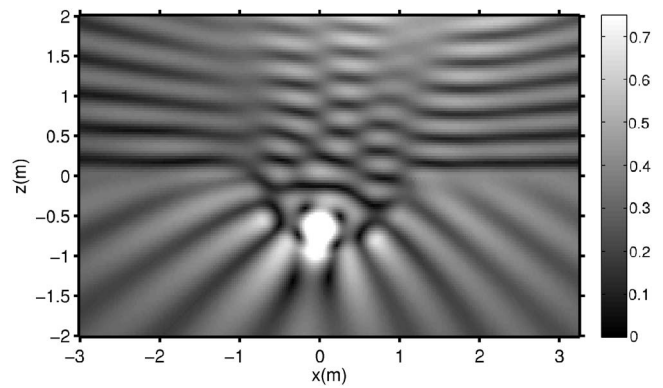


FIG. 4. Computed two-dimensional slice of three-dimensional pressure field using the superposition approach at 2500 Hz. The source is located off the plot at a distance of 4 m and 10° off the vertical.

effect upon the propagation and there is a strong focusing feature near the bottom. It is also important to note that there are no observable discontinuities in going across the boundary $r=1$ where r denotes the distance from the origin. The values for $x < 0$ are computed by using $\theta = \pi$ in the azimuthal sum.

The pressure field is predicted to be somewhat ill behaved near $x = \pm 1$ m, $z = 0$ m. This is because there is a discontinuity in the boundary condition for $p_n / \rho(\theta)$ across the line $z=0$. In Fig. 5 we show the real part of the amplitudes for the exterior (the first 122 points) and interior (the last 122 points) for $m=2$ and $f=2500$ Hz. As can be seen, the coefficients become larger in amplitude and more erratic as a function of position toward $z=0$. However, as seen from Fig. 4, they do combine to give a nicely continuous solution across the boundary $r = 1$ m.

C. A general seabed feature

We now consider the more general situation of an axisymmetric feature with the upper half space consisting of

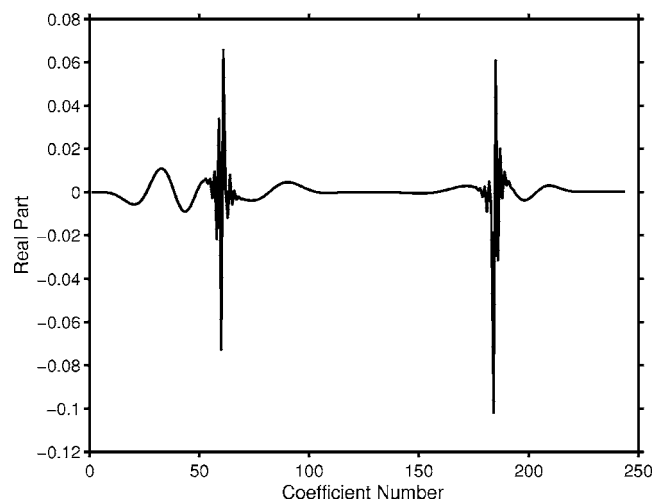


FIG. 5. The real part of the solution for $m=2$ for the point source amplitudes for the superposition approach at 2500 Hz. The values near 61 and 183 represent the solution near $r=1$ m and $z=0$ m.

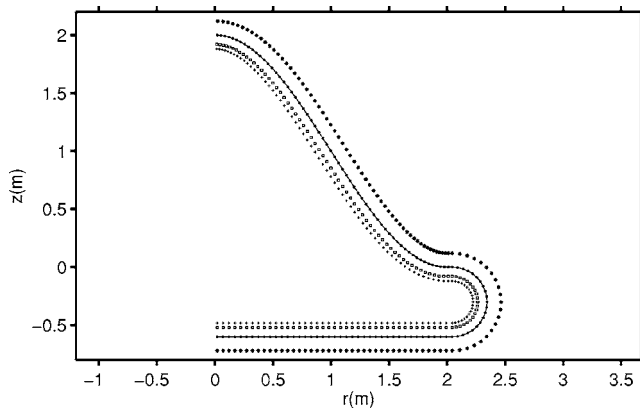


FIG. 6. Two-dimensional vertical cross section of bathymetric feature. The portion above $z=0$ is the actual feature. A smooth bounding curve is added in the bottom. The line with dots is the curve along which the continuity equations are imposed (at the dots). The plus signs and squares are the source points interior to the domain which represent the scattered fields in the exterior; the half space Green's functions (in the azimuthal domain) are used for these sources. The exterior sources, indicated by an asterisk, represent the fields within the domain and these sources use the free space Green's function with the interior value of sound speed (complex in the case of attenuation). The positions are shown for 500 Hz.

one fluid c_1, ρ_1 overlying the seabed c_2, ρ_2 . We consider the following analytical expression for the deformation of the seabed:

$$z(r) = \frac{a}{2} \left(1 + \cos\left(\frac{2\pi r}{w}\right) \right). \quad (31)$$

We will consider $a=2$ and $w=4$; this is a "bump" which has a height of 2 m and a radial extent of 2 m. It is important to recall that the actual feature results from rotating this cross section about the z axis, so that the true lateral extent in a horizontal slice is 4 m. A point source in the water column at 10 m range (here, $\sqrt{r^2+z^2}$) and at a height of 1 m above the seabed is considered. For the upper medium, $c_1=1500$ m/s, $\rho_1=1$ g/cm³, and for the lower sediment $c_2=1800$ m/s and $\rho_2=1.5$ g/cm³. The lower medium has an attenuation of 0.3 dB/ λ which is implemented by using a complex-valued c_2 . The two-dimensional slice of the scattering feature, the additional bounding curve in the bottom, and the source points are shown in Fig. 6. The number of azimuthal terms used in the computation varied with respect to frequency according to the relation $n_m = 6\pi f / 1500 + 6$. This results in 37 (38, including $m=0$) azimuthal terms for the frequency of 2500 Hz. There are 121 interior and exterior point sources. The distance of these sources from the curve is important in obtaining accurate solutions and it was found that optimal position varied somewhat with frequency. For lower frequencies, the distance was greater than at higher frequencies. This is reasonable if one considers the longer wavelength at lower frequencies. In the following we compute the pressure fields for frequencies between 500 and 2500 Hz in steps of 25 Hz. Two sets of interior point sources (but with the same coefficient for each radially displaced pair) were used. For $500 \text{ Hz} \leq f < 1200 \text{ Hz}$, the interior point sources were displaced 0.12 and 0.08 m. For frequencies greater than or equal to 1200 Hz, the displacements were 0.09 and 0.07 m. The external point sources were placed at the greater of the

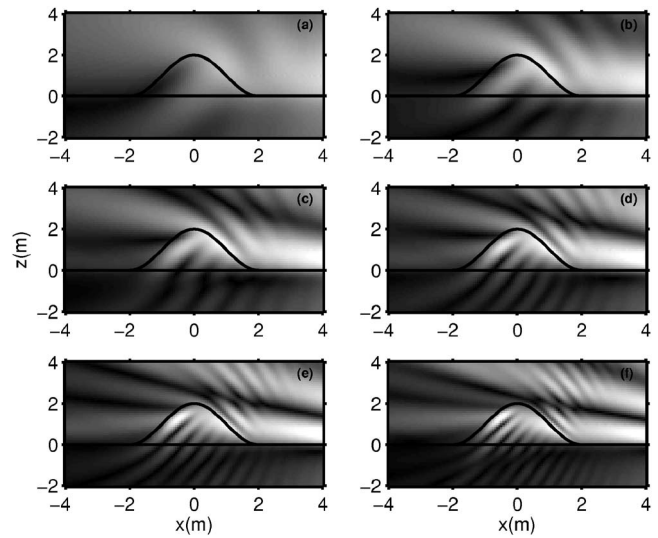


FIG. 7. Two-dimensional vertical ($y=0$ m) cross section of the total pressure field for six different frequencies: (a) 500 Hz; (b) 900 Hz; (c) 1300 Hz; (d) 1700 Hz; (e) 2100 Hz; and (f) 2500 Hz. The point source is to the right of the grid. The gray scale represents the absolute values of the pressure fields in the interval $[0 0.3]$.

two distances away from the boundary. An example of this discretization is shown in Fig. 6 for the 500 Hz case. This sequence of distances yielded very good scattering results as a function of frequency. It is interesting to note that for the pair of interior point sources closest to the line $z=0$, one of the point sources is above $z=0$ and one is below. We use the Green's function representation appropriate for the specified individual source location. The parameters used in the numerical implementation of the wave number integrals were described previously.

In Figs. 7 and 8 we show the computed pressure fields for six different frequencies for a vertical grid of receiver locations and then for a horizontal grid at a fixed vertical distance of $z=0.50$ m. In these figures there are no observable discontinuities of the field across the scattering surface and the additional bounding curve in the bottom. It is interesting to note that since the feature is penetrable, some energy propagates into the bump and out the other side. In Fig. 8 the circle denotes the boundary of the bathymetric feature at this altitude above the seabed and hence the acoustic parameters within this circle correspond to those of the seabed. As one would expect, the interference patterns between the direct and reflected incident fields and the scattered field become more complicated for higher frequencies.

In Figs. 9 and 10, time domain computations are shown. The frequency results were transformed from the frequency domain to the time domain by using the expression

$$p(x, y, z, t) = \Re \left\{ \sum_{n=1}^N p(x, y, z; f_n) \exp(-i2\pi f_n t) S(f_n) \right\} \quad (32)$$

where for the source weighting we used

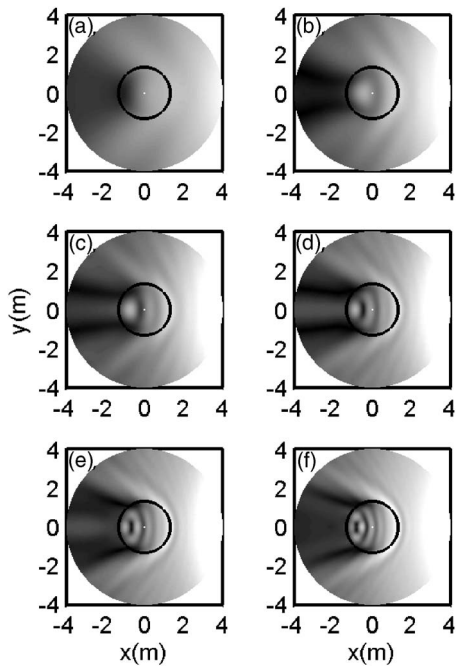


FIG. 8. Two-dimensional horizontal cross section ($z=0.5$ m) of the total pressure field for six different frequencies: (a) 500 Hz; (b) 900 Hz; (c) 1300 Hz; (d) 1700 Hz; (e) 2100 Hz; and (f) 2500 Hz. The point source is to the right of the figure (off the grid). The gray scale represents the absolute values of the pressure fields in the interval $[0, 0.25]$.

$$S(f) = \exp(-(f - 1500)^2 / (2 \times 500^2)). \quad (33)$$

In Fig. 9 the resulting pulses are shown for a vertical plane for six instances of time. The pulse is propagating from the right to the left of the plot. The incident pulse, in the seabed, is slightly ahead of the pulse in the water because of the higher sound speed in the seabed. The pulse can be seen to travel up the scattering features and also through the feature. A reflected wavefront begins to appear in panel(d). In Fig. 10

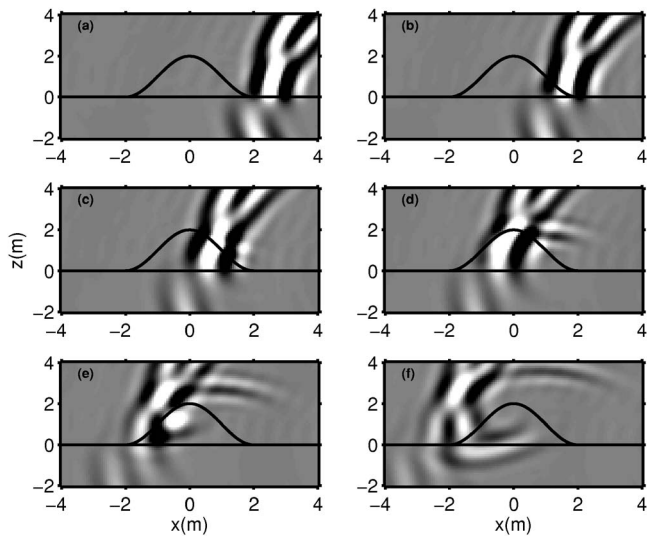


FIG. 9. Two-dimensional vertical ($y=0$ m) cross section of the pressure field (pulse computation) for six different times: (a) 4.9 ms; (b) 5.5 ms; (c) 6.1 ms; (d) 6.7 ms; (e) 7.3 ms; and (f) 7.9 ms. The point source is to the right of the figure (off the grid). The gray scale represents the real-valued time domain pressure fields in the interval $[-3.0, 3.0]$ (black is negative, white is positive).

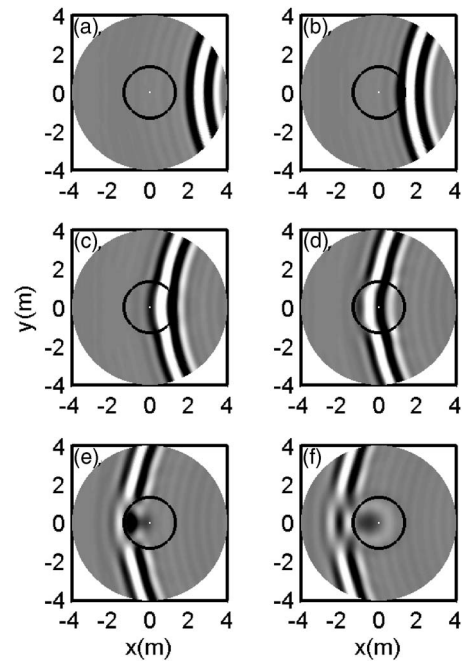


FIG. 10. Two-dimensional horizontal ($z=0.50$ m) cross section of the pressure field (pulse computation) for six different times: (a) 4.9 ms; (b) 5.5 ms; (c) 6.1 ms; (d) 6.7 ms; (e) 7.3 ms; and (f) 7.9 ms. The point source is to the right of the figure (off the grid). The gray scale represents the real-valued time domain pressure fields in the interval $[-3.0, 3.0]$ (black is negative, white is positive).

the propagation of the pulse in a horizontal plane can be seen. In panel(c) the wavefront is just entering into the feature at this depth and begins to pull ahead of the wavefront on the outside. In panel(f) a backscattered wavefront within the feature can be observed.

The scattering feature, just considered, was 2 m in height. Even with this small height, 37 azimuthal orders were used for a frequency of 2500 Hz. A much larger version of the same scattering feature can be considered for lower frequencies. For example, at 25 Hz we can increase the height of the feature to 200 m as well as increasing its lateral extent. One would expect that one should obtain the same pressure fields (amplitude reduced by a factor of 100) if all the spatial dimensions of the original problem for 2500 Hz are increased by a factor of 100 (including the source distance). Thus to compute the field scattered by this large feature at 25 Hz, we simply scale the spatial dimensions of the parameters used in the 2500 Hz computation. The only parameter we change for the numerical evaluation of the wave number integrals is the contour offset which is reduced from 2 to 0.02 in Eq. (26). The resulting pressure field (multiplied by a factor of 100) is shown in the top plot of Fig. 11 and the corresponding 2500 Hz computation is shown in the bottom plot. As can be seen, the two results are in excellent agreement. This example illustrates: (a) the applicability of the method to model large feature scattering at low frequencies and (b) the numerical consistency between the high- and low-frequency computations.

V. SUMMARY

In this paper we have described an efficient and accurate approach for computing the pressure field scattered from a

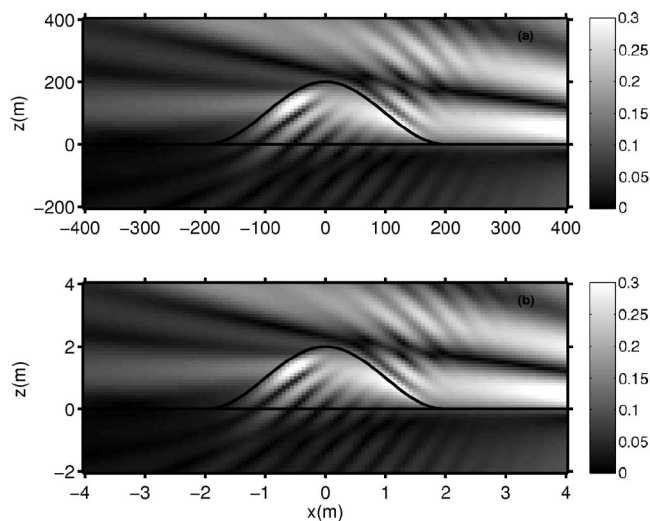


FIG. 11. Two-dimensional vertical ($y=0$ m) cross section of the total pressure field for: (a) 25 Hz with the spatial dimensions and pressure amplitude scaled by a factor of 100 (b) 2500 Hz and small scattering feature. The gray scale represents the absolute values of the pressure fields in the interval [0 0.3].

three-dimensional seabed feature. The feature was taken to be azimuthally symmetric and the deformation to the surrounding flat seabed was taken to be of one polarity, positive—a “bump” and negative—a “basin.” The computational method solved a sequence of two-dimensional problems and then used Fourier synthesis to construct the full three-dimensional solution. A benchmark solution for the case of a hemispherical deformation with a density-contrast seabed was derived. This type of solution was used to show that the wave superposition/scattering chamber approach was very accurate for this case. The problem of irregular frequencies for the superposition method was addressed by the introduction of a second set of interior point sources. This more general method was then used to solve the scattering from a 2 m high, penetrable “bump.” The pressure field was accurately computed for a large range of frequencies for both vertical and horizontal planes. At the highest frequency, 37 azimuthal orders were computed. The multifrequency computations allowed for the computation of the pressure field in the time domain for an incident pulse. Finally, we concluded with an example of a low-frequency (25 Hz) computation for a 200 m high scattering feature.

¹M. D. Collins and S. Chin-Bing, “A three-dimensional parabolic equation model that includes the effects of rough boundaries,” *J. Acoust. Soc. Am.* **87**, 1104–1109 (1990).

²G. J. Orris and M. D. Collins, “The spectral parabolic equation and three-dimensional backscattering,” *J. Acoust. Soc. Am.* **96**, 1725–1731 (1994).

³M. F. Levy and A. A. Zaporozhets, “Target scattering calculations with the parabolic equation method,” *J. Acoust. Soc. Am.* **103**, 735–741 (1998).

⁴D. Zhu and L. Bjorno, “A three-dimensional, two-way, parabolic equation model for acoustic backscattering in a cylindrical coordinate system,” *J. Acoust. Soc. Am.* **108**, 889–898 (2000).

⁵J. F. Lingeitch, M. D. Collins, M. J. Mills, and R. B. Evans, “A two-way parabolic equation that accounts for multiple scattering,” *J. Acoust. Soc. Am.* **112**, 476–480 (2002).

⁶R. B. Evans, “Stepwise coupled mode scattering of ambient noise by a cylindrically symmetric seamount,” *J. Acoust. Soc. Am.* **119**, 161–167 (2006).

⁷J. Fawcett, “Coupled-mode modeling of acoustic scattering from three-dimensional, axisymmetric objects,” *J. Acoust. Soc. Am.* **102**, 3387–3393 (1997).

⁸M. I. Taroudakis, “A coupled-mode formulation for the solution of the Helmholtz equation in water in the presence of a conical sea-mount,” *J. Comput. Acoust.* **4**, 101–121 (1996).

⁹F. J. Sanchez-Sesma and F. Luzon, “Seismic response of three-dimensional alluvial valleys for incident P,S, and Rayleigh waves,” *Bull. Seismol. Soc. Am.* **85**, 269–284 (1995).

¹⁰P. Gerstoft and H. Schmidt, “A boundary element approach to ocean seismoacoustic facet reverberation,” *J. Acoust. Soc. Am.* **89**, 1629–1642 (1991).

¹¹J. A. Fawcett, “An efficient three-dimensional boundary integral equation for solving azimuthally symmetric scattering problems in the oceanic waveguide,” *J. Acoust. Soc. Am.* **94**, 2307–2314 (1993).

¹²J. Murphy, G. Li, S. Chin-Bing, and D. King, “Multifilament source model for short-range underwater acoustic problems involving penetrable ocean bottoms,” *J. Acoust. Soc. Am.* **99**, 845–850 (1996).

¹³M. Imhof, “Computing the elastic scattering from inclusions using the multiple multipoles method in three dimensions,” *Geophys. J. Int.*, **156**, 287–296 (2004).

¹⁴A. Sarkissian, “Multiple scattering effects when scattering from a target in a bounded medium,” *J. Acoust. Soc. Am.* **96**, 3137–3144 (1994).

¹⁵J. Fawcett, “A scattering-chamber approach for solving finite rough surface scattering problems,” *J. Acoust. Soc. Am.* **118**, 1348–1357 (2005).

¹⁶A. T. Abawi and M. B. Porter, “The use of the equivalent source technique in calculation of scattering from underwater targets,” *Proceedings of Boundary Influences In High Frequency, Shallow Water Acoustics*, University of Bath, U.K., September 5–9, 2005, edited by N. G. Pace and P. Blondel, pp. 341–347.

¹⁷A. T. Abawi and M. B. Porter, “Propagation in an elastic wedge using the virtual source technique,” *J. Acoust. Soc. Am.* **121**, 1374–1382 (2007).

¹⁸I. Lucifredi and H. Schmidt, “Subcritical scattering from buried elastic shells,” *J. Acoust. Soc. Am.* **120**, 3566–3583 (2006).

¹⁹R. Lim, “Acoustic scattering by a partially buried three-dimensional elastic obstacle,” *J. Acoust. Soc. Am.* **104**, 769–782 (1998).

²⁰D. R. Burns, “Acoustic and elastic scattering from a seamount in three dimensions—A numerical study,” *J. Acoust. Soc. Am.* **92**, 2784–2791 (1992).

²¹R. A. Stephen and S. A. Swift, “Modelling seafloor geoacoustic interaction with a numerical scattering chamber,” *J. Acoust. Soc. Am.* **96**, 973–990 (1994).

²²V. Twersky, “On the non-specular reflection of plane waves of sound,” *J. Acoust. Soc. Am.* **22**, 539–546 (1950).

²³V. Twersky, “On the nonspecular reflection of sound from planes with absorber bosses,” *J. Acoust. Soc. Am.* **23**, 336–338 (1951).

²⁴H. Sakurai, M. Ohki, K. Motojim, and S. Kozaki, “Scattering of Gaussian beam from a hemispherical boss on a conducting plane,” *IEEE Trans. Antennas Propag.*, **52**, 892–894 (2004).

²⁵J. Grannell, J. Shirron, and L. Couchman, “A hierarchic p-version boundary-element method for axisymmetric acoustic scattering and radiation,” *J. Acoust. Soc. Am.* **95**, 2320–2329 (1994).

²⁶M. Zampolli, A. Tesei, and J. Blottmann, III, “Scattering from axially symmetric objects: theory, FEMLAB model implementation and results,” NURC Technical Report No. NURC-FR-2006-010, April 2006.

²⁷P. R. Stepanishen, “Acoustic axisymmetric radiation and scattering from bodies of revolution using the internal source density and Fourier methods,” *J. Acoust. Soc. Am.* **102**, 726–732 (1997).

²⁸A. Tesei, M. Zampolli, J. Fawcett, and D. S. Burnett, “Verification of a 3-D structural-acoustic finite-element tool against thin-shell scattering models,” *Proceedings of the 7th European Conference on Underwater Acoustics, ECUA 2004*, edited by D. G. Simons, Delft, The Netherlands, 5–8 July 2004, pp. 431–436.

²⁹D. Chu, “Exact solution for a density contrast shallow water wedge using normal coordinates,” *J. Acoust. Soc. Am.* **87**, 2442–2450 (1990).

³⁰J. Fawcett, “The computation of the scattered pressure field from a cylinder embedded between two half-spaces,” *J. Acoust. Soc. Am.* **99**, 2435–2438 (1996).

³¹F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, Springer, New York, 2000).

³²R. Jeans and I. C. Mathews, “The wave superposition method as a robust technique for computing acoustic fields,” *J. Acoust. Soc. Am.* **92**, 1156–1166 (1992).

Passive reverberation nulling for target enhancement

H. C. Song,^{a)} W. S. Hodgkiss, W. A. Kuperman, K. G. Sabra, and T. Akal
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238, USA

M. Stevenson
NATO Undersea Research Centre, La Spezia, Italy

(Received 19 December 2006; revised 29 May 2007; accepted 25 September 2007)

Echo-to-reverberation enhancement previously has been demonstrated using time reversal focusing when knowledge of the channel response between a target and the source array elements is available. In the absence of this knowledge, direct focusing is not possible. However, active reverberation nulling still is feasible given observations of reverberation from conventional source array transmissions. For a given range of interest, the response between the source array elements and the dominant sources of boundary reverberation is provided by the corresponding reverberation from this range. Thus, an active transmission can be projected from the source array which minimizes the energy interacting with the boundaries at a given range while still ensonifying the waveguide between the boundaries. As an alternative, here a passive reverberation nulling concept is proposed. In a similar fashion, the observed reverberation defines the response between the source array elements and the dominant sources of boundary reverberation at each range and this is used to drive a range-dependent sequence of projection operators. When these projection operators subsequently are applied to the received data vectors, reverberation can be diminished. The improvement in target detectability is demonstrated using experimental data with an echo repeater simulating the presence of a target. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2799508]

PACS number(s): 43.30.Gv, 43.30.Hw [DRD]

Pages: 3296–3303

I. INTRODUCTION

Reverberation from rough ocean boundaries often limits the performance of active sonar systems in shallow water environments.¹ Recently, a time reversal (TR) approach for reverberation mitigation has been investigated in theory² and subsequently demonstrated experimentally³ to reduce prominent reverberation returns by retransmission from a time reversal mirror (TRM). The basic idea is that a time-gated portion of reverberation provides the transfer function between the TRM and the sea surface/bottom boundaries for a given range cell, which can be exploited either for focusing back to the corresponding interface^{4,5} or reverberation nulling.^{2,3} The motivation behind reverberation nulling is to minimize the acoustic energy incident on the boundaries in a specific range cell while still projecting energy into the water column for the ensonification of targets.

Reverberation nulling occurs naturally with the TR method when focusing acoustic energy on a target due to shadowing of the boundaries below and above the focus in a waveguide (see Fig. 2(b) in Ref. 6). Indeed, Kim *et al.*⁶ demonstrated the potential of echo-to-reverberation enhancement using the TR approach over a simple broadside (BS) transmission. First, with a probe source (PS) at 4.7 km range from the TRM in 110-m-deep water, the increased level of ensonification at the PS and the decreased level of reverberation at the TRM were measured separately and combined to esti-

mate the monostatic echo-to-reverberation enhancement. Second, improved target detectability was confirmed bistatically using an artificial target of unknown target strength (TS) located next to the PS. Here, “monostatic” and “bistatic” refer to the use of a receiving array collocated with or separated from the source array. In Sec. III, we will show the results of a monostatic echo-to-reverberation enhancement experiment as opposed to our earlier bistatic demonstration.⁶

Although TR focusing is a useful demonstration of waveguide physics, it is unlikely that the target actually will be in the immediate vicinity of a PS.⁷ In this case, it is not possible to focus directly on the target. If a priori knowledge of target range is available, however, focusing on a target still is feasible using an iterative time reversal approach without a PS by iteratively retransmitting a selected portion of data which contains the target echo.^{5,8}

Instead of boosting the target echo by TR focusing, we also can achieve echo-to-reverberation enhancement by actively reducing the reverberation given observations of reverberation from conventional source array transmissions. In this paper, the terms “active” and “passive” refer to the TRM: “active” when a TRM attempts to focus on a target or null out reverberation by transmitting an environmentally dependent excitation vector signal while “passive” when a TRM simply transmits conventional BS signals followed by processing of the returned time series. For a given range of interest, the response between the source array elements and the dominant sources of boundary reverberation is provided by the corresponding reverberation from this range. Thus, an active transmission can be projected from the source array which minimizes the energy interacting with the boundaries

^{a)}Author to whom correspondence should be addressed. Electronic mail: hcsong@mpl.ucsd.edu

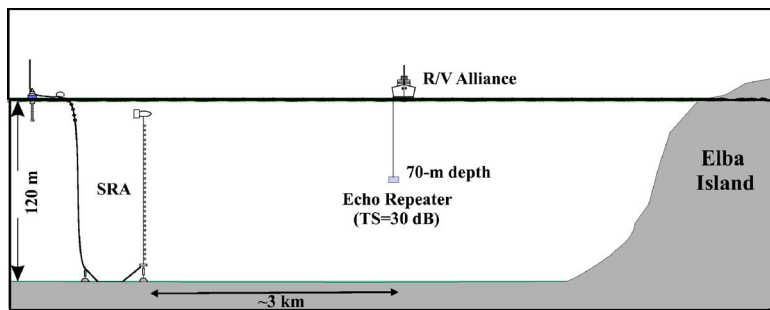


FIG. 1. (Color online) Experimental configuration for reverberation measurements carried out north of Elba Island off the west coast of Italy. The echo repeater (ER) was deployed from the R/V Alliance at 70 m depth and about 3 km away from the SRA. The ER was drifting along with the R/V Alliance.

at a given range while still ensonifying the waveguide between the boundaries.² While active reverberation nulling also has been demonstrated experimentally,³ the range over which boundary interaction is suppressed is limited and thus different excitation vectors must be transmitted for each range unless the specific range of a target is known.

As an alternative, we consider a passive reverberation nulling concept which operates only on the returning reverberation from conventional source array transmissions without a PS and a priori knowledge of target range. In a similar fashion, the observed reverberation defines the response between the source array elements and the dominant sources of boundary reverberation at each range. A collection of target-free responses can be used to drive a range-dependent sequence of projection operators based on the eigendecomposition of the data covariance matrix.^{9,10} When these projection operators subsequently are applied to the received data in the presence of a target, reduction of reverberation results in enhancing the target echo. This approach essentially is equivalent to the eigenvector-based projection method in radar¹⁰⁻¹³ where strong interferences are suppressed by projecting the array data onto the subspace that is orthogonal to the interference subspace. In reverberation-limited active sonar conditions,¹ reverberation can be treated as a strong interference although it represents a collection of distributed sources at the interface rather than a single point source. Despite the analogy, we are not aware of any previous work reported in the literature applying the projection approach to low or mid-frequency reverberation nulling. Hodgkiss and Alexandrou¹⁴ proposed an adaptive least-squares algorithm to reject high-frequency under-ice reverberation, while Cox¹⁵ exploited Doppler and applied an adaptive beamforming algorithm with a towed horizontal array to detect a weak moving target in the presence of strong bottom reverberation.

This paper is organized as follows. Section II describes the reverberation experiment conducted during FAF-04. Section III demonstrates monostatic echo-to-reverberation enhancement of the TR approach using an echo repeater simulating a target. Finally, Sec. IV demonstrates improved target detectability resulting from the passive reverberation nulling approach.

II. REVERBERATION EXPERIMENT

A time reversal experiment (FAF-04) was conducted jointly with the NATO Undersea Research Center in July 2004 both north and south of Elba Island, off the west coast

of Italy. The reverberation experiment reported in this paper was carried out in a flat region of 120-m-deep water north of Elba on July 22 (JD204). The source-receive array (SRA or TRM) consists of 29 transducers spanning a 78 m aperture (2.786 m interelement spacing from 32 to 110 m) as shown in Fig. 1. The maximum source level of each element was 178 dB *re* 1 μ Pa. An echo repeater (ER) simulating a target was deployed at 70 m depth which was drifting along with the R/V Alliance. The target strength (TS) of the ER was set to 30 dB. A sound speed profile collected during the experiment is displayed in Fig. 2 featuring an extended thermocline down to 70 m depth. We confine our interest to narrowband continuous wave (cw) pulses as in our previous papers.^{2,3,6}

III. ECHO-TO-REVERBERATION ENHANCEMENT

In this section, monostatic echo-to-reverberation enhancement is demonstrated using the echo repeater shown in Fig. 1. First, consider a conventional BS transmission as a base line.

The SRA simply transmits a 100 ms cw pulse at 3500 Hz on all elements simultaneously and that BS transmission is captured by the ER. Next, the SRA repeats the BS transmission while the ER transmits the previously captured signal simultaneously with an appropriate time delay simulating a target echo. Immediately after the BS transmission,

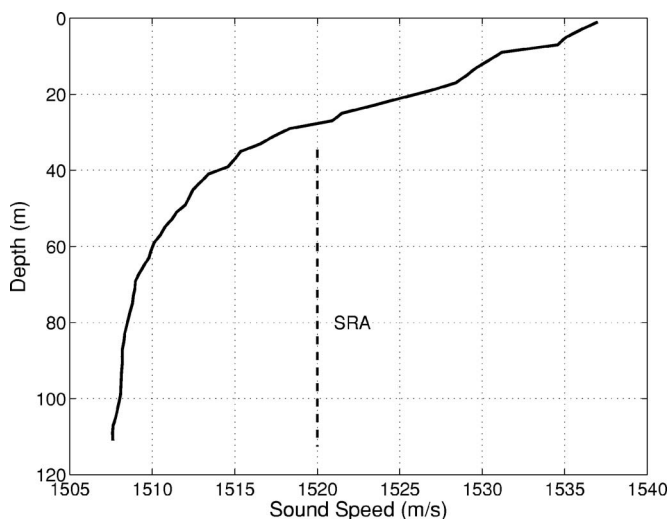


FIG. 2. The sound speed profile measured during the reverberation experiment on July 22, 2004 (JD204) along with the depth coverage of the source-receive array (SRA). The downward refracting profile with an extended thermocline down to 70 m depth suggests that the reverberation return mainly is due to interaction with the bottom.

the SRA starts recording the monostatic reverberation return which contains the target echo as well. An example of the measured monostatic reverberation field at the SRA plus ER transmission is shown in Fig. 3(a) when the ER was about 3 km away from the SRA. Note that three of the SRA transducers (Channels 7, 22 and 25 from the bottom) were disabled during the experiment indicated in thick blue horizontal lines in Figs. 3(a) and 3(b).

Next, consider the case of TR implementing active focusing. A 100 ms cw pulse at 3500 Hz is transmitted initially from the ER (acting as a PS), and is received and time reversed by the SRA. The time-reversed signal from the SRA then is refocused on and captured by the ER. As in the BS transmission, the SRA repeats the TR transmission while the ER transmits the captured signal with an appropriate time delay. The monostatic reverberation field plus ER transmission recorded by the SRA is shown in Fig. 3(b). The presence of a target around 3.7 s corresponding to the ER range of 3 km is indicated clearly as opposed to Fig. 3(a) where the target echo is hardly visible. Note that the dynamic range of Fig. 3(a) is 5 dB higher than that of Fig. 3(b) due to the difference in the transmitted level. In other words, approximately 5 dB less energy is transmitted by the SRA for the TR transmission compared to the conventional BS transmission as previously observed in Ref. 6.

Finally, Fig. 3(c) shows a comparison between the BS and TR transmissions. The levels are obtained by incoherent averaging of the 26 active channels. The TR level (red) has been increased by 5 dB to compensate for the lower level transmission by the SRA. The echo-to-reverberation enhancement is about 7–8 dB over the BS transmission in addition to an overall reverberation reduction of 5 dB. This result is consistent with the one reported in Ref. 6 where the increased ensonification level at the target (5 dB) and the decreased reverberation level (3 dB) were measured separately without using an echo repeater.

IV. PASSIVE REVERBERATION NULLING

Although the active TR focusing approach shows a significant echo-to-reverberation enhancement, it is assumed that the channel response from a target is known to the SRA by using a probe source in the vicinity of the target.^{6,7} If a priori knowledge of target range is available, either active reverberation nulling² or an iterative time reversal approach⁸ still can be employed to enhance the target echo as described in Sec. I. In this section, we consider the more typical situation where neither a PS nor a priori knowledge of target range is available. Here we propose a passive reverberation nulling approach to improve target detection from BS transmissions containing a target echo as shown in Fig. 3(a). The approach essentially is equivalent to the orthogonal projection methods found in adaptive beamforming algorithms^{10,13} where strong interferers are suppressed.

During the FAF-04 reverberation experiment described in Sec. II, a 100 ms cw BS transmission was carried out every 15 s over approximately 4 min (JD204 11:20:15–11:23:45) while the ER at 3 km range retransmitted the captured echo every other BS transmission (i.e., 30 s). As a

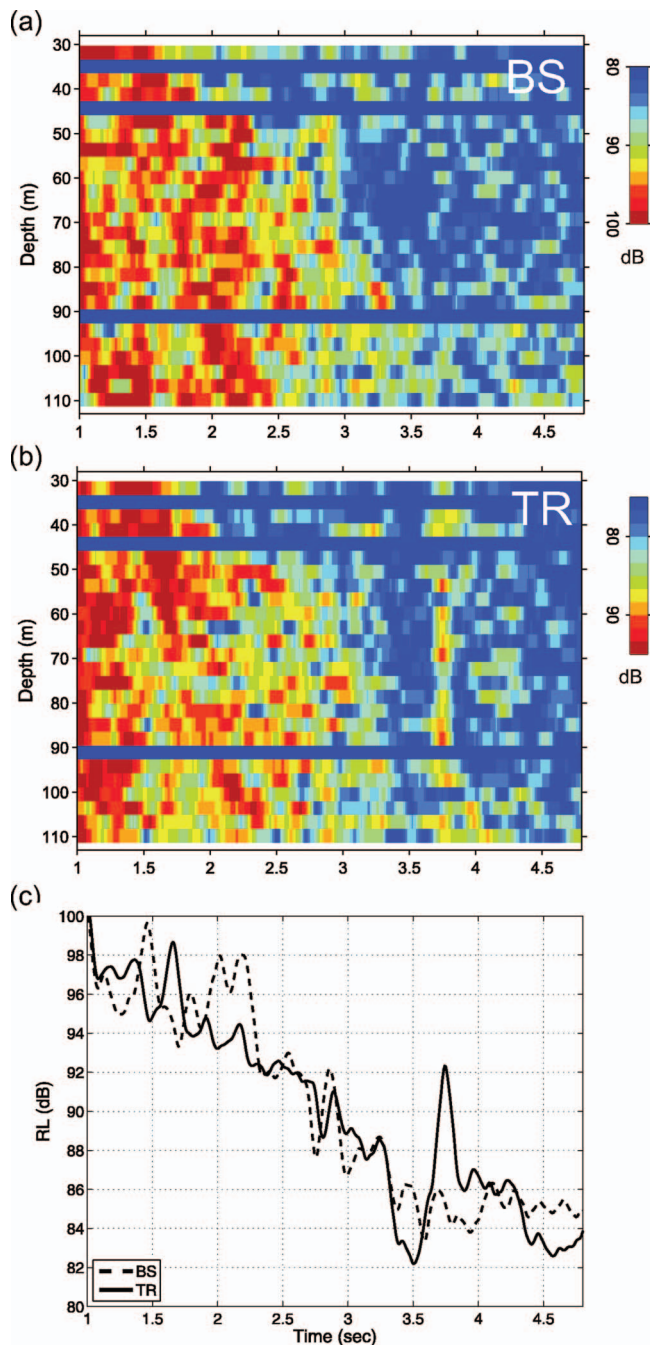


FIG. 3. Measured backscattered field at the SRA when the ER transmission simulating a target is present: (a) BS transmission and (b) TR transmission. Note that the BS dynamic range is 5 dB higher than that of TR due to the difference in the transmitted level. Plot (c) shows the corresponding reverberation level incoherently averaged across the array elements where the TR curve is offset by 5 dB for comparison purposes. The TR result clearly shows the presence of a target around 3.7 s corresponding to the ER range of 3 km at this time.

result, a total of 16 BS reverberation returns were collected, half of which included the target echo as well. Without any signal processing, the incoherent reverberation levels are shown in the upper left and right panels of Fig. 4: (a) reverberation alone and (b) echo plus reverberation. Interestingly, the target at around 3.9 s happens to occur very close to the strong reverberation return around 4.1 s, making it difficult to determine the presence of the target in Fig. 4(b). We also

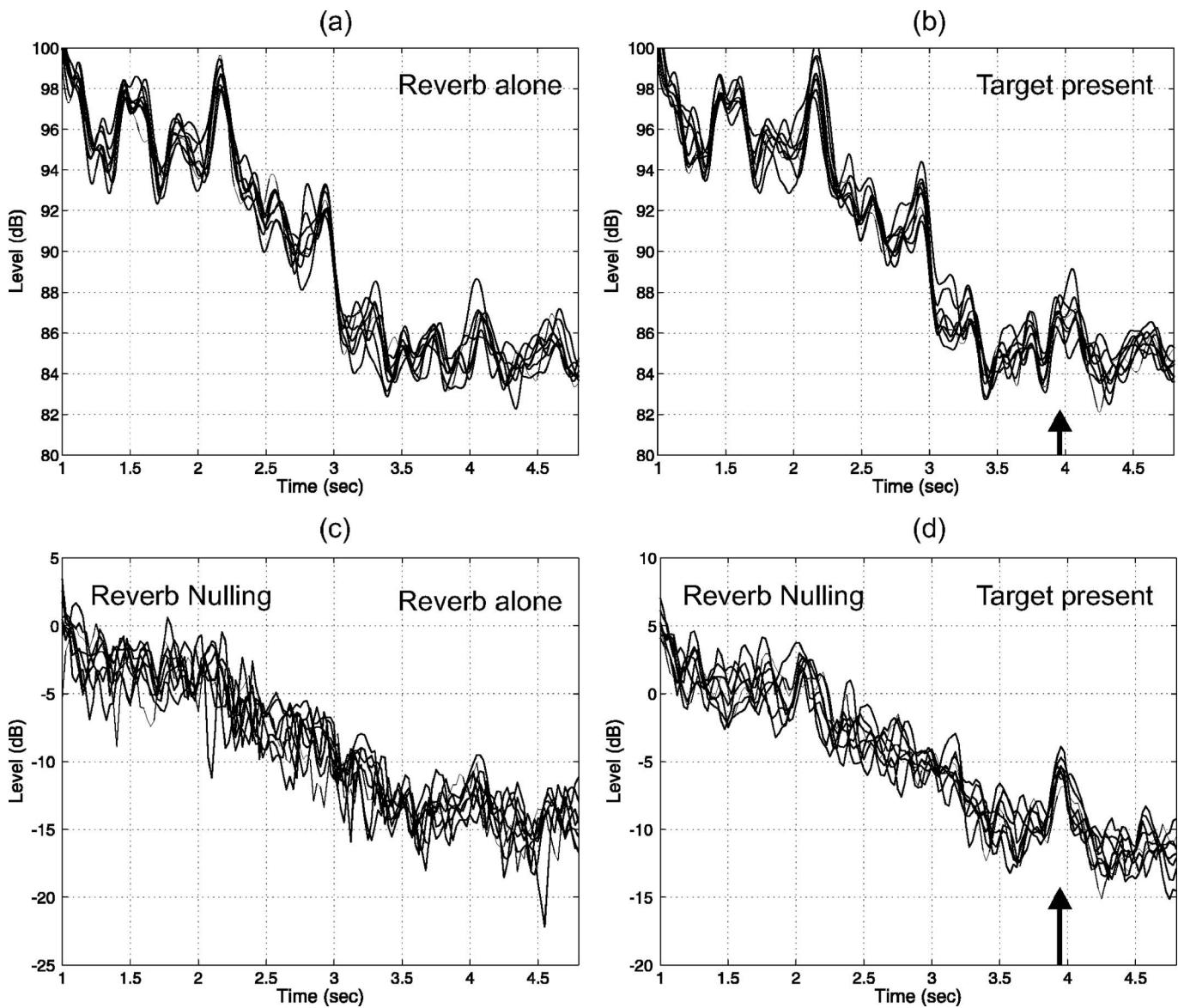


FIG. 4. Reverberation returns from BS transmissions before (top) and after (bottom) passive reverberation nulling (see text for details). The upper two plots display the incoherent average of reverberation levels: (a) reverberation alone and (b) echo plus reverberation. The target echo should be around 3.9 s (arrow). The lower two plots show the results after reverberation nulling has been applied to: (c) reverberation alone when $\mathbf{r}=\mathbf{x}$ and (d) echo plus reverberation when $\mathbf{r}=\mathbf{s}+\mathbf{x}$. Plot (d) shows improved target detectability resulting from reverberation nulling over plot (b) before nulling. Note that after reverberation nulling the absolute level has no meaning in the lower two plots.

will show results of target detection from echo plus reverberation returns later in Sec. IV B. Note that this portion of the experiment was carried out half an hour earlier than the monostatic echo-to-reverberation data collection reported in Sec. III where the ER is at around 3.7 s. The BS transmission loss (TL) from the SRA is displayed in Fig. 5(a) indicating that the reverberation return mainly is due to interaction with the sea floor in a downward refracting environment (see Fig. 2). Here we do not attempt to reproduce the experimental results shown in Fig. 4(a) [or Fig. 5(b)] except to capture the general features of reverberation from a propagation perspective.

A. Projection operator

The basic idea behind reverberation nulling is based on the observation that in the absence of a target \mathbf{s} , a time-

windowed segment of reverberation data transformed into the frequency domain \mathbf{x} can be represented as

$$\mathbf{x} = \mathbf{h} + \mathbf{n}, \quad (1)$$

where \mathbf{h} the transfer function vector between a TRM array and the corresponding sea surface and/or bottom boundaries for that range cell.² Note that \mathbf{h} incorporates the reverberation scattering strength as well as the source level as a scale factor. The additive white noise \mathbf{n} is assumed uncorrelated with the reverberation \mathbf{h} . Here boldface lower case letters denote N -dimensional vectors and N is the number of TRM elements.

In practice, a number of target-free reverberation returns \mathbf{x}_j (i.e., snapshots) are collected to form a data covariance matrix $\hat{\mathbf{R}}$ and its eigen-decomposition:

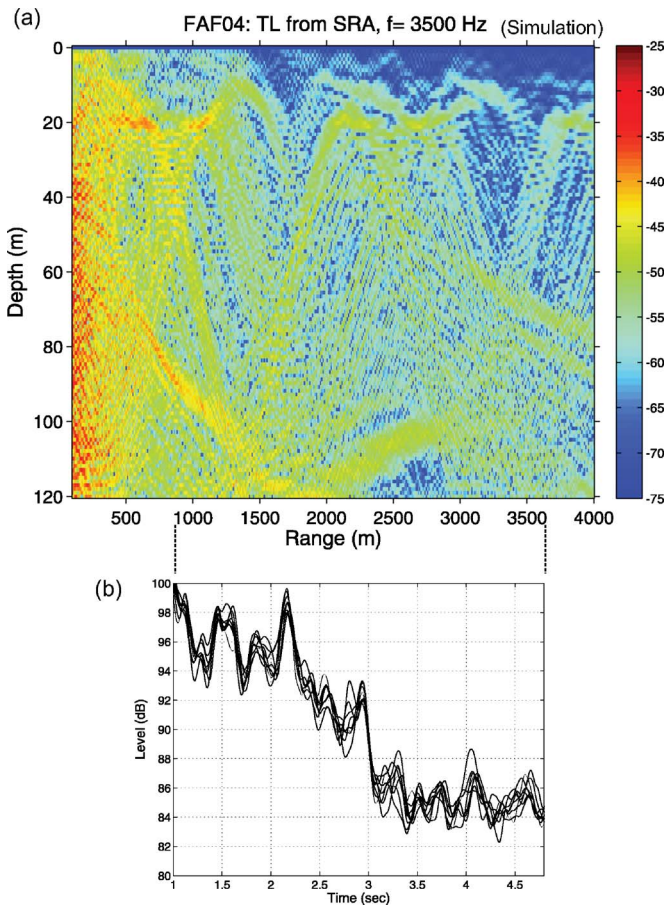


FIG. 5. (Color online) (a) BS transmission loss (TL) from the SRA using the sound speed profile displayed in Fig. 2 (b) Reverberation return shown in Fig. 4(a). The two plots are lined up approximately by the corresponding round-trip travel time. The TL plot indicates that the reverberation return mainly is due to interaction with the sea floor including the peak at around 4.1 s.

$$\hat{\mathbf{R}} = \frac{1}{L} \sum \mathbf{x}_j \mathbf{x}_j^H = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H + \mathbf{V} \mathbf{\Omega} \mathbf{V}^H, \quad (2)$$

where H denotes a conjugate transpose, L the number of snapshots, $\mathbf{\Lambda}$ a diagonal matrix containing K largest eigen-

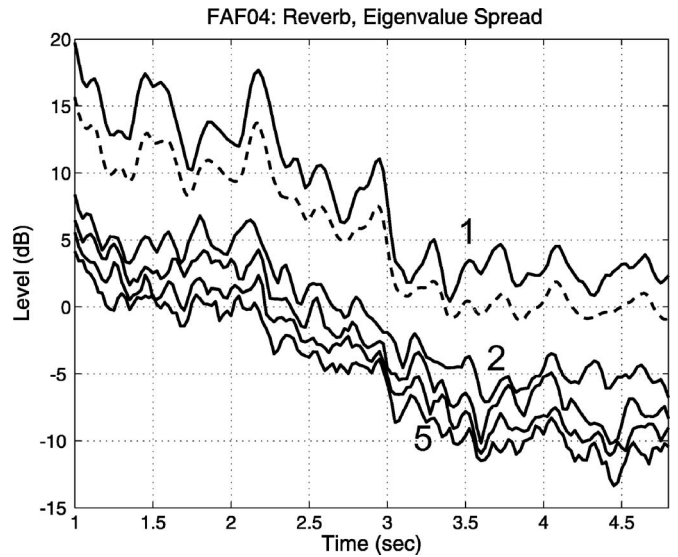
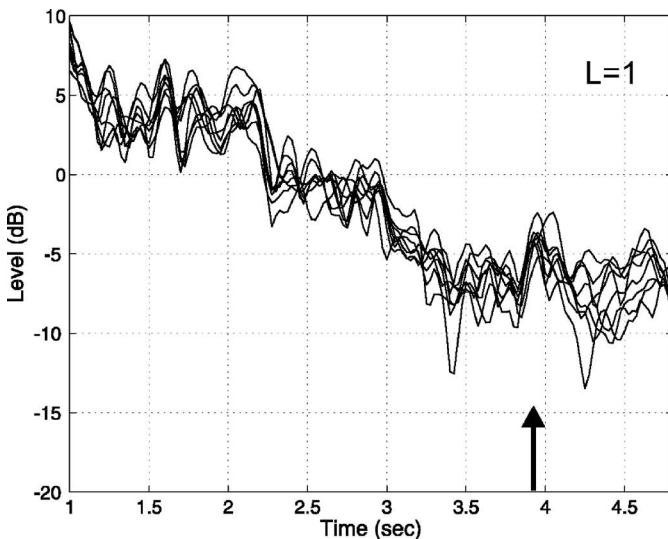


FIG. 6. Eigenvalue spread (1–5) of the data covariance matrix $\hat{\mathbf{R}}$ from a total of eight ($L=8$) target-free snapshots (BS transmissions) as a function of time. The first eigenvalue (top) is about 10 dB higher than the second. As a reference, the average of the incoherent energy from Fig. 4(a) is superimposed with an offset of 80 dB (dotted line) below the first eigenvalue. The similarity between the first eigenvalue and the average indicates that the first eigenvalue captures the reverberation component.

values, and \mathbf{U} an $N \times K$ matrix of the corresponding eigenvectors (\mathbf{u}_i , $i=1, \dots, K$) spanning the K -dimensional reverberation subspace. Some of the eigenvalues may be zero so that $\hat{\mathbf{R}}$ may be less than full (e.g., for a snapshot-deficient case when $L < N$).¹⁶

For reverberation nulling, a complementary subspace that is orthogonal to the reverberation subspace is obtained in terms of an $N \times N$ projection matrix \mathbf{P} such that¹³

$$\mathbf{P} = \mathbf{I} - \mathbf{U} \mathbf{U}^H = \mathbf{I} - \sum_{i=1}^K \mathbf{u}_i \mathbf{u}_i^H, \quad (3)$$

where \mathbf{I} is an identity matrix. It should be mentioned that the projection matrix can be deduced from the inverse of $\hat{\mathbf{R}}$

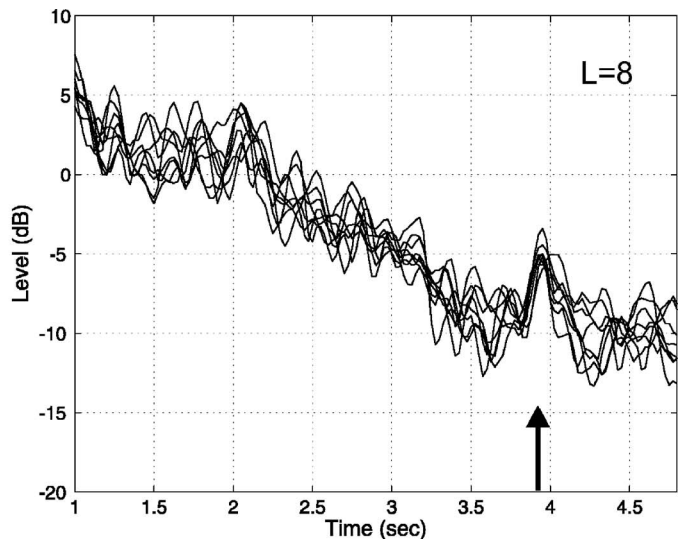


FIG. 7. Echo-to-reverberation enhancement after reverberation nulling for different numbers of snapshots: (a) $L=1$ and (b) $L=8$. The improvement with more snapshots is visible. $K=1$ is used.

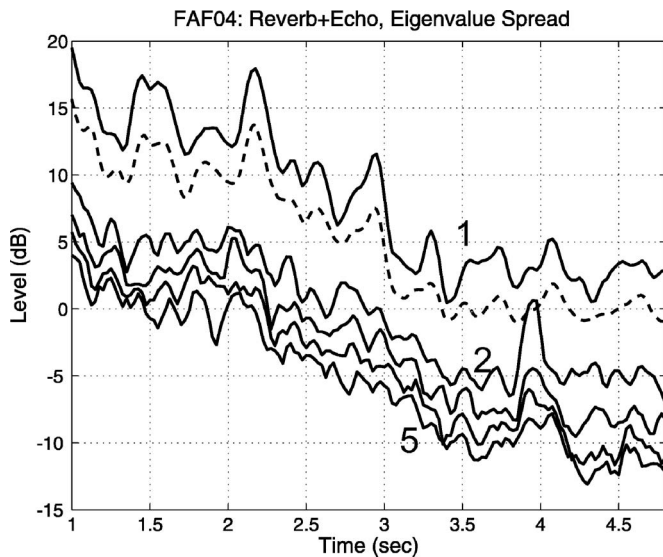


FIG. 8. Eigenvalue spread (1–5) of the data covariance matrix $\hat{\mathbf{R}}$ from a total of eight ($L=8$) echo plus reverberation returns as a function of time. The first eigenvalue looks similar to the one shown in Fig. 6. The incoherent average of the energy from Fig. 4(a) is superimposed with an offset of 80 dB (dotted line). On the other hand, the second eigenvalue shows a strong peak at 3.9 s at the range of the target.

which arises in adaptive algorithms when the $(N-K)$ small eigenvalues are negligible as compared to the K significant eigenvalues.⁹

Ideally the rank of $\hat{\mathbf{R}}$ will be equal to 1 (i.e., $K=1$) in a stationary environment since reverberation is treated as a single interference in Eq. (1). Thus, the first eigenvector \mathbf{u}_1 will be proportional to the transfer function vector \mathbf{h} . For a single snapshot case ($L=1$), \mathbf{u}_1 is assumed approximately proportional to the transfer function vector \mathbf{h} for a high reverberation-to-noise ratio.^{2,3} Our experiment produced eight target-free snapshots ($L=8$) while the ER simulating a target was drifting along with the R/V Alliance. For a small

number of snapshots, the first eigenvector alone may not necessarily represent the reverberation subspace due to eigenvalue spreading.¹⁰

Since target-free observations are used to construct the projection matrix and the number of snapshots ($L=8$) is much smaller than the number of degrees of freedom (DOF) equal to the number of active array elements ($N=26$), any number of $K \leq L$ can be used for the reverberation subspace which is confirmed in Sec. IV B. This is not surprising because the 26-element array covers the water column and provides the spatial diversity¹⁷ which allows for simultaneous focusing (or localization) on a target within the water column and the bottom interface. In other words, all the target-free reverberation is confined to the bottom interface and the remaining $(N-L)$ DOF subspace is sufficient to capture the signal component which is orthogonal to the reverberation. When the echo plus reverberation returns are used to construct the projection matrix, however, K must be chosen carefully as discussed in the next subsection. Below we investigate the impact of L , K , as well as N on the detection performance using the 16 BS transmission data.

B. Target detection

The projection matrix \mathbf{P} can be applied directly to echo plus reverberation data in order to improve the target detectability. Since a priori knowledge of target range is not assumed, each segment of the reverberation data is transformed into the frequency domain by fast Fourier transform and is advanced by a sliding (moving) window. Specifically, the duration of the time window Δ is selected 1.3 times the pulse length $\tau=100$ ms as suggested in Ref. 2 and each segment is advanced by 25 ms.

Figure 6 shows the first five eigenvalues of the data covariance matrix $\hat{\mathbf{R}}$ as a function of time (or equivalently range) from the eight target-free BS transmissions ($L=8$) displayed in Fig. 4(a). As expected for a strong single interfer-

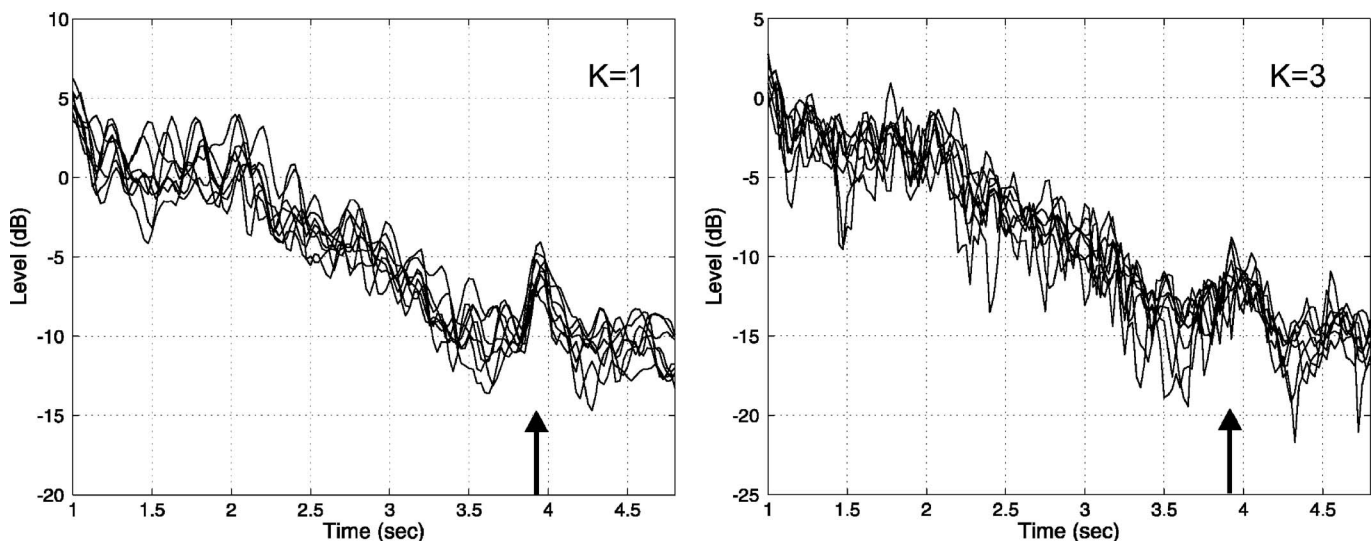


FIG. 9. Echo-to-reverberation enhancement after reverberation nulling from eight ($L=8$) echo plus reverberation snapshots: (a) $K=1$ and (b) $K=3$. $K=1$ still provides target detection comparable to one in Fig. 4(d) which uses target-free observations. However, the detection performance deteriorates with $K=3$ by rejecting the signal component as well.

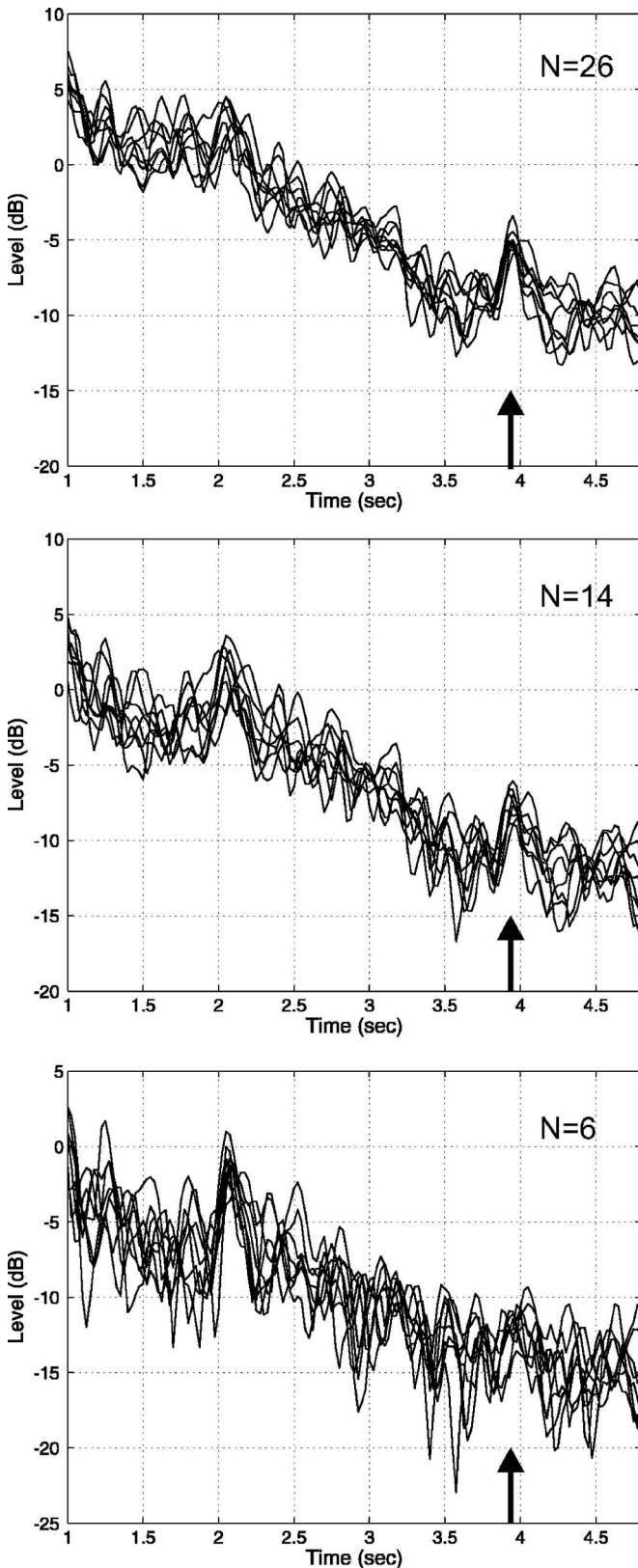


FIG. 10. Echo-to-reverberation enhancement after reverberation nulling for different numbers of receive array elements: $N=26$ (full), $N=14$ (half), and $N=6$ (quarter). The array elements are selected from the bottom. $K=1$ and $L=8$ from target-free observations. Half the array $N=14$ still shows a target echo at 3.9 s while a false target emerges at around 2.1 s.

ence, the first eigenvalue is dominant and is about 10 dB higher than the second eigenvalue. The dotted line represents the average level of the incoherent energy in Fig. 4(a) super-

imposed with an offset of 80 dB below the first eigenvalue. It follows closely the first eigenvalue confirming that the first eigenvalue captures the reverberation. Due to the large number of DOF ($N=26$) and small number of snapshots ($L=8$) as discussed in Sec. IV A, it is found that any number of $K \leq 8$ provides similar performance.

The lower two plots of Fig. 4 show the results after reverberation nulling $\|\mathbf{Pr}\|$ using $K=3$: (c) reverberation alone when $\mathbf{r}=\mathbf{x}$ and (d) echo plus reverberation when $\mathbf{r}=\mathbf{s}+\mathbf{x}$. Two observations can be made. First, the prominent reverberation returns in Fig. 4(a) have diminished and are flattened out in Fig. 4(c). Second, plot (d) clearly demonstrates improved target detectability after nulling at around 3.9 s (about 5 dB over the background) as compared to plot Fig. 4(b) before reverberation nulling. Note that after the projection operation the absolute level (dB) has no meaning in the lower two plots. It also should be mentioned that the passive reverberation nulling approach can be applied even when multiple targets are present at various ranges.

The impact of the number of snapshots L is illustrated in Fig. 7 which compares results when (a) $L=1$ and (b) $L=8$. Since only a single eigenvalue is available for $L=1$, $K=1$ is applied to both cases for comparison purposes. While a single snapshot still provides reasonable performance suggesting a high reverberation-to-noise ratio of the data, the improvement using more snapshots is evident in Fig. 7(b).

Up to now, it is assumed that the target-free observations are available to construct the data covariance matrix $\hat{\mathbf{R}}$ and then the projection matrix for nulling. In practice, however, we may not know whether the measured reverberation contains a target echo. Consider the case when we have a total of eight echo plus reverberation returns. The observation vector then can be written as

$$\mathbf{x} = \mathbf{h} + \mathbf{s} + \mathbf{n}. \quad (4)$$

If the eigenvalues corresponding to reverberation and signal are separable, a projection matrix still can be derived by selecting an appropriate number of K such that the K -dimensional subspace spans the reverberation alone but excludes the signal component.¹⁰ The first five eigenvalues are shown in Fig. 8 from the eight ($L=8$) echo plus reverberation returns. The first eigenvalue looks about the same as one in Fig. 6. On the other hand, the second eigenvalue shows a strong peak around at 3.9 s. This suggests that the first eigenvector $K=1$ spans the reverberation space while the second eigenvector represents the target space. Figure 9 compares results when (a) $K=1$ and (b) $K=3$ are employed to construct the projection matrix which is then applied to the individual echo plus reverberation returns. $K=1$ shows target detection comparable to the result shown in Fig. 4(d) which uses target-free observations. On the other hand, the detection performance deteriorates with $K=3$ since the resulting projection matrix rejects the target component as well.

Finally, we investigate the impact of the number of array elements used for processing the received time series. Recall that the full source array was used for the BS transmissions. Figure 10 shows the performance when $N=26$ (full), $N=14$ (half), and $N=6$ (quarter). The array elements are selected

from the bottom. The eight ($L=8$) target-free snapshots are used to construct the data covariance matrix as before and $K=1$. Half the array ($N=14$) still shows target detection at 3.9 s while a false target emerges at around 2.1 s as well. With a further decrease down to $N=6$, the target echo eventually disappears and the distinct peak at 2.1 s likely would lead to a false alarm.

V. CONCLUSION

Bottom backscattering potentially can be used as a surrogate probe source for reverberation nulling. A time-gated portion of the reverberation return from conventional BS transmissions provides an estimate of the transfer function between a TRM array and the corresponding range cell boundaries. Active reverberation nulling projects an excitation weight vector that is in the complementary subspace orthogonal to the focusing vector, minimizing the acoustic energy interacting with the boundaries in a specific range cell. However, the range over which boundary interaction is suppressed is limited and different excitation vectors must be transmitted for each range of interest. Alternatively, passive reverberation nulling applies an orthogonal projection onto the echo plus reverberation data to suppress the reverberation component without need for a specialized active retransmission. The signal processing employed essentially is equivalent to the orthogonal projection method developed in adaptive beamforming algorithms. The improved target detection resulting from passive reverberation nulling is demonstrated experimentally with reverberation data and an echo repeater simulating a target in shallow water.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research under Contract No. N00014-01-D-0.043-D06.

- ¹R. Urlick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1983).
- ²H. C. Song, S. Kim, W. Hodgkiss, and W. A. Kuperman, "Environmentally adaptive reverberation nulling using a time reversal mirror," *J. Acoust. Soc. Am.* **116**, 763–768 (2004).
- ³H. C. Song, W. Hodgkiss, W. A. Kuperman, P. Roux, T. Akal, and M. Stevenson, "Experimental demonstration of adaptive reverberation nulling using time reversal," *J. Acoust. Soc. Am.* **118**, 1381–1387 (2005).
- ⁴J. F. Lingeitch, H. C. Song, and W. A. Kuperman, "Time reversed reverberation focusing in a waveguide," *J. Acoust. Soc. Am.* **111**, 2609–2614 (2002).
- ⁵K. G. Sabra, P. Roux, H. C. Song, W. S. Hodgkiss, W. Kuperman, T. Akal, and M. Stevenson, "Experimental demonstration of iterative time-reversed reverberation focusing in a rough waveguide: Application to target detection," *J. Acoust. Soc. Am.* **120**, 1305–1314 (2006).
- ⁶S. Kim, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, G. F. Edelmann, and T. Akal, "Echo-to-reverberation enhancement using a time reversal mirror," *J. Acoust. Soc. Am.* **115**, 1525–1531 (2004).
- ⁷H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "A time-reversal mirror with variable range focusing," *J. Acoust. Soc. Am.* **103**, 3234–3240 (1998).
- ⁸H. C. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal, C. Ferla, and D. R. Jackson, "Iterative time reversal in the ocean," *J. Acoust. Soc. Am.* **105**, 3176–3184 (1999).
- ⁹H. Cox and R. Pitre, "Robust dmr and multi-rate adaptive beamforming," in *Proceedings of 31st Asilomar Conference*, 920–924 (1997).
- ¹⁰H. Cox, "Multi-rate adaptive beamforming (MRABF)," in *Sensor Array and Multichannel Signal Processing Workshop*, 306–309 (2000).
- ¹¹T. Citron and T. Kailath, "Eigenvector methods and beamforming—a first approach," in *Proceedings of 17th Asilomar Conference* (1983).
- ¹²I. Kirsteins and D. Tufts, "On the probability density of signal-to-noise ratio in improved adaptive detector," in *Proceedings of ICASSP-85*, 572–575 (1985).
- ¹³A. Gershman, V. Turchin, and V. Zverev, "Experimental results of localization of moving underwater signal by adaptive beamforming," *IEEE Trans. Signal Process.* **43**, 2249–2257 (1995).
- ¹⁴W. Hodgkiss and D. Alexandrou, "Under-ice reverberation rejection," *IEEE J. Ocean. Eng.* **OE-10**, 285–289 (1985).
- ¹⁵H. Cox, "Space-time processing for suppression of bottom reverberation," in *Proceedings of 29th Asilomar Conference*, 1296–1299 (1996).
- ¹⁶H. C. Song, W. A. Kuperman, W. Hodgkiss, P. Gerstoft, and J. Kim, "Null broadening with snapshot-deficient covariance matrices in passive sonar," *IEEE J. Ocean. Eng.* **28**, 250–261 (2003).
- ¹⁷P. Roux, W. Kuperman, W. Hodgkiss, H. Song, T. Akal, and M. Stevenson, "Non-reciprocal implementations of time reversal in the ocean," *J. Acoust. Soc. Am.* **116**, 1009–1015 (2004).

Determining dominant scatterers of sound in mixed zooplankton populations

Andone C. Lavery^{a)}

Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Peter H. Wiebe

Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Timothy K. Stanton

Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Gareth L. Lawson

Hopkins Marine Station of Stanford University, Pacific Grove, California 93950

Mark C. Benfield

Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana 70803

Nancy Copley

Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 6 April 2006; revised 28 August 2007; accepted 7 September 2007)

High-frequency acoustic scattering techniques have been used to investigate dominant scatterers in mixed zooplankton populations. Volume backscattering was measured in the Gulf of Maine at 43, 120, 200, and 420 kHz. Zooplankton composition and size were determined using net and video sampling techniques, and water properties were determined using conductivity, temperature, and depth sensors. Dominant scatterers have been identified using recently developed scattering models for zooplankton and microstructure. Microstructure generally did not contribute to the scattering. At certain locations, gas-bearing zooplankton, that account for a small fraction of the total abundance and biomass, dominated the scattering at all frequencies. At these locations, acoustically inferred size agreed well with size determined from the net samples. Significant differences between the acoustic, net, and video estimates of abundance for these zooplankton are most likely due to limitations of the net and video techniques. No other type of biological scatterer ever dominated the scattering at all frequencies. Copepods, fluid-like zooplankton that account for most of the abundance and biomass, dominated at select locations only at the highest frequencies. At these locations, acoustically inferred abundance agreed well with net and video estimates. A general approach for the difficult problem of interpreting high-frequency acoustic scattering in mixed zooplankton populations is described. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2793613]

PACS number(s): 43.30.Pc, 43.30.Sf, 43.30.Ft [KGF]

Pages: 3304–3326

I. INTRODUCTION

High-frequency acoustic scattering techniques provide a unique remote sensing capability to rapidly and synoptically investigate zooplankton distributions on similar spatial and temporal scales to the physical processes influencing them. Zooplankton are key components of pelagic food webs and yet determining their temporal and spatial distributions and abundances is an ongoing challenge. Zooplankton distributions are inherently complex, patchy at many scales, and influenced by physical processes at many different scales, from small turbulence scales (mm to cm) (Rothschild and

Osborn, 1988; Seuront *et al.*, 2001) to basin scales (tens to hundreds of km) (Haury *et al.*, 1978). Adding further complexity, zooplankton aggregations are often highly heterogeneous and some zooplankton are active swimmers, resulting in distributions that are not simply a passive response to the physical processes influencing them, such as turbulence or advection, but an intricate coupling of behavior and physical forcing.

Over the last two decades, the use of high-frequency acoustic scattering techniques has become more routine for synoptic studies of zooplankton populations from centimeter to kilometer scales and across seasonal time scales (Wiebe *et al.*, 1996, 1997; Brierley *et al.*, 1998; Pieper *et al.*, 2001; Lawson *et al.*, 2004), a task not as easily achievable by traditional net or optical sampling techniques alone. Although

^{a)}Author to whom correspondence should be addressed. Electronic mail: alavery@whoi.edu

traditional single-frequency echosounders are frequently used for visualizing zooplankton populations, there remain inherent difficulties associated with the interpretation of the acoustic scattering returns even when direct and coincident measurements of the scattering sources are available. Difficulties associated with the interpretation of the acoustic returns are reduced in regions in which a single zooplankton taxon dominates the scattering, e.g., *Euphausia superba* in Antarctic regions (Lascara *et al.*, 1999), or the relative abundance of zooplankton taxa remains constant (Benfield *et al.*, 1998). Holliday and Pieper (1995) review the circumstances when single-frequency acoustic scattering returns can be successfully interpreted. Multi-frequency acoustic scattering techniques expand the range of conditions under which it is possible to interpret the acoustic data in terms of relevant biological parameters, such as animal size or abundance (Holliday and Pieper, 1980, 1995; Costello *et al.*, 1989; Pieper *et al.*, 1990; Napp *et al.*, 1993; Wiebe *et al.*, 1997; Korneliussen and Ona, 2002; Warren *et al.*, 2003; Mair *et al.*, 2005; Trevorrow *et al.*, 2005; Lawson *et al.*, 2006).

A major complicating factor in the use of high-frequency sound to characterize zooplankton is the diverse array of organisms present in the water column (Fig. 1). Zooplankton aggregations typically span a wide range of sizes, shapes, orientations, and material properties: all parameters that influence the scattering of sound and that can be difficult to quantify. In order to understand the scattering from the wide diversity of zooplankton, it is convenient to categorize zooplankton into three groups based on the boundary conditions and anatomy (Stanton *et al.*, 1994, 1998a, b): 1) weakly scattering fluid-like zooplankton, for which the density and sound speed contrasts are small relative to the surrounding water, 2) zooplankton with hard elastic shells, which scatter sound efficiently relative to fluid-like zooplankton of similar size, and 3) zooplankton with gas inclusions that can give rise to strong scattering resonances. The vast majority of zooplankton biomass and numerical abundance is made up of fluid-like zooplankton, typically copepods. However, there are many locations where fluid-like zooplankton collectively only make up a small fraction of the scattered energy, which is instead dominated by less abundant, but more efficient scatterers of sound, such as pteropods (hard elastic shell category) and siphonophores (gas-bearing category). There are also other potentially confounding factors, such as the presence of fish, squid, bubbles, and suspended sediments that may contribute to the scattering. Further compounding difficulties in the interpretation is the fact that microstructure has been shown to contribute significantly to scattering, although only under some, possibly limited, conditions (Warren *et al.*, 2003), and can be co-located with the zooplankton. As a result of the often simultaneous presence of many different types of scatterers in mixed zooplankton populations, the estimation of biologically meaningful quantities, such as animal abundance or size, from measurements of high-frequency volume backscattering, called the “inverse problem,” is generally highly under-determined.

In order to address some of the challenges associated with the interpretation of high-frequency acoustic scattering

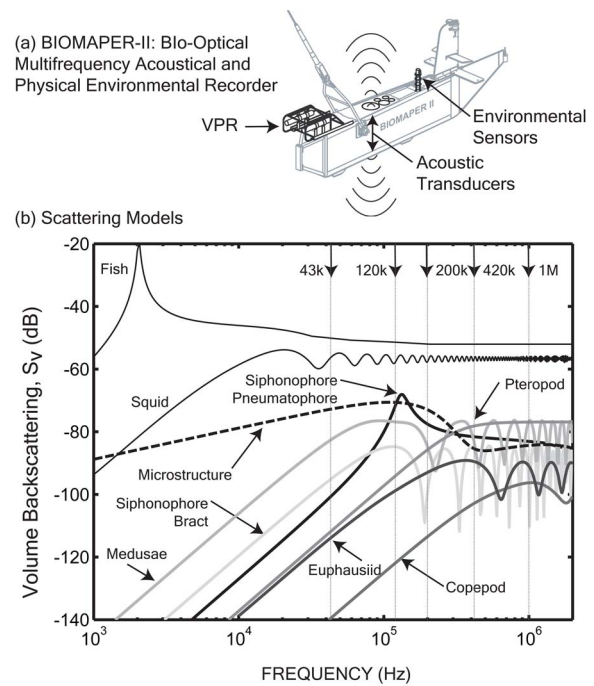


FIG. 1. (Color online) (a) BIOMAPER-II. (b) Volume backscattering as a function of frequency for the major biological scatterers observed in the GoM, assuming a numerical abundance of 1 organism/m³. The five BIOMAPER-II frequencies are indicated by vertical lines. The predictions are based on mean lengths in the 75–100 m depth range of MOCNESS 4 in Jordan Basin, GoM, in December 1999: pneumatophore=0.27 mm, siphonophore bract=9.86 mm, copepod=1.53 mm, euphausiid=9.79 mm, and pteropod=1.15 mm. The medusa length (16.53 mm) is based on the deepest net as none were observed at mid depths. A DWBA-based prolate spheroid scattering model was used for squid (length=9 cm, width = 1.2 cm, density contrast $g=1.043$, sound speed contrast $h=1.053$) (Iida *et al.*, 2006), averaged over a normal distribution of angles of orientation with a mean of 0° (corresponding to broadside incidence) and a standard deviation of 30°. The hybrid model with a 1-cm-diam gas bubble was used to represent the scattering from swim-bladdered fish. Also included are microstructure predictions using a dissipation rate of turbulent kinetic energy $\varepsilon = 1 \times 10^{-6} \text{ m}^2/\text{s}^3$ and the maximum temperature and salinity gradients observed for CTD 2 in Jordan Basin, GoM, in December 1999, representing an upper-bound estimate for the contribution to scattering from turbulent microstructure.

from mixed zooplankton populations, a towed instrument platform, BIOMAPER-II (Bio-Optical Multi-frequency Acoustical and Physical Environmental Recorder), was developed (Wiebe *et al.*, 2002) that allows coincident measurement of multi-frequency acoustic backscattering, temperature, salinity, depth, and high-resolution zooplankton video images (Fig. 1). BIOMAPER-II has been used to perform broad-scale surveys of mixed zooplankton populations in the deep basins of the Gulf of Maine (GoM) and results from these surveys are presented in this study. Supplementing the data collected with this instrument are almost-coincident net samples and conductivity, temperature, and depth (CTD) profiles.

The almost-simultaneous use of net and video sampling techniques (often referred to as “ground truthing”) with the collection of acoustic data helps constrain the otherwise highly under-determined inverse problem and increases the range of circumstances under which dominant scatterers can be identified. Comparison of observed to predicted volume scattering (based on scattering models that incorporate net

and/or video measurements of animal size, shape, and abundance), known as the “forward problem,” allows dominant water-column scatterers to be determined. Once these have been determined, volume backscattering measurements can then be used to invert for biologically relevant parameters, such as size or abundance, at least at some restricted locations.

In this study, the volume sampled acoustically with BIOMAPER-II is sufficiently large that a two-dimensional cross section, or “curtain,” of the ocean interior is imaged, and a significant fraction of the zooplankton taxa present in the water column is included in the sampling volume, typically involving multiple trophic levels. This is in contrast to zooplankton video imaging instruments, such as the Video Plankton Recorder (Davis *et al.*, 1992), and some acoustic instruments (Holliday and Pieper, 1995; Pieper *et al.*, 1990, 2001; Costello *et al.*, 1989, Napp *et al.*, 1993), which sample smaller volumes (up to four orders of magnitude smaller), require profiling, and result in one-dimensional measurements of the water column. In addition, instruments with small sampling volumes typically do not adequately sample less abundant and/or larger taxa, such as gas-bearing siphonophores, and are instead optimized for sampling small and abundant scatterers, such as copepods. Yet, typical zooplankton surveys use hull-mounted or towed acoustic platforms (such as BIOMAPER-II) with relatively large acoustic sampling volumes, and thus sample the less abundant and/or larger taxa that often dominate the scattering, at least over a range of frequencies. In addition, the less abundant and/or larger taxa are important to sample in order to understand the entire pelagic food web. However, the contribution to scattering from multiple different taxa can also lead to additional complexities in interpreting the returns. Finally, though less abundant and/or larger taxa are sampled by many net systems, net techniques often have coarse depth resolution, destroy fragile individuals, suffer from selective sampling due to avoidance or escapement, and require time-consuming analysis. As a result of the inherent limitations of any one technique, the combined use of multi-frequency acoustic, video, and net sampling techniques increases the likelihood of correctly identifying dominant scatterers and determining biologically meaningful parameters.

The specific objectives of this work are to: 1) Assess the limitations of existing scattering models and their input parameters, and the implications of these limitations to the determination of dominant scatterers in mixed zooplankton populations. To reduce discrepancies between predicted and observed volume backscattering, minor modifications are made to existing scattering models for gas-bearing and elastic-shelled zooplankton. 2) Use direct video and net sampling techniques together with the scattering models to make forward predictions of expected backscattering in order to determine dominant water-column scatterers, and deduce the distribution and variability of the dominant water-column scatterers. 3) Compare predicted volume backscattering with measured volume backscattering at multiple frequencies, locations, and depths. 4) Perform simple inversions for biologically relevant parameters in regions where the scattering is shown by the forward predictions to be dominated by a

single zooplankton taxon. 5) Make recommendations for conditions under which these methods for quantitative studies of mixed zooplankton populations can be used.

II. SCATTERING MODELS AND PARAMETERS

There are many potential water-column scattering sources, including zooplankton and micronekton (e.g., shrimp, fish, and squid), phytoplankton, suspended sediments, bubbles, and microstructure. Throughout this work, shrimp and zooplankton are referred to collectively as “zooplankton.” Accurate scattering models for different scattering sources (Fig. 1), together with information obtained from net and video sampling techniques, are key to determining dominant water-column scatterers. In this section, the models used to compare measured to predicted volume backscattering from zooplankton and microstructure are described. Sensitivity analyses of some of the necessary parameters are discussed elsewhere in the literature (e.g. Stanton and Chu, 2000; Lavery *et al.*, 2001).

A. Basic equations

For a single bounded target, the backscattering amplitude, f_{bs} , is a measure of the efficiency with which the target scatters sound in the backscattering direction, and is a function of the acoustic frequency, orientation of the target relative to the incident wave, the size and shape of the target, and the density ($g = \rho_{target}/\rho$, where ρ_{target} is the density of the target and ρ is the water density) and sound speed ($h = c_{target}/c$, where c_{target} is the sound speed of the target and c is the water sound speed) contrasts between the target and surrounding water. The far-field backscattered energy is often expressed in terms of the target strength (TS) with units of decibel (dB) relative to 1 m² and is given by $TS = 10 \log_{10} \sigma_{bs} = 10 \log_{10} |f_{bs}|^2$, where $\sigma_{bs} = |f_{bs}|^2$ is the differential backscattering cross section and differs from the often-used scattering cross section σ by a factor of 4π ($\sigma = 4\pi\sigma_{bs}$). For spherical targets of radius a , the reduced target strength (RTS) is given by $RTS = 10 \log_{10} (\sigma_{bs}/\pi a^2)$. Mean TS and RTS are defined as $\langle TS \rangle = 10 \log_{10} \langle \sigma_{bs} \rangle$ and $\langle RTS \rangle = 10 \log_{10} \langle \sigma_{bs} / \pi a^2 \rangle$.

During most field experiments, it is not the scattering from a single individual that is measured, but instead the average over many individual targets. So long as the phases from the echoes of the individual targets are random, and there is no attenuation or multiple scattering, then the average echo energy from the aggregation is equal to the sum of the echo energy from each individual, averaged over an ensemble of independent realizations, for example, averaged over animal orientation, length, or a combination of the two. The echo integration procedure results in an estimate of the volume backscattering coefficient, s_v , with units of m²/m³, which corresponds to the scattered echo energy at the receiver from the aggregation of scatterers normalized by the scattering volume. For an aggregation of zooplankton

$$s_V(f_q, d_k) = \frac{1}{V} \sum_{i=1}^{N_k} \sum_{j=1}^{M_k} \langle \sigma_{bs}^{ij}(f_q, d_k, \Psi_{ij}) \rangle, \quad (1)$$

where f_q is the frequency, V_k is the volume of water sampled in the depth range d_k , N_k is the number of zooplankton of a particular taxon in the depth range d_k , M_k is the number of zooplankton taxa in the depth range d_k , and $\langle \sigma_{bs}^{ij}(f_q, d_k, \Psi_{ij}) \rangle$ is the backscattering cross section of each individual of size i , taxon j , at frequency f_q , in the depth range d_k . The term $\langle \dots \rangle$ represents an average over angles of orientation. The parameter Ψ_{ij} includes the dependence of σ_{bs} on a number of taxon-specific parameters not explicitly written into the equation, such as shape and material properties. All parameters are assumed to be constant for all individuals of a given taxon, other than size. The volume backscattering strength is given by $S_V = 10 \log_{10} s_V$, which has units of decibel (dB) relative to an inverse meter. Multiple sources of scattering in the same sampling volume are accounted for by incoherently adding their contributions (e.g., $S_V^{\text{total}} = 10 \log_{10}(s_V^{\text{zooplankton}} + s_V^{\text{microstructure}})$).

B. Turbulent microstructure

Under certain circumstances it is possible for turbulent oceanic microstructure to result in acoustic backscattering levels comparable to those observed for zooplankton (Thorpe and Brubacker, 1983; Seim *et al.*, 1995; Seim, 1999; Goodman, 1990; Lavery *et al.*, 2003; Ross and Lueck, 2003; Warren *et al.*, 2003). Failure to account for the contribution to scattering from microstructure can lead to overestimates of zooplankton numerical abundance, as well as difficulties in interpreting frequency-dependent scattering spectra. The volume backscattering coefficient for turbulent microstructure is given by (Lavery *et al.*, 2003; Ross *et al.*, 2004)

$$s_V = \left(\frac{k^4}{K^3} \right) q \left(\frac{\nu}{\varepsilon} \right)^{\frac{1}{2}} \left\{ A^2 \chi_T \exp\left(-q \frac{K^2}{k_{BT}^2}\right) + B^2 \chi_S \exp\left(-q \frac{K^2}{k_{BS}^2}\right) + AB(\chi_T \chi_S)^{\frac{1}{2}} \exp\left(-q \frac{K^2}{k_{BTS}^2}\right) \right\}, \quad (2)$$

where k is the acoustic wave number, $K=2k$ is the Bragg wave number in the backscattering direction, $A = c^{-1}(\partial c / \partial T) + \rho^{-1}(\partial \rho / \partial T)$, and $B = c^{-1}(\partial c / \partial S) + \rho^{-1}(\partial \rho / \partial S)$. The term ε is the dissipation rate of turbulent kinetic energy [m^2/s^3], ν [m^2/s] is the molecular viscosity, and q is a universal constant ($q=3.7$, Oakey, 1982). $k_{BT} = (\varepsilon / (\nu \kappa_T^2))^{1/4}$, $k_{BS} = (\varepsilon / (\nu \kappa_S^2))^{1/4}$, and $k_{BTS} = (\varepsilon / (\nu \kappa_{TS}^2))^{1/4}$, where κ_T [m^2/s] and κ_S [m^2/s] are the molecular diffusivities for temperature and salt, and $\kappa_{TS} = (\kappa_T + \kappa_S) / 2$. The dissipation rates of temperature and salinity variance are given by $\chi_T = 2\Gamma \varepsilon N^{-2} (dT/dz)^2$ [m^2/s] and $\chi_S = 2\Gamma \varepsilon N^{-2} (dS/dz)^2$ [m^2/s], where $\Gamma=0.2$ is the mixing efficiency (Gregg, 1987), N is the buoyancy frequency ($N^2 = -g/\rho \partial \rho / \partial z$, g is the acceleration due to gravity), and dT/dz and dS/dz are the vertical temperature and salinity gradients averaged over 1 m intervals.

C. Weakly scattering fluid-like zooplankton

The scattering models and parameters used in this study for weakly scattering fluid-like zooplankton are summarized in Table I. A scattering model based on the distorted-wave Born approximation (DWBA) (Stanton *et al.*, 1998a, b; Stanton and Chu, 2000) is used for most of these zooplankton for all frequencies and angles of orientations. It is adequate to model the shape of many elongated fluid-like zooplankton as uniformly bent and tapered cylinders or prolate spheroids averaged over a distribution of angles (Stanton and Chu, 2000). Though there have been a number of studies aimed at quantifying typical angular distributions (Kils, 1981; Chu *et al.*, 1993; Endo, 1993; Miyashita *et al.*, 1996; McGehee *et al.*, 1998; Benfield *et al.*, 2000; Lawson *et al.*, 2006) and material properties for certain zooplankton (Foote, 1990; Chu *et al.*, 2000, 2003; Chu and Wiebe, 2005), these factors continue to result in great uncertainty in predicting scattering from many fluid-like zooplankton.

D. Elastic-shelled zooplankton

Elastic-shelled zooplankton, such as thecosome (i.e., shelled) pteropods, can give rise to high scattering levels compared to fluid-like zooplankton of a similar size (Stanton *et al.*, 1994). Pteropods have hard, rough, spiral, elastic, aragonite shells, with a large discontinuity called the opercular opening. The shelled pteropod found most commonly in the GoM, *Limacina retroversa*, is modeled as being spherical in shape as it generally supports only a slight elongation (length-to-width ratio $\beta_D < 3$). As water-column zooplankton are close to neutrally buoyant, pteropod shells are generally quite thin, potentially as thin as a few micrometers (Lalli and Gilmer, 1989). Little is known regarding the *in situ* swimming orientation of many pteropods, though there is evidence that they tend to swim preferentially with the opercular opening facing up (Gallager *et al.*, 1996), which could result in differences between volume backscattering measurements with upward- and downward-facing acoustic systems.

The only published scattering models for elastic-shelled zooplankton have been developed by Stanton and colleagues (Stanton *et al.*, 1994, 1998a, b, 2000). Stanton *et al.* (1994) collected laboratory scattering data for individual pteropod shells and found that a high-pass dense fluid-sphere model with an empirically derived reflection coefficient ($R=0.5$) best fit the data. This model is commonly used in the literature, yet the reflection coefficient is lower than the value based on the actual material properties for aragonite ($g=2.84$, $h=3.98$; $R=0.84$).

In later studies, Stanton *et al.* (1998b, 2000) developed ray-based scattering models that incorporated subsonic zeroth-order antisymmetric Lamb waves and the effects of shell roughness. However, these models are valid for $ka > 1$ and include a number of heuristic parameters that are unknown for the current study, and thus are inappropriate for the current study for which $ka=0.03-2.5$. Finally, Stanton *et al.* (2000) were able to reproduce averaged scattering data for elastic-shelled gastropods (benthic organisms similar in shape to pteropods) by using an averaged model based on an

TABLE I. Scattering models and associated model parameters used to predict high-frequency acoustic scattering from different fluid-like zooplankton. The DWBA uniformly-bent cylinder model was used for the majority of elongated fluid-like zooplankton (Eq. (6) in Stanton *et al.*, 1998b; Eq. (6) in Stanton and Chu, 2000). A slight tapering (taper parameter=10) was applied to the cylinder ends (Eq. (2) in Lawson *et al.*, 2006, following Chu *et al.*, 1993). As the scattering is not particularly sensitive to the length-to-radius-of-curvature ratio (ρ_L) (Stanton *et al.*, 1993), $\rho_L=3$ is used throughout this study. Averages were performed over a normal distribution of angles of orientation, with the mean and standard deviation determined from the literature when possible. As the *in situ* angular distribution is not known for most zooplankton taxa, unless otherwise stated, a mean of 0° (corresponding to broadside incidence or horizontally oriented) and a standard deviation of 30° was used. In some cases, the model parameters have been validated through comparison with laboratory or *in situ* measurements. All lengths (L) in the table are in millimeters. The length-to-width ratio, β_D , should not be confused with the length-to-radius ratio, β , used by Stanton *et al.* (1998b).

Taxon (Scattering model)	Length-to-width ratio (β_D)	Orientation (Mean, STD)	Density contrast (g)	Sound speed contrast (h)
Euphausiids and Decapod Shrimp (DWBA uniformly-bent cylinder)	10.5 ^a	N(20,20) ^{b, R1}	$g=5.485L/10^4+1.002, L>25$ $g=1.016, L<25$ ^{R2}	$h=5.942L/10^4+1.004, L>25$ $h=1.019, L<25$ ^{R2}
Larval Crustaceans ^c (DWBA uniformly-bent cylinder)	2.55 ^d	N(0,30)	1.058 ^{R3}	1.058 ^{R3}
Amphipods ^{c, R4} (DWBA uniformly-bent cylinder)	3.00 ^d	N(0,30)	1.058 ^{R3}	1.058 ^{R3}
Ostracods ^c (DWBA uniformly-bent cylinder)	2.55 ^d	N(0,30)	1.03 ^{R5}	1.03 ^{R5}
Chaetognaths and Polychaetes ^c (DWBA uniformly-bent cylinder)	17.15 ^d	N(0,30)	1.03 ^{R5}	1.03 ^{R5}
Gymnosome Pteropods (<i>Clione</i>) ^c (DWBA uniformly-bent cylinder)	1.83 ^d	N(0,30)	1.03 ^{R5}	1.03 ^{R5}
Salps ^{c, R6, R7} (DWBA uniformly-bent cylinder)	4.0 ^d	N(0,30)	1.004 ^{R6}	1.004 ^{R6}
Copepods (DWBA prolate spheroid ^{R10})	2.55 ^d	N(90,30) ^{e, R8}	1.02 ^{R5}	1.058 ^{R9}
Medusae ^{R11} (DWBA two prolate spheroidal surfaces ^{R5})	NA	NA	1.02 ^f	1.02 ^f
Eggs (High-pass fluid sphere ^{g, R13})	NA	NA	0.979 ^{R12}	1.017 ^{R12}

R1: Chu *et al.*, 1993

R2: Lawson *et al.*, 2004

R3: Table I in Lawson *et al.* (2004)

R4: Trevorrow and Tanaka, 1997

R5: Personal communication Dezhong Chu

R6: Stanton *et al.*, 1994

R7: David *et al.*, 2001

R8: Benfield *et al.*, 2000

R9: Chu *et al.*, 2000

R10: Fig. 12 in Stanton and Chu, 2000

R11: Mutlu, 1996; Monger *et al.*, 1998; Brierley *et al.*, 2004

R12: Chu *et al.*, 2003

R13: Stanton, 1989

^aMeasurements performed by Joe Warren (personal communication) for a subset of euphausiids in Wilkinson Basin in the GoM in October 1999. This value is larger by almost a factor of 2 than values used previously in the literature (Warren *et al.*, 2003; Fielding *et al.*, 2004) leading to significant decreases in the predicted scattering.

^bA recent study has indicated that the distribution of euphausiid orientations in the Western Antarctic Peninsula has a mean that is close to 0° (Lawson *et al.*, 2006). This change results in small changes in the contribution to scattering from euphausiids and does not affect the conclusions arrived at here, particularly as euphausiids did not greatly contribute to the predicted scattering in the GoM. In addition, the changes of the measured pitch and roll of BIOMAPER-II were sufficiently small that they do not significantly affect the scattering predictions for euphausiids, or any other zooplankton.

^cThere have been limited acoustic studies specific to these zooplankton and the DWBA uniformly-bent cylinder model was used as they have no known gas inclusions and their material properties appear similar to those of better studied fluid-like zooplankton.

^dThese length-to-width ratios were based on measurements of a sub-sample of zooplankton from the Western Antarctic Peninsula (Table I of Lawson *et al.*, 2004). It is expected that these parameters will depend on season and geographic location.

^eOther studies have used a normal distribution of orientations for copepods with a mean of 0° (broadside incidence or horizontally oriented) and a standard deviation of 30° (Lawson *et al.*, 2004). This distribution was also investigated, but due to the averaging the differences in the predicted scattering between the two distributions of orientations were small and did not affect the conclusions of this study.

^fInferred from comparison of the Monger *et al.* (1998) data to the model predictions, personal communication Dezhong Chu.

^g $\sigma_{bs} = a^2(ka)^4 \alpha_{\pi s}^2 / (1 + 4(ka)^4 \alpha_{\pi s}^2 / R^2)$ where $\alpha_{\pi s} = (1 - gh^2) / 3gh^2 + (1 - g) / (1 + 2g)$ and $R = (gh - 1) / (gh + 1)$.

idealized spherical fluid-filled elastic shell, which has an exact modal series solution (Goodman and Stern, 1962) and is valid for all ka , where the average is over a range of radii and shell thickness (Fig. 11 in Stanton *et al.*, 2000).

Three different models for elastic-shelled pteropods were used in this study (Fig. 2). *MODEL 1*: A high-pass dense fluid-sphere model (in Table I) with $g=h=1.73$ and $R=0.5$. This model best fits the Stanton *et al.* (1998a; Fig. 4)

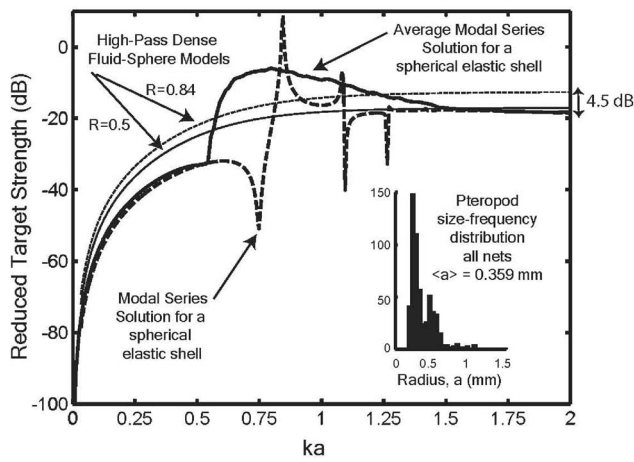


FIG. 2. Reduced target strength (RTS) for elastic-shelled pteropods as a function of ka based on the high-pass dense fluid-sphere model with $R=0.5$ (thin solid line) and $R=0.84$ (thin dashed line), the exact modal series solution for an individual spherical aragonite shell (thick dashed line) with a fractional shell thickness (shell thickness/mean radius) of 2.3%, corresponding to shell thicknesses of 4–9 μm for the range of measured pteropod radii in the GoM, and an averaged modal series solution for a spherical aragonite shell, with the average taken over a Gaussian distribution of shell radii and thickness (thick solid line). The value for the fractional shell thickness is consistent with those found in the literature (Lalli and Gilmer, 1989; Chu and Stanton, 1998), as well as with restricted measurements performed by the authors for crushed shell parts. For large values of ka , the difference in the reduced target strength predicted by the high-pass dense fluid-sphere model with reflection coefficients R_1 and R_2 is $10 \log_{10}(R_1^2/R_2^2)$, or 4.5 dB for $R_1=0.5$ and $R_2=0.84$. For these predictions, the density contrast between the surrounding water and the shell was 2.84 and 1.022 between the surrounding water and the fluid interior. The compressional sound speed contrasts between the surrounding water and the shell and fluid interior were 3.98 and 1.04, respectively. The shear sound speed contrast was 2.34. The averages were performed over a Gaussian distribution of shell radii with a 20% standard deviation (s.d.), and a Gaussian distribution of shell thickness centered around a fractional shell thickness of 2.3% and with a 10% s.d., spanning ± 2 s.d. from the mean (the approach taken by Stanton *et al.*, 2000). These parameters were chosen to give a reasonable fit to published values of the average RTS, -18.2 dB over a range of ka values from 1.16 to 1.88, deduced from laboratory measurements of scattering from pteropods (Fig. 4 in Stanton *et al.*, 1998a). The inset shows the size-frequency distributions for all pteropods observed in the GoM.

pteropod data when averaged over all angles of orientation. **MODEL 2:** A high-pass dense fluid-sphere model with $g=2.84$ and $h=3.98$ for aragonite, for which $R=0.84$. **MODEL 3:** An averaged exact modal series solution for a fluid-filled elastic aragonite shell (following Stanton *et al.*, 2000).

Finally, there are other zooplankton with hard elastic shells, including foraminifera and radiolarians. These zooplankton are more complex and irregular in shape than pteropods, are typically an order of magnitude smaller, and can have shells made of aragonite, calcite, or silica. A high-pass dense fluid-sphere model was used for these zooplankton with $g=2.147$ and $h=3.979$ based on values for fused silica.

E. Gas-bearing zooplankton

Siphonophores are fragile gelatinous zooplankton that often possess small gas inclusions called pneumatophores. The siphonophore most commonly observed in the GoM, *Nanomia cara*, typically has a single pneumatophore filled with carbon monoxide gas ($g=0.0012$ and $h=0.22$) (Benfield *et al.*, 2003, and references therein). The nongaseous weakly

scattering tissue of the siphonophores is composed of numerous gelatinous parts, including nectophores and bracts. Scattering by the pneumatophores is significantly stronger than the scattering from the tissue over a broad frequency range (Fig. 5 in Stanton *et al.*, 1998a, Warren *et al.*, 2001), particularly at the resonance frequency of the gas inclusion.

In this study, a hybrid scattering model for a fluid-filled sphere has been used for the siphonophore pneumatophores in which a simple model that includes damping was used for $ka < 0.1$, and the exact modal series solution for a fluid sphere (Anderson, 1950) was used for $ka > 0.1$. For $ka < 0.1$ the backscattering cross section is given by (Weston, 1967)

$$\sigma_{\text{bs}} = \frac{a^2}{(1 - k_0^2/k^2)^2 + 1/Q^2}, \quad (3)$$

where a is the bubble radius and k_0 is the acoustic wave number at the resonance frequency, given by

$$k_0 = \frac{(3\gamma)^{1/2}}{ac} \left(\frac{P_0(1 + 0.1z)}{\rho} \right)^{1/2}, \quad (4)$$

where z is depth, $\rho=1027$ kg/m^3 , $P_0=1.013 \times 10^5$ Pa is the pressure at the surface, and γ ($\gamma=1.4$) is the ratio of gaseous specific heats. Q is the quality factor ($Q=\delta^{-1}$, where δ is the damping constant). A typical value for swim-bladdered fish is $Q=5$ (Diachok, 2001), and for lack of data, it was also assumed here that $Q=5$. The depth dependence has been included by assuming a sphere of constant volume (Benfield *et al.*, 2003). This damped model is only valid at small ka as an inherent assumption is that the scattering is spherically symmetric. At high ka values, the total scattering cross section, σ , for an ideal gas-filled sphere should be independent of frequency (ignoring narrow resonances) and converge to the geometrical scattering cross section (πa^2). Yet for the damped model, σ converges to $4\pi a^2$, a factor of 4 (~ 6 dB) too large. The difference between the damped model and the exact modal series solution is larger than 5% for $ka > 0.1$. As the range of ka values spanned in this study is 0.0034–3.99, the hybrid approach was necessary.

The exact modal series solution for a fluid sphere ($g=h=1.02$) was also used to describe the scattering from siphonophore body parts. Lengths were converted to radii of the spheres of equivalent volume using an empirically derived scaling factor ($L=\gamma a_{\text{csr}}$, where $\gamma=0.4$ for nectophores and $\gamma=0.29$ for bracts). These scaling factors were derived for siphonophore body parts collected in the Western Antarctic Peninsula (Lawson *et al.*, 2004), and could differ according to siphonophore genera and location. In previous work (Stanton *et al.*, 1998b) siphonophore body parts have also been modeled as cylinders. The contributions to scattering from the siphonophore pneumatophores and body parts were added incoherently. Though siphonophore parts are weakly scattering, their dimensions are relatively large compared to more abundant copepods or pteropods (mean measured length for bracts=8.2 mm, with individuals as large as 19.5 mm). As a result, the Rayleigh-to-geometric scattering transition occurs at lower frequencies for siphonophore body parts than for copepods or pteropods. Thus, over the range of

frequencies for which the large siphonophore body parts have reached the geometric scattering regime but more abundant, smaller zooplankton have not, the contribution to scattering from siphonophore body parts can be larger than the contribution from copepods or pteropods, even when the siphonophores occur in relatively low abundances.

F. Other sources of scattering

In this study, scattering from organisms that were rarely observed in the net samples, including fish and fish larvae (ten individuals observed in all 56 nets, only one myctophid), starfish (nine individuals), crabs (two individuals), and bivalves (one individual), has not been included. Scattering from diatoms was also omitted, as they were small (average length 0.3 mm) and only observed at some locations with low abundances. Scattering from bubbles can be also important close to the surface, potentially dominating the scattering over a wide range of frequencies. As there was no method of quantifying the bubble size distribution close to the surface, analysis of the surface scattering layer was not included in this study. While suspended sediments are also known to contribute to water-column scattering over the shallow waters of Georges Bank (Wiebe *et al.*, 1997; Pershing *et al.*, 2001), no suspended sediment was observed in the deep basins of the GoM, and thus it is not included in forthcoming analyses.

III. METHODS

The results presented here involve the collection of near-coincident multi-frequency acoustic data, net tows, CTD profiles, and video images of zooplankton. Central to the program is the towed instrument platform BIOMAPER-II that collects along-track video images of plankton and multi-frequency acoustic backscattering (Fig. 1). Additional ground truthing is provided by CTD profiles and depth-resolved oblique net tows. The methods used to collect and analyze data collected with these instruments are outlined in this section, together with a description of how this information is combined with the scattering models to make volume scattering predictions.

A. BIOMAPER-II towed instrument platform

BIOMAPER-II is specifically designed to conduct synoptic, high-resolution, multi-frequency zooplankton acoustic surveys (Wiebe *et al.*, 2002). The key components of this instrument are the five upward- and five downward-looking transducers (43, 120, 200, 420, and 1000 kHz), manufactured by Hydroacoustic Technologies Inc. (HTI, Seattle, WA), a single camera Video Plankton Recorder (VPR) (Davis *et al.*, 1992), and sensors for measuring temperature, conductivity, pressure, heading, pitch, and roll. Full water-column acoustic coverage in shelf waters can often be achieved at the four lower frequencies with BIOMAPER-II at depth since each pair of acoustic transducers has one downward- and one upward-facing transducer. BIOMAPER-II is normally towed in an undulating fashion, up and down through the water column, from a few meters below the sea surface to within 10 or 20 m from the seafloor,

in order to collect VPR images and environmental data at all depths. A Global Positioning System (GPS) receiver synchronized the different data. BIOMAPER-II is typically towed at speeds of 4–7 knots, except when biological net samples are collected and the tow speed is below 2.5 knots.

Volume backscattering is collected from sequential transmissions from all ten transducers in 1 m depth intervals to a range of 200 m at 43 and 120 kHz, 150 m at 200 kHz, 100 m at 420 kHz, and 35 m at 1 MHz. The 1 MHz backscattering data are not used in this study due to their reduced range. A 10 kHz bandwidth, linear frequency modulated (chirp) signal was used at a repetition rate of 2.5 pings/s. Echo integration was performed every 12 s to obtain volume backscattering, corresponding to a horizontal resolution of approximately 30–40 m, depending on vessel speed. Combined noise levels resulting from the ship, ambient, and system noise, were collected as a function of depth with BIOMAPER-II in “passive listening” mode. Noise thresholds, which vary with range and frequency (Korneliussen, 2000), were set by adding 6 dB to the measured noise profiles. Backscattering levels were then compared to these noise thresholds on a ping-by-ping basis, prior to echo integration, and bins in which the backscattering did not exceed the noise were set to zero. The transducers were split-beam and had full beamwidths (–3 dB to –3 dB) of 7° for the 43 kHz and 3° for the 120–420 kHz. The corresponding sampling volumes of a 1-m-thick bin centered at 100 m ranged from 21 to 86 m³.

The acoustic system on BIOMAPER-II was calibrated by HTI in September 1999, immediately prior to this study, using both standard transducers and standard targets, for source level, receive sensitivity, and beam patterns. These calibrations were supplemented with standard-target calibrations, using 20 and 38.1 mm tungsten carbide (6% cobalt) spheres and excluding off-axis returns, performed at Woods Hole Oceanographic Institution (WHOI) based on the practices established by Foote *et al.* (1987). Depth-dependent calibrations were not performed, but the transducer backings were made of a noncompressible synthetic material to minimize the effect of depth-dependent changes in the transducer performance. In addition, changes in the scattering from the sea floor as a function of BIOMAPER-II depth, down to approximately 200 m, were small, generally smaller than 0.5 dB, which will not affect the conclusions arrived at in this study. Studies performed by Kloser (Fig. 2 in Kloser, 1996) and Gauthier and Rose (Fig. 4 in Gauthier and Rose, 2002) support the conclusion that depth-dependent changes in transducer calibrations with this type of backing will not be significant over the range of depths investigated here. Furthermore, for the majority of the data analyzed in this paper, which involve data collected when BIOMAPER-II was towed at the surface during the net tows, there are no depth-dependent effects that need to be considered.

A region of enhanced backscattering was observed in the vicinity of BIOMAPER-II, particularly at 43 and 120 kHz, while BIOMAPER-II was towed up and down through the water column at depths with strong gradients in the temperature and salinity. Regions in which enhanced backscattering were observed are not used in a quantitative way in this

study: the enhancement was not observed during the net tows when BIOMAPER-II was towed at the surface.

B. Physical and biological sampling

1. Physical sampling

The ship's CTD (Sea Bird 911 plus), sampling at 24 Hz, with typical descent rates of 1 m/s, was used to determine many of the parameters needed to make predictions of scattering from microstructure. Though BIOMAPER-II collected physical environmental data, as it was towed up and down through the water column, calculation of vertical gradients would need to account for horizontal gradients. In addition, the sampling rate was low (0.25 Hz), corresponding to at best 6–8 m horizontal resolution and 40 cm vertical resolution.

2. Biological sampling: Nets

Depth-resolved zooplankton samples were collected at select locations using a 1-m² Multiple Opening/Closing Net and Environmental Sensing System, MOCNESS (Wiebe *et al.*, 1985). The MOCNESS system was towed obliquely while the BIOMAPER-II instrument was at the surface collecting acoustic data. The MOCNESS was equipped with nine 335 μm mesh nets and environmental sensors including a flow meter, temperature, conductivity, and depth. The first net (net 0) sampled the entire water column down to approximately 10 m above the bottom, and the remaining eight nets sampled quantitatively, with sampling strata dependent on the water depth: typically, the upper 100 m was sampled in 25 m intervals, with a few 50 m intervals at intermediate depths, and 25 m intervals at the deeper depth ranges. The samples were preserved upon recovery in 5% buffered formalin. The "silhouette" method developed by Ortner *et al.* (1979) and modified by Davis and Wiebe (1985) was used to measure the size (typically length) of each individual organism in each net sample. Between 100 and 300 m³ of water were filtered by each net. Occasionally, the number of organisms in a particular net was excessively large, making it unfeasible to count every individual. These large samples were sequentially split into two equal portions, sometimes more than once.

3. Biological sampling: Video images

The VPR is a high-magnification underwater video system that records images of plankton using an analog video camera and strobe light (Davis *et al.*, 1992), and is mounted to the front end of the BIOMAPER-II instrument. The volume imaged by this system is 5.1 ml at 60 Hz (3×10^{-4} m³/s), many times smaller than the acoustic or MOCNESS sampling volumes. The postprocessing of the images involved digitization and target detection using user-defined criteria for size, focus, and brightness (Benfield *et al.*, 2003). Targets that met these criteria were sorted into different taxonomic categories, enumerated and measured together with the location, time, and depth at which they were observed. For copepods, the size measured in the video images was the smallest dimension, which corresponds approximately to the width of the organisms. The length of the

copepods was then determined using the length-to-width ratio $\beta_D=2.55$. This approach was taken to minimize the effect of ambiguities in the orientation of individual copepods in the field of view of the camera.

C. Predicting volume backscattering: The forward calculation

1. Predicting volume backscattering from turbulent microstructure

The predicted volume backscattering for turbulent microstructure was based on Eq. (2) (Sec. II B), with all the necessary model parameters determined from the CTD data except the dissipation rate of turbulent kinetic energy, ε . Many of the model parameters, such as A and B , are relatively constant as a function of depth and their depth-averaged values were used. Temperature and salinity gradients and the buoyancy frequency were averaged over 1 m depth bins. The CTD profiles were carefully processed to remove spikes in the temperature and salinity gradients resulting from the heaving motion of the vessel. Measurements of ε typically require specialized microstructure instruments, which were not available in this study. Under some circumstances it is possible to determine ε by identifying the outer scales of overturning eddies from CTD profiles (Thorpe, 1977). However, in regions of weak density gradients, low dissipation rates, and high sea states, the inferred values of ε can be compromised. Instead of pursuing this approach, an upper-bound estimate for ε ($\varepsilon=10^{-6}$ m²/s³) based on available information in the published literature (Burgett *et al.*, 2001) was used to estimate an *upper-bound* contribution to scattering from microstructure at each location and depth. The actual contribution to scattering from microstructure is expected to be lower than the upper-bound prediction.

2. Predicting volume backscattering from zooplankton: Nets

The predicted volume backscattering based on the MOCNESS samples was calculated using Eq. (1) at the four BIOMAPER-II frequencies ($f_q=43, 120, 200,$ and 420 kHz), and at each of the eight MOCNESS depth ranges, d_k , where k varies from 1 to 8. The backscattering cross section was calculated for each individual organism sampled by the MOCNESS based on its measured length. Though no differences are expected in the scattering from upward- versus downward-facing transducers for any weakly scattering zooplankton (Lavery *et al.*, 2001), this may not be the case for elastic-shelled zooplankton. However, as BIOMAPER-II was towed at the surface during the MOCNESS tows, only acoustic data from the downward-looking transducers are used for comparisons of predicted and observed scattering. For the comparison of measured and predicted volume backscattering, the measured volume backscattering was averaged over the same time intervals and range of depths as were sampled by each of the MOCNESS nets in a given profile. The offset between the MOCNESS and the BIOMAPER-II was not accounted for as it was no larger than 200 m and, at select locations where the offset was accounted for, it did not significantly affect the results. The

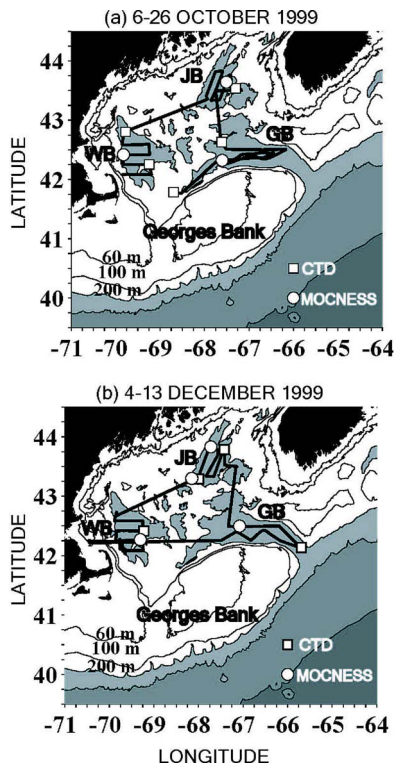


FIG. 3. (Color online) (a) October and (b) December, 1999, actual survey tracks superimposed on the bathymetry of the Gulf of Maine (GoM). The location of the three major basins in the GoM, Wilkinson Basin (WB), Jordan Basin (JB), and Georges Basin (GB) are indicated. The squares show the locations of the CTD profiles and the circles show the locations of the MOCNESS tows. Gale conditions in Georges Basin in October, 1999, resulted in significant deviations from the planned survey grid lines, in addition to BIOMAPER-II having to be towed at a depth of 100 m for large distances.

scattering predictions for the shallowest MOCNESS nets (nets 8), spanning the top 25 m of the water column, were not compared to the measured scattering in this depth range, though predictions were made, since elevated scattering was often observed, probably due to bubbles generated by breaking waves or the ship's wake, and it was not possible to quantitatively evaluate their contribution to the observed scattering.

3. Predicting volume backscattering from zooplankton: Video images

Predictions of volume backscattering based on the VPR images proceeded in much the same way as for the MOCNESS. However, as the acoustic data were averaged over 1 m depth bins, the predicted volume backscattering based on Eq. (1) and the VPR images were also binned into 1 m depth bins. The VPR images from the upward and downward parts of the tow were combined, and the acoustic data were averaged horizontally across the duration of the tow.

IV. GENERAL OBSERVATIONS IN THE GULF OF MAINE

Selected results are presented from two acoustic surveys of the deep basins of the Gulf of Maine (GoM) performed with the BIOMAPER-II towed instrument platform during

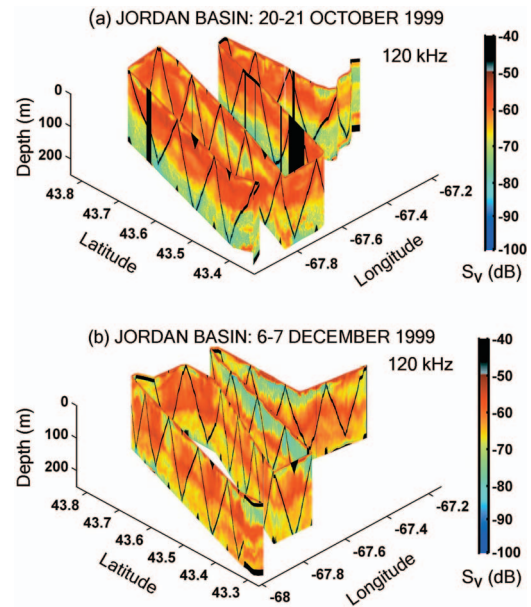


FIG. 4. Volume backscattering at 120 kHz as a function of depth, latitude and longitude in Jordan Basin, GoM, during (a) October and (b) December 1999. The black line is the BIOMAPER-II trajectory.

16–26 October (cruise EN330) and 4–13 December (cruise EN331), 1999, on board the RV Endeavor (Fig. 3). Hundreds of kilometers of along-track acoustics data and associated direct physical and biological data were collected in the three deep basins in the GoM: Wilkinson (WB), Jordan (JB), and Georges (GB), each with typical depths between 200 and 300 m.

A. Acoustical observations

The volume scattering strength during the 1999 surveys of the deep basins of the GoM was characterized by a complex horizontal and vertical structure, with spatial structures observed over a range of different scales, from a few meters to tens of kilometers (Fig. 4). At many locations the scattering increased monotonically with increasing frequency. At some locations and depths, such as the pervasive scattering layer that was often observed at the seasonal thermocline (Brooks, 1996) (Fig. 5), the scattering varied nonmonotonically with frequency, largest at 120 kHz, smallest at 43 kHz, and broadly similar at 200 and 420 kHz. High scattering levels were observed close to the sea surface, probably due to bubbles. A deeper scattering layer was observed occasionally, typically during daylight hours, and the volume scattering associated with this deeper layer was approximately constant at all frequencies.

B. Physical observations

At least one CTD profile was performed in each basin during the October and December 1999 surveys (Table II). The seasonal thermocline was located at approximately 40–60 m in October and 70–100 m in December 1999. The temperature and salinity stratification in the early fall was significantly larger than in the late fall, particularly in Wilkinson Basin.

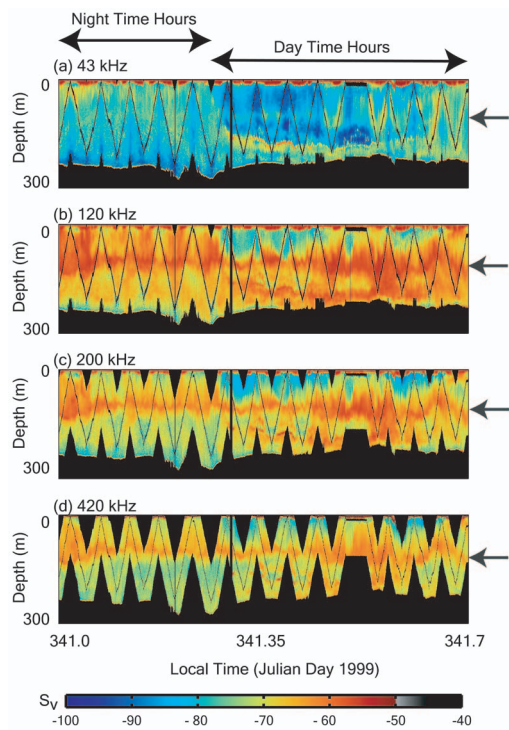


FIG. 5. Volume backscattering as a function of depth and local time on 7 December 1999 in Jordan Basin, GoM, showing the frequency dependence of the daily vertical migration of zooplankton. The arrows on the right indicate the approximate depth of the seasonal thermocline, which was located at approximately 40–60 m in October and 70–100 m in December. The diel migration of zooplankton was observed all frequencies. The largest differences between the day and night scattering were observed at 43 kHz, particularly for the deeper scattering layer.

C. Biological observations

1. Biological observations: Nets

One night tow per basin has been analyzed for taxon composition and size distribution for each survey, in addition to a day tow in Jordan Basin in December 1999 (Table III, Fig. 3). Copepods made up the majority of the numerical abundance (Fig. 6) and biomass (Fig. 7) of zooplankton at most locations. Euphausiids contributed most significantly to biomass in December 1999 in Wilkinson and Georges Basin. Other weakly scattering zooplankton only contributed significantly to biomass at select locations. Pteropods made up an insignificant portion of the observed biomass at any location, while the biomass due to siphonophores was only significant at a few locations.

It is difficult to accurately assess the distribution and abundance of siphonophores as they are often overlooked or destroyed by net systems. However, siphonophore parts, including pneumatophores, nectophores, and bracts, were observed in most of the MOCNESS samples. There were only four nets (nets 1, 3, 6, and 8) in Wilkinson Basin in October and one net (net 1) in Jordan Basin in December in which no pneumatophores were observed. The vertical distribution of pneumatophores peaked at depths that corresponded to the seasonal thermocline, with abundances of 3–5 pneumatophores/m³ (Fig. 8).

A number of previous studies have assumed that all observed siphonophores have pneumatophores of a single size (Warren *et al.*, 2001, Trevorrow *et al.*, 2005). In this study, the measured pneumatophore size distribution is used to compare observed versus predicted scattering. There is evidence based on the adult pneumatophores collected with the MOCNESS tows that the pneumatophores support a slight eccentricity, $\beta_D < 3$, though it is unclear exactly how much

TABLE II. Dates (mm-dd-year), times, and locations of the CTD profiles performed during the October and December 1999 surveys of the deep basins of the GoM. WB=Wilkinson Basin, JB=Jordan Basin, GB=Georges Basin, and YD=Julian Year Day.

	Local time	Local date	Latitude (N)	Longitude
October 1999 CTD 1 (WB)	12:53 (Daytime CTD)	10-17-1999 (YD 290)	42.254	-69.2412
October 1999 CTD 2 (WB)	12:28 (Daytime CTD)	10-20-1999 (YD 294)	42.8028	-69.7773
October 1999 CTD 3 (JB)	17:17 (Evening CTD)	10-21-1999 (YD 295)	43.5367	-67.2208
October 1999 CTD 4 (GB)	12:54 (Daytime CTD)	10-22-1999 (YD 296)	42.6297	-67.6182
October 1999 CTD 5 (GB)	11:15 (Daytime CTD)	10-25-1999 (YD 298)	41.767	-68.7067
December 1999 CTD 1 (WB)	14:23 (Daytime CTD)	12-05-1999 (YD 339)	42.4128	-69.2270
December 1999 CTD 2 (JB)	14:35 (Daytime CTD)	12-06-1999 (YD 340)	43.2700	-68.0007
December 1999 CTD 3 (JB)	12:34 (Daytime CTD)	12-07-1999 (YD 341)	43.7847	-67.4208
December 1999 CTD 4 (GB)	13:32 (Daytime CTD)	12-09-1999 (YD 343)	42.1367	-65.6933

TABLE III. Dates (mm-dd-year), times, and location of the MOCNESS tows performed during the October and December 1999 surveys of the deep basins of the GoM. WB=Wilkinson Basin, JB=Jordan Basin, GB=Georges Basin, and YD=Julian Year Day.

MOCNESS	Local time	Local date	Latitude (N)	Longitude	Closest CTD
MOC 2, WB, October 1999	01:34 (Night Tow)	10-18-1999 (YD 291)	42.4172	-69.8229	CTD 1, WB, October 1999
MOC 5, JB, October 1999	01:10 (Night Tow)	10-21-1999 (YD 295)	43.6426	-67.5096	CTD 3, JB, October 1999
MOC 6, GB, October 1999	21:44 (Night Tow)	10-24-1999 (YD 298)	42.3215	-67.6027	CTD 4, GB, October 1999
MOC 2, WB, December 1999	11:58 (Night Tow)	12-04-1999 (YD 338)	42.2687	-69.3062	CTD 1, WB, December 1999
MCO 4, JB, December 1999	13:10 (Day Tow)	12-06-1999 (YD 340)	43.3174	-68.0000	CTD 2, JB, December 1999
MCO 5, JB, December 1999	22:59 (Night Tow)	12-06-1999 (YD 340)	43.8303	-67.7225	CTD 3, JB, December 1999
MCO 6, GB, December 1999	22:55 (Night Tow)	12-08-1999 (YD 343)	42.4960	-67.0767	CTD 4, GB, December 1999

NUMERICAL ABUNDANCE

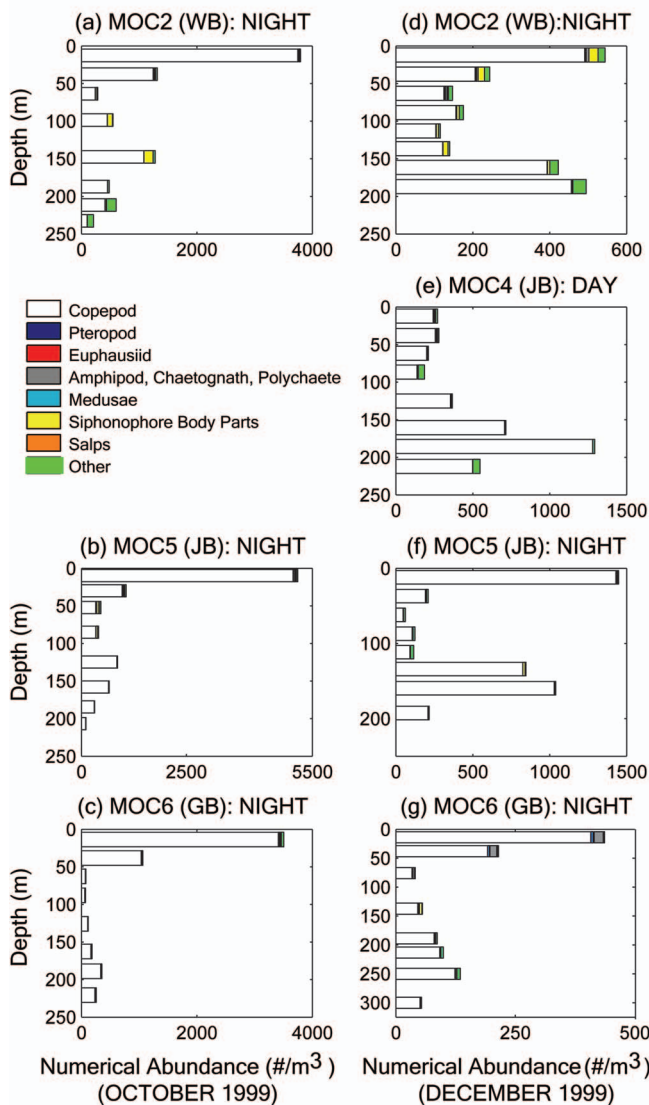


FIG. 6. Numerical abundance as a function of depth determined from the quantitative analysis of the MOCNESS net tows.

of the actual pneumatophore is filled with gas, or the extent to which the process of raising them to the surface from the depth at which they were captured and preserving them for later analysis has altered the shape. Based on length (L) and width (W) measurements of a sub-sample of the adult pneumatophores, it was determined that $W(\text{mm})=0.256L(\text{mm})+0.085$. In this study, the pneumatophore radius is given by the equivalent spherical radius of a sphere of the same volume as a prolate ellipsoid with major and minor axes given by the measured length and width, $a=(LW^2)^{1/3}/2$.

2. Biological observations: Video images

Due to the vast number of VPR images collected (~5 million per day) only the images for one full upward and downward tow in Jordan Basin on 6 December 1999 have been fully analyzed for all taxonomic categories and sizes. This location was chosen as it immediately followed profiles by the MOCNESS and CTD systems. At this loca-

BIOMASS

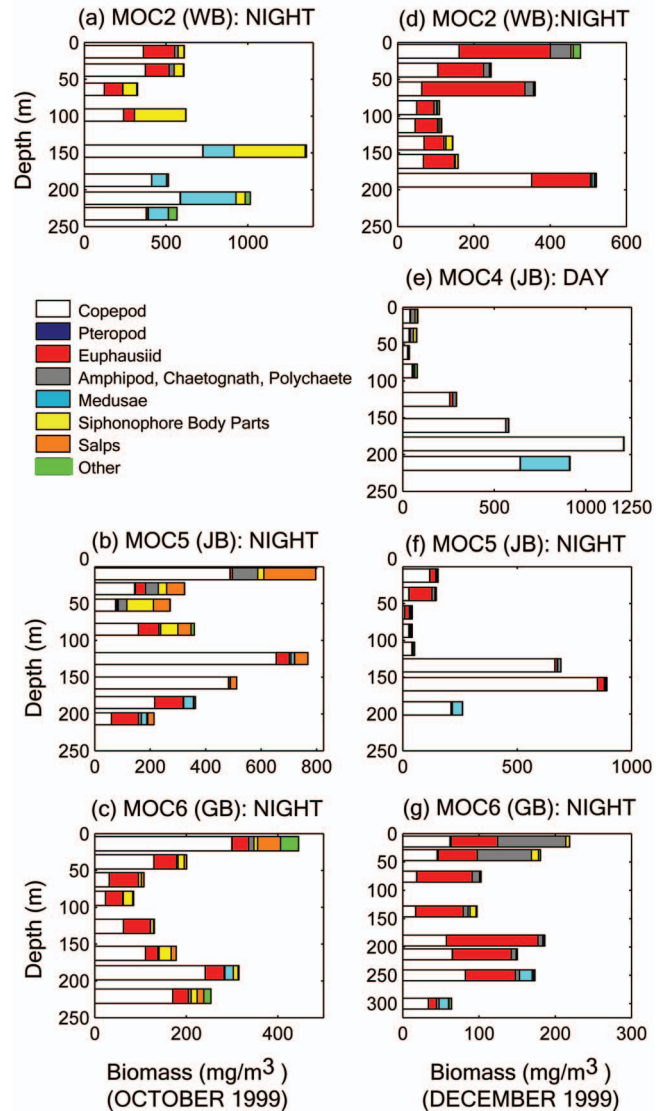


FIG. 7. Biomass as a function of depth determined from the quantitative analysis of the MOCNESS net tows.

tion, the biomass generally increased with depth and copepods made up the majority of the biomass at all depths. As a result of the small sampling volume of the VPR, the biomass and abundance only agree with the MOCNESS for small, abundant zooplankton, namely copepods. For all other zooplankton, the VPR estimates of abundance were significantly lower than the corresponding MOCNESS estimates. Previous studies have shown better agreement between VPR and MOCNESS estimates of pteropod abundance (Benfield *et al.*, 1996), but that success was not reproduced here, possibly due to lower pteropod abundance in this study or the inherent patchiness in the distribution.

V. DETERMINING DOMINANT SCATTERERS

In this section, predicted dominant biological scatterers (scatterers that make up more than 50% of the total predicted scattering) are identified based on the composition and size of zooplankton in all the available MOCNESS net tows

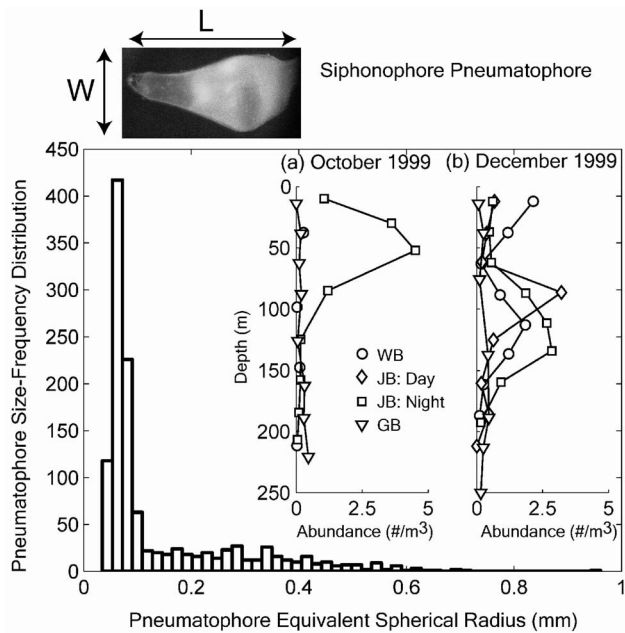


FIG. 8. Pneumatophore size-frequency distribution for all the pneumatophores observed in the deep basins of the GoM during October and December, 1999. A total of 1417 pneumatophores were measured, 121 in October and 1076 in December. The distribution is strongly peaked at a radius of 0.075 mm, corresponding to a resonance frequency of 45 kHz at the surface and 143 kHz at 100 m, approximately spanning the three lowest BIOMAPER-II frequencies. Inset: Numerical abundance of pneumatophores as a function of depth (a) in October 1999 and (b) in December 1999. The pneumatophore abundance is generally largest at the seasonal thermocline, which was deeper in December than October. The size distribution of pneumatophores did not change significantly from October to December, though the pneumatophore abundance was lower in October than in December. The image shows a pneumatophore collected by the MOCNESS.

(Figs. 9(a)–9(d)) and compared to the measured multi-frequency backscattering data collected at the same time and location with BIOMAPER-II towed at the surface (Figs. 10–12). The predicted dominant biological scatterers are also identified based on the composition and size of zooplankton from video images at a select location. In addition, an upper-bound contribution to volume backscattering from turbulent microstructure is calculated based on the CTD data. Scattering predictions are performed over the frequency range from 10 kHz to 2 MHz, for comparison with the four BIOMAPER-II frequencies within that range. At locations where a single dominant scatterer could be identified, simple inversions for size and/or abundance are performed. These results are presented in this section.

A. Determining dominant scatterers: Turbulent microstructure

The predictions of scattering from turbulent microstructure based on the CTD data reveal that at most depths and locations the predicted scattering generally decreases with frequency by approximately 10–20 dB over the range of BIOMAPER-II frequencies. Though the magnitude of the change varied depending on the exact temperature and salinity gradients measured by the CTD (which vary with depth) and on the value of ϵ used to make the predictions, this

decrease was predicted across the frequency range of interest at most locations, depths, and values of ϵ (Fig. 13).

Internal waves are the most likely cause for elevated values of ϵ . In an earlier study, multi-frequency acoustic observations of an internal wave in Wilkinson Basin in October 1997 (Warren *et al.*, 2003) suggested that the observed acoustic scattering spectrum in areas of strong temperature gradients and elevated values of ϵ are not consistent with scattering from zooplankton, showing a generally decreasing trend of volume backscattering with increasing frequency. Scattering from non-gas-bearing zooplankton of the sizes typically observed in the GoM tends to show a generally increasing trend with increasing frequency over the range of frequencies relevant to this study. A similar analysis was performed at the locations of the two internal waves that were observed in the current study (one in Jordan Basin and one in Georges Basin in December 1999), and at other select locations, including at the locations of the MOCNESS tows (Fig. 11). Generally positive slopes were observed, which is consistent with scattering from zooplankton and not microstructure. It is concluded that scattering from microstructure was not a large contributor to volume scattering during the October and December 1999 surveys, and this contribution is not included in the forthcoming analyses.

B. Determining dominant biological scatterers: Nets

The total predicted scattering of biological origin at 43 kHz was strongly dominated by pneumatophores at most depths and locations (Fig. 9(a)). At 120 kHz, pneumatophores were still the dominant scatterers (Fig. 9(b)), but the relative contribution to scattering from siphonophore body parts had increased. At some locations where copepods were very abundant and made up most of the biomass, the contribution to the total predicted scattering from copepods became apparent at 120 kHz, though it was still small. At 200 kHz, the contribution to total scattering from pneumatophores was still significant, though only at certain locations and depths, and the contribution to total scattering from copepods dominated at some locations (Fig. 9(c)). At 420 kHz, the overall dominant predicted scatterers were copepods (Fig. 9(d)), though there were select locations where the contribution from pneumatophores was still significant, and relatively important contributions were evident from elastic-shelled pteropods in the near-surface nets. In general, the relative contribution to total predicted scattering from pneumatophores decreased with increasing frequency, while the contribution from fluid-like and elastic-shelled zooplankton increased. Though there was a significant contribution to biomass from euphausiids at some locations, their relative contribution to scattering was small at all frequencies, as the frequencies at which they would normally dominate were instead dominated by pneumatophores.

Comparison between the observed and predicted volume backscattering based on the MOCNESS tows illustrates that there is a general trend of under predicting the observed scattering (Fig. 12), except at 43 kHz. In fact, the predicted scattering was approximately 10 dB lower than the observed scattering if the pneumatophores were not included, and the

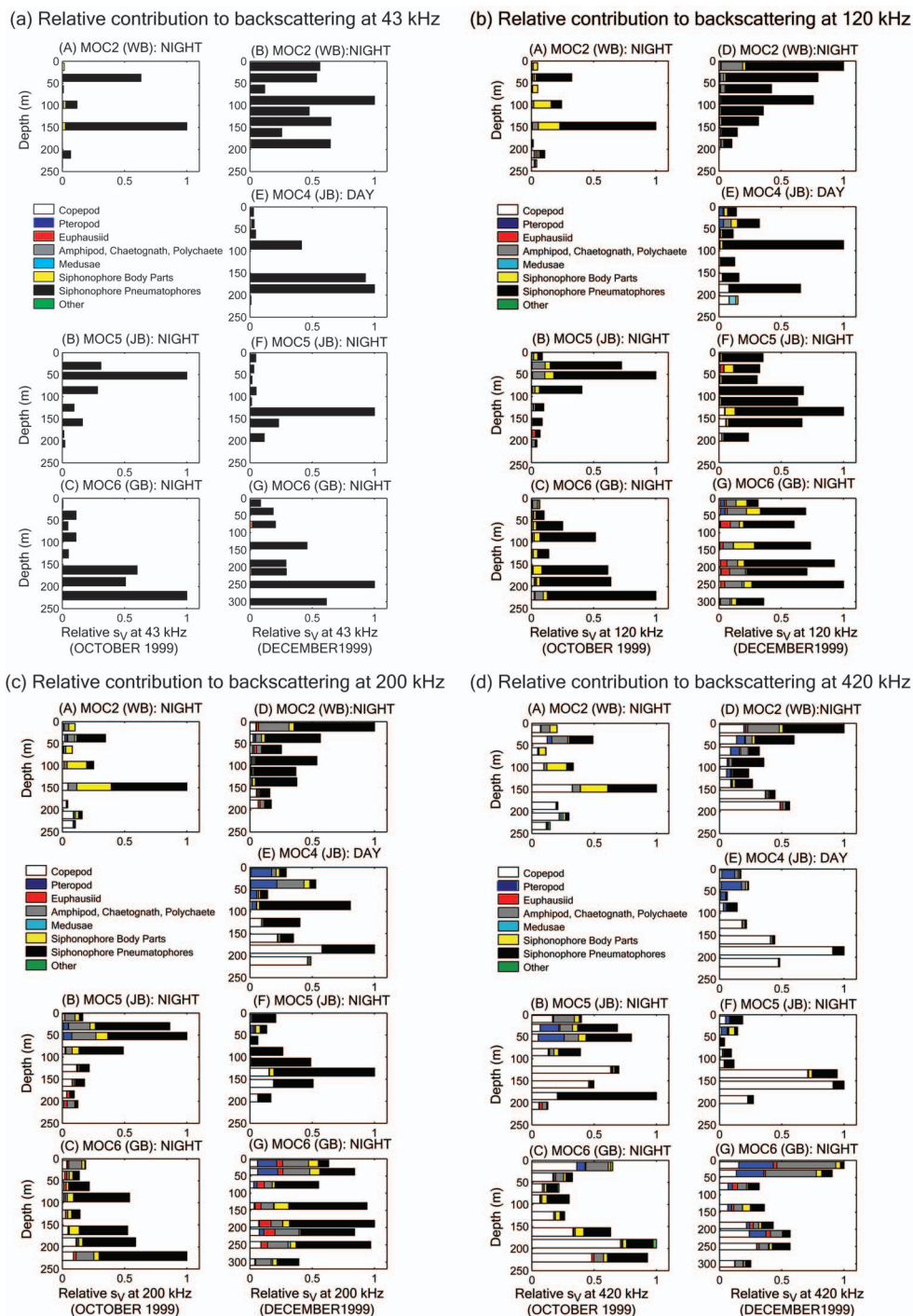


FIG. 9. Predicted dominant biological scatterers at a) 43 kHz, b) 120 kHz, c) 200 kHz, and d) 420 kHz for all MOCNESS tows. These predictions are based on pteropod model 1. Predictions based on pteropod models 2 and 3 showed qualitatively similar results.

difference was relatively independent of location, depth, or frequency. Predicted volume backscattering for both the day and night tows appeared to under predict the observed scattering by about the same amount (results not shown). Once the pneumatophores were included in the predictions the scattering levels were in closer agreement, though the general trend of under prediction remained.

1. Scattering dominated by fluid-like zooplankton

Fluid-like zooplankton, and copepods in particular, dominated the total predicted volume backscattering at

420 kHz at most depths except near the surface where elastic-shelled pteropods were also important. At 420 kHz, the general shape of the depth dependence of the total predicted volume backscattering (Fig. 9(d)) closely resembled the corresponding shape of the measured biomass (Fig. 7). This was particularly clear during the daytime net tow (MOCNESS 4) in Jordan Basin in December 1999 at depths below 150 m (nets 1–3), where copepods made up a very large fraction of the biomass and contributed over 90% of the total predicted scattering. A simple single-frequency calculation of copepod abundance has been performed at this loca-

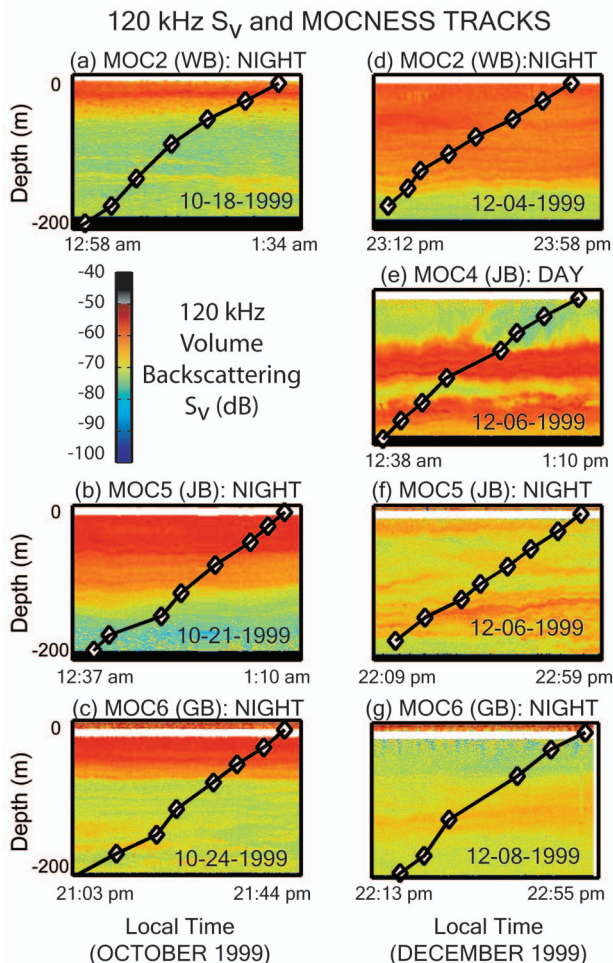


FIG. 10. Volume backscattering at 120 kHz as a function of depth at the locations of each of the seven MOCNESS tows. BIOMAPER-II was towed at the surface (<25 m depth) during all the MOCNESS tows. The black lines show the trajectory of the MOCNESS system, with the diamonds indicating the mid depths between the opening and closings of the nets.

tion, based on the measured mean volume backscattering and the mean copepod size measured in the nets. As BIOMAPER-II was towed at the surface during the deployment of the MOCNESS system, and the range on the 420 kHz transducers was 100 m, there were no acoustic data at 420 kHz collected simultaneously with the MOCNESS data below 100 m. As a result, the mean 420 kHz volume backscattering at the location of the first tow-yo directly after the MOCNESS was used. This calculation has been performed for nets 2 (175–200 m) and 3 (150–175), with mean 420 kHz S_V values of -68.7 and -68.5 dB, respectively. The mean copepod lengths in these depth ranges (2.37 mm in net 2 and 2.22 mm in net 3) give TS values of -94.8 and -95.8 dB, respectively. The copepod abundance is given by $s_V/\sigma_{bs}=10^{(S_V-TS)/10}$ resulting in inferred abundances of $410/m^3$ and $540/m^3$ for nets 2 and 3, correspondingly, compared to the measured abundances of $1290/m^3$ and $715/m^3$.

2. Scattering dominated by elastic-shelled zooplankton

The contribution from elastic-shelled pteropods to the total predicted scattering was most important at shallow

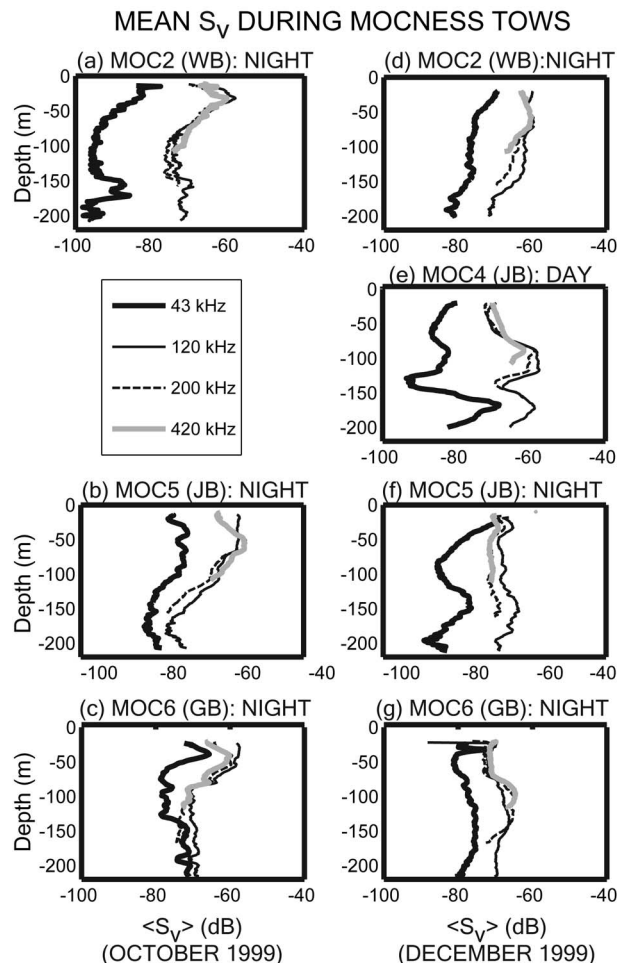


FIG. 11. Mean volume backscattering at all four BIOMAPER-II frequencies as a function of depth at the locations of each of the seven analyzed MOCNESS tows (Fig. 10). BIOMAPER-II was towed at the surface (<25 m depth) during the MOCNESS tows.

depths and at high frequencies, namely 200 and 420 kHz (Figs. 9(c) and 9(d)). The predicted contribution to scattering from pteropods has been made using the three models outlined in Sec. II D. Model 3 is exploratory in nature and was only applied to three MOCNESS tows (MOCNESS 4–6, December 1999) as the predictions based on this model are more computationally intensive. This exploratory model was attempted since, unlike the high-pass models, it does not monotonically increase with frequency and may reproduce the nonmonotonic scattering behavior observed at some locations more successfully.

The predicted scattering from pteropods never dominated (i.e., was not >50% of the total predicted scattering) at all frequencies (Table IV). However, it did dominate at some frequencies, reaching a maximum of 70% of the total predicted scattering (based on model 1) at 420 kHz in the 25–50 m depth bin (net 7) of the daytime MOCNESS tow (MOC 4) in Jordan Basin in December 1999 (Fig. 14). At this location and depth, the total predicted scattering based on all biological scatterers reproduced the measured scattering reasonably well at all frequencies. As expected, the contribution to scattering from pteropods increased with increasing frequency: The predicted scattering based on the

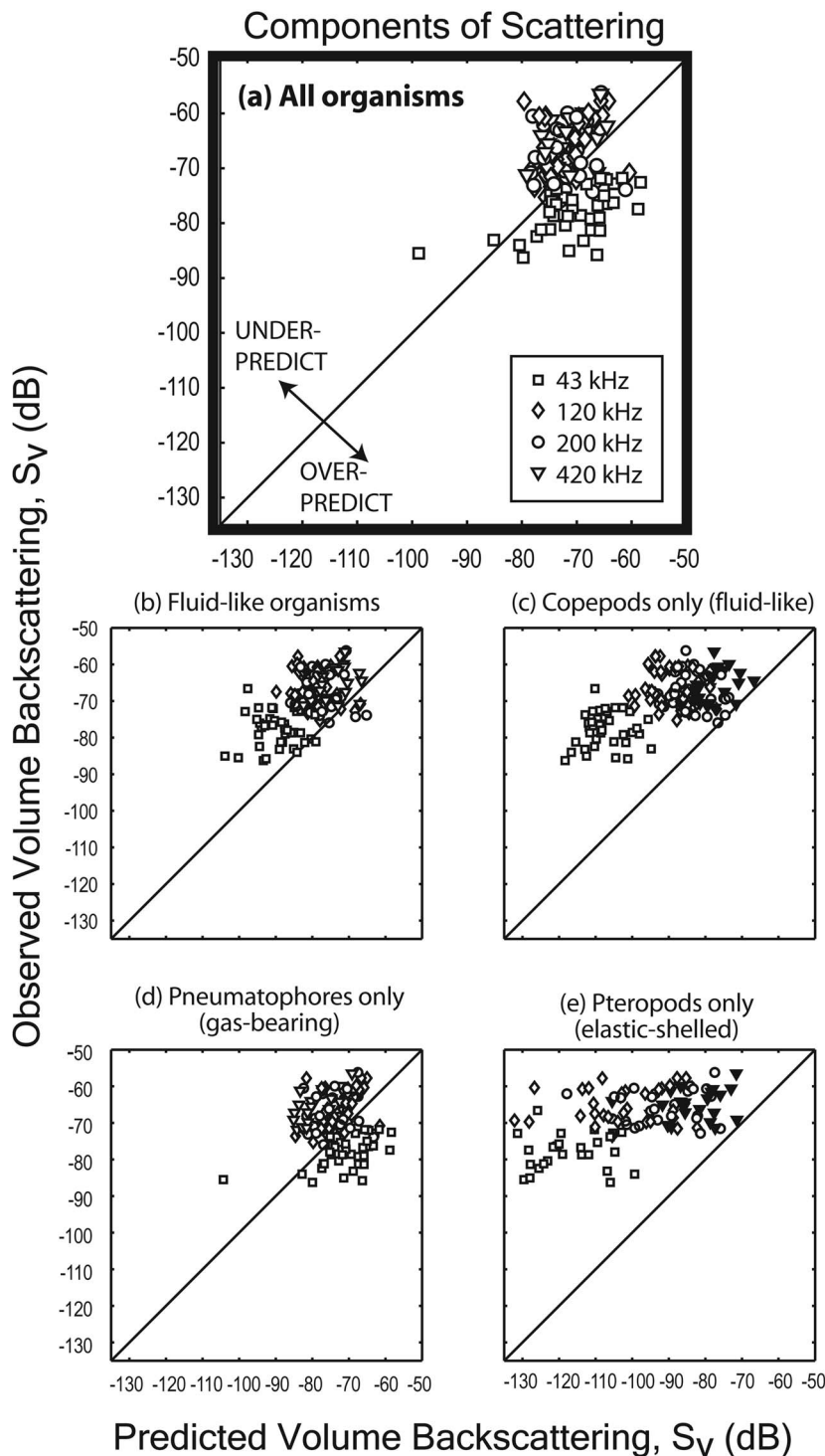


FIG. 12. Observed versus predicted volume backscattering at all frequencies for all MOCNESS tows. (a) All zooplankton, (b) fluid-like zooplankton only, (c) copepods (fluid-like) only, (d) pneumatophores (gas-bearing siphonophores) only, and (e) pteropods (elastic-shelled) only. The 43 kHz data from MOC 2 in October 1999 have been discarded due to suspect noise profiles. The black lines show the one-to-one linear relationship that would be obtained if the predicted and observed scattering agreed perfectly. The solid black triangles in (c) and (d) highlight the 420 kHz data.

pteropods alone was significantly smaller than the observed backscattering at 43 and 120 kHz, regardless of model, but the predicted scattering from pteropods based on both high-pass models (models 1 and 2) was in relatively good agreement with the measured scattering at 200 and 420 kHz. Pteropod model 3 significantly over-predicted the scattering at 200 and 420 kHz. As the predicted scattering from pteropods dominated at 420 kHz at this location and depth, based on all models, a single-frequency inference of pteropod abundance was performed, assuming a known pteropod size

(mean size in MOC4 net 7). The predictions compared favorably to the measured MOCNESS pteropod abundance (Table V).

3. Scattering dominated by gas-bearing zooplankton

Siphonophore pneumatophores were important contributors to volume scattering throughout the deep basins of the GoM. Pneumatophores were particularly abundant at the depth of the seasonal thermocline, where a strong scattering layer was often observed with volume scattering levels

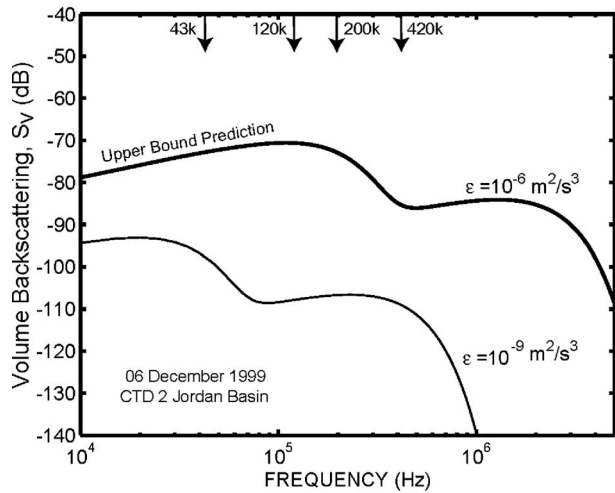


FIG. 13. Predicted volume backscattering for turbulent microstructure as a function of frequency based on parameters estimated from CTD 2 on 6 December 1999. These predictions are based on an assumed value of $\epsilon = 10^{-6} \text{ m}^2/\text{s}^3$ (thick line) and on the maximum temperature and salinity gradients observed at any depth, giving an upper bound prediction. Actual values are expected to fall well below this prediction. Predictions based on $\epsilon = 10^{-9} \text{ m}^2/\text{s}^3$ (thin line) are shown to indicate that the shapes of the curves are generally similar, independent of the exact value of ϵ . The scattering generally decreases across the BIOMAPER-II frequency range.

reaching a maximum at 120 kHz. A particularly strong scattering layer at the seasonal thermocline was observed in Jordan Basin on 6 December 1999. MOCNESS 4 and CTD 2 were performed in close succession at this location, with BIOMAPER-II collecting acoustic data while at the surface (Fig. 15). A second deeper scattering layer was also observed.

Though relatively strong temperature gradients were observed at the depths corresponding to the two scattering layers (Fig. 16(a)), the measured scattering spectra for the layers are not consistent with the predicted scattering from turbulent microstructure. The total biomass and numerical abundance of zooplankton based on the MOCNESS samples reached a minimum at the location of the scattering layer at the seasonal thermocline and a maximum at the location of the deeper scattering layer (Fig. 16(b)). However, there was an elevated numerical abundance of siphonophore pneumatophores in the scattering layer at the seasonal thermocline (Fig. 16(b)), though the numerical abundance of pneumatophores was small ($<4/\text{m}^3$) compared to the total abundance of zooplankton at this depth ($186/\text{m}^3$). This suggests that two different mechanisms (both of biological origin) are giving rise to the scattering layers: the strong and pervasive scattering layer at the seasonal thermocline due predominantly to pneumatophores and the deeper scattering

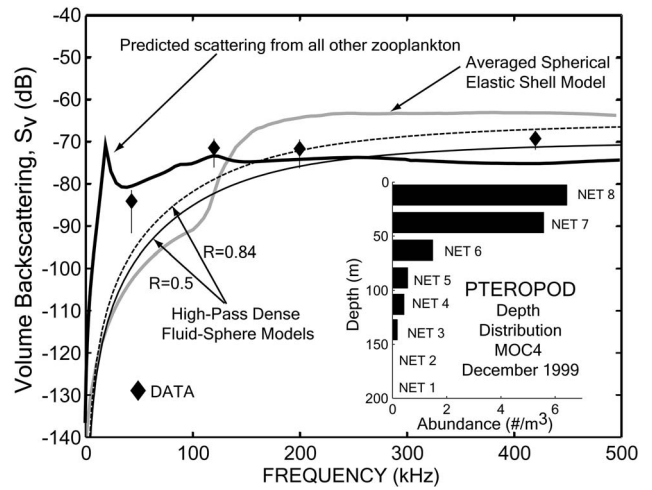


FIG. 14. Predicted volume backscattering in Jordan Basin, GoM, on 6 December 1999 based on the measured distribution of pteropods in MOCNESS 4 Net 7 (25–50 m). Predictions are based on three models: the high-pass dense fluid-sphere model with reflection coefficients of $R=0.5$ (thin solid line) and $R=0.84$ (thin dashed line), and the averaged modal series solution for a spherical aragonite shell (thick gray line). The mean measured S_V values for this depth range are shown by the solid diamonds. The error bars indicate the standard deviation of the mean measured S_V values. The total predicted S_V based on all the remaining zooplankton for this depth stratum is shown (thick black line). The inset shows the vertical distribution of pteropod abundance at this location.

layer due to an elevated abundance of other zooplankton. Though there are differences between the predicted and observed scattering at this location, the predicted scattering if the pneumatophores are not included does not even qualitatively reproduce the observed scattering trends, and underpredicts the scattering at 120 kHz by almost 35 dB (results not shown). Elevated scattering at the seasonal thermocline has been observed at multiple acoustic frequencies previously (e.g., Holliday and Pieper, 1980), though no satisfactory explanation of its origin has been suggested.

As the predicted scattering in the layer at the seasonal thermocline is dominated ($>50\%$ of the total predicted scattering) by pneumatophores at all frequencies, a simple least squares inversion was performed in which the scattering was assumed to arise solely from pneumatophores of a single size, and the radius and abundance were varied to give the best possible fit to the measured S_V data (Fig. 17). The pneumatophore diameter (0.15 mm) obtained by this simple inversion was virtually identical to the median pneumatophore diameter measured by the MOCNESS in that depth stratum. However, the inferred abundance was significantly higher ($216/\text{m}^3$) than the measured abundance ($3.2/\text{m}^3$). These kinds of simple inversions have been performed previously

TABLE IV. Number of MOCNESS nets in which the contribution from copepods, pteropods, and siphonophores to the total predicted volume backscattering of biological origin was larger than 50%. The quantities in the parentheses represent these numbers as a percentage of the total number of nets: The number of MOCNESS tows analyzed in this study was 7, and each tow had eight nets, thus the total number of nets analyzed was 56.

Organism	43 kHz	120 kHz	200 kHz	420 kHz
Copepods (fluid-like)	0 (0%)	2 (3.6%)	7 (12.5%)	22 (39.3%)
Pteropods (elastic-shell)	0 (0%)	0 (0%)	1 (1.8%)	3 (5.4%)
Siphonophore Pneumatophores (gas-bearing)	49 (85.7%)	43 (76.7%)	30 (53.6%)	9 (16.1%)

TABLE V. Acoustically inferred zooplankton size and/or abundance for different zooplankton taxa that dominate the scattering at one or more frequencies. If the scattering was dominated at all frequencies by a single type of scatterer, a four-frequency inversion for size (assuming a single size) and numerical abundance was performed. Otherwise, a single-frequency inversion for numerical abundance was performed assuming a known size and based on the frequency at which the organisms most dominated the total predicted scattering. The measured organism size corresponds to the mean size in the depth bin indicated, estimated by the corresponding sampling method. For copepods and euphausiids size corresponds to length, and for pteropods and pneumatophores size corresponds to diameter. The frequencies that were used to perform the inversions are highlighted. Where applicable, the depth range of the measurements is noted.

Organism	Fraction of total biomass	Fraction of total numerical abundance	Fraction of contribution to scattering at each frequency	Measured Size (mm)	Acoustically inferred size (mm)	Measured abundance (#/m ³)	Acoustically inferred abundance (#/m ³)	Sampling technique, location, date, depth
Copepods ^a (Fluid-like)	99.8%	98.9%	0.1% (43 kHz) 12% (120 kHz) 58% (200 kHz) 91% (420 kHz)	2.37	NA	1290	410	MOCNESS Jordan Basin, GoM December 1999 MOC 4, Net 2: 175–200 m
Copepods ^a (Fluid-like)	97.2%	98.8%	0.03% (43 kHz) 17% (120 kHz) 63% (200 kHz) 92% (420 kHz)	2.22	NA	715	540	MOCNESS Jordan Basin, GoM December 1999 MOC 4, Net 3: 150–175 m
Copepods ^b (Fluid-like)	See note ^c	See note ^c	See note ^c	2.45	NA	930	355 ^d	VPR Jordan Basin, GoM December 1999 175–200 m
Copepods ^b (Fluid-like) (Video: VPR)	See note ^c	See note ^c	See note ^c	2.46	NA	645	370 ^d	VPR Jordan Basin, GoM December 1999 150–175 m
Pteropods ^a (Elastic shell)	5.9%	2.0%	1% (43 kHz) 11% (120 kHz) 40% (200 kHz) 70% (420 kHz)	0.996	NA	5.57	10.6 (model 1) 4.3 (model 2) 0.66 (model 3)	MOCNESS Jordan Basin, GoM December 1999 MOC 4, Net 7: 25–50 m
Siphonophore Pneumatophores ^c (Gas-bearing)	NA	1.7%	99% (43 kHz) 97% (120 kHz) 92% (200 kHz) 65% (420 kHz)	0.15	0.15	3.2	216	MOCNESS Jordan Basin, GoM December 1999 MOC 4, Net 5: 75–100 m
Euphausiids ^f (Fluid-like) (Warren, 2001; Warren <i>et al.</i> , 2003)	7%	<1%	50% (43 kHz) 80% (120 kHz) 75% (200 kHz) 65% (420 kHz)	14.7	15	1	11	MOCNESS Jordan Basin, GoM December 1999 MOC 9, Net 5: 60–80 m

^aAbundance estimated from single-frequency (420 kHz) inversion, assuming a known size.

^b1st VPR tow, performed shortly after MOC 4. Abundance estimated from single-frequency (420 kHz) inversion, assuming a known copepod size, as estimated from the VPR video images.

^cDue to the small VPR sampling volume, the total biomass and numerical abundance of zooplankton were overwhelmingly dominated by copepods. As a result, the VPR-based predictions of volume scattering were also dominated at all frequencies by copepods.

^dThe VPR estimates of copepod length were larger than the MOCNESS (MOC 4) estimates of copepod length, potentially due to ambiguities in the orientation of the copepods in the VPR images, due to the 2-D projection of a 3-D object onto the video image plane. TS = -94.2 dB for a copepods of length 2.45–2.46 mm.

^eAbundance and size inferred from 4-frequency inversion.

^fThese studies involved data collected with the same multi-frequency towed instrument platform, BIOMAPER-II, in an earlier survey of Wilkinson Basin in the Gulf of Maine in October 1997. Direct biological sampling was performed using the MOCNESS to ground truth the acoustic data.

with some success (Warren *et al.*, 2003; Trevorrow *et al.*, 2005). The assumption of a single pneumatophore size is partially justified as the observed pneumatophore size distribution was peaked around a single size. Inversion of the multi-frequency data for the location of the resonance frequency (or the Rayleigh-to-geometric scattering transition) provides a relatively accurate method for determining size. In contrast, absolute scattering levels are subject to the vagaries of calibration, in addition to which there is uncertainty in the abundance measured by net tows due to issues such as avoidance and destruction of fragile individuals. Thus, these simple inversions typically predict size with greater accuracy than these inversions predict abundance.

C. Determining dominant biological scatterers: Video images

Volume backscattering predictions were performed based on the VPR video images collected during the first BIOMAPER-II tow in Jordan Basin on 6 December 1999 (Fig. 15). The biomass and numerical abundance of zooplankton as determined by the VPR were overwhelmingly dominated by copepods at all depths. A total of only 13 images were not copepods, divided almost equally into euphausiids, chaetognaths, ostracods, and larvaceans (all fluid-like scatterers), two elastic-shelled pteropods, and three gas-bearing siphonophores. The predicted volume backscattering

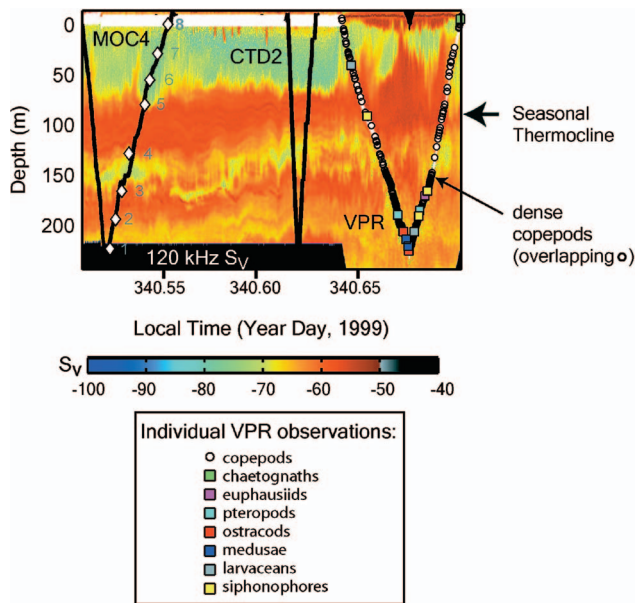


FIG. 15. Volume backscattering at 120 kHz as a function of depth and local time in Jordan Basin, GoM, on 6 December 1999. The MOCNESS 4 and CTD 2 profiles are shown. The depths of the opening and closing of the 8 MOCNESS nets are indicated by open diamonds. The depths of all the individual VPR observations of different zooplankton taxa are marked. MOCNESS 4 was completed at 13:45 while the first BIOMAPER-II towyo was completed at 16:45 (sunset was at 15:57). The BIOMAPER-II track line is indicated in white. The time elapsed from the beginning of the MOCNESS to the end of the first towyo is approximately 4.5 h.

from copepods (Fig. 18), based on the VPR images, is in better agreement with the observed scattering at the higher frequencies and at the deeper depths, where copepods were most abundant, though the predicted scattering at 420 kHz based on just the copepods was a few dB larger than the observed scattering at the deeper depths. The predicted scattering based on the remaining taxa did not contribute significantly to scattering at most depths or frequencies. At the few depths where siphonophores were observed, there are spikes in the predicted scattering, in some cases many dB above the predicted backscattering from copepods, and that were not observed in the measured volume backscattering, a result that can be easily explained by the large differences in the acoustic and VPR sampling volumes. There is little evidence that any type of zooplankton observed by the VPR can account for the scattering layer observed at the seasonal thermocline, though the deeper scattering layer can be explained by an elevated abundance of copepods, consistent with the presence of elevated biomass and numerical abundance at depth caused by overwintering copepods, specifically *Calanus finmarchicus*.

As copepods dominated the biomass, numerical abundance, and total predicted volume backscattering at this location (based on the VPR images), it is possible to perform a simple single-frequency inference of copepod abundance based on the measured volume backscattering and VPR-based estimate of copepod length. As the predicted and measured S_V values were in closest agreement at 420 kHz and below 150 m depths, the numerical abundance of copepods was inferred acoustically at all depths in 1 m bins between 150 and 200 m based on the measured 420 kHz S_V data (Fig.

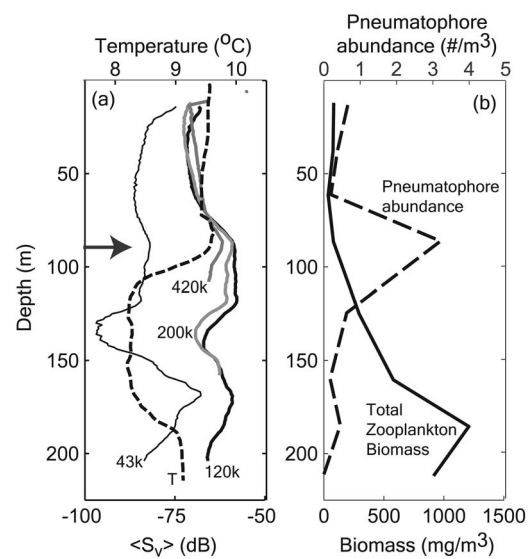


FIG. 16. (a) Mean volume backscattering as a function of depth at 43, 120, 200, and 420 kHz in Jordan Basin, GoM, on 6 December 1999 during the deployment of MOCNESS 4. Two strong scattering layers are observable at 120 kHz: A layer approximately 40 m wide centered at 105 m, and a deeper layer, approximately 20 m wide, centered at 170 m. The arrow indicates the approximate depth of the seasonal thermocline. The volume backscattering at 43 kHz is significantly weaker than at the other frequencies. The deeper layer is not fully observed at 200 kHz, while at 420 kHz the shallower layer is only partially observable. Superimposed on the mean S_V values is the temperature profile obtained by the nearby CTD 2 (thick dashed line corresponding to the scale on the top axis). (b) Total zooplankton biomass estimated from MOCNESS 4 (thick line, bottom scale) and numerical abundance of pneumatophores (dashed line, top scale). It can be seen that there is a peak in the biomass at 185 m, and a peak in the numerical abundance of pneumatophores at 85 m, corresponding to net 5, spanning the depth range from 75 to 100 m.

18(d)) and TS values based on the mean copepod lengths in the 1 m depth bins. These S_V values were also used in Sec. V B 1 to calculate the numerical abundance of copepods based on copepod lengths from the nearby MOCNESS data. For ease of comparison with the numerical abundance based on

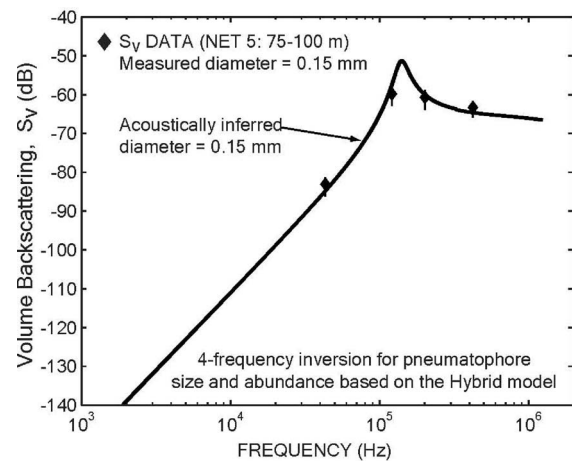


FIG. 17. Observed volume backscattering (diamonds) in Jordan Basin, GoM, on 6 December 1999 in the depth stratum spanning 75–100 m (MOC 4 net 5), and the results of a four-frequency least squares inversion (based on the hybrid model) for pneumatophore radius and abundance (solid line). This simple inversion assumes the scattering is due to pneumatophores of a single size. The error bars indicate the standard deviation of the mean measured S_V values.

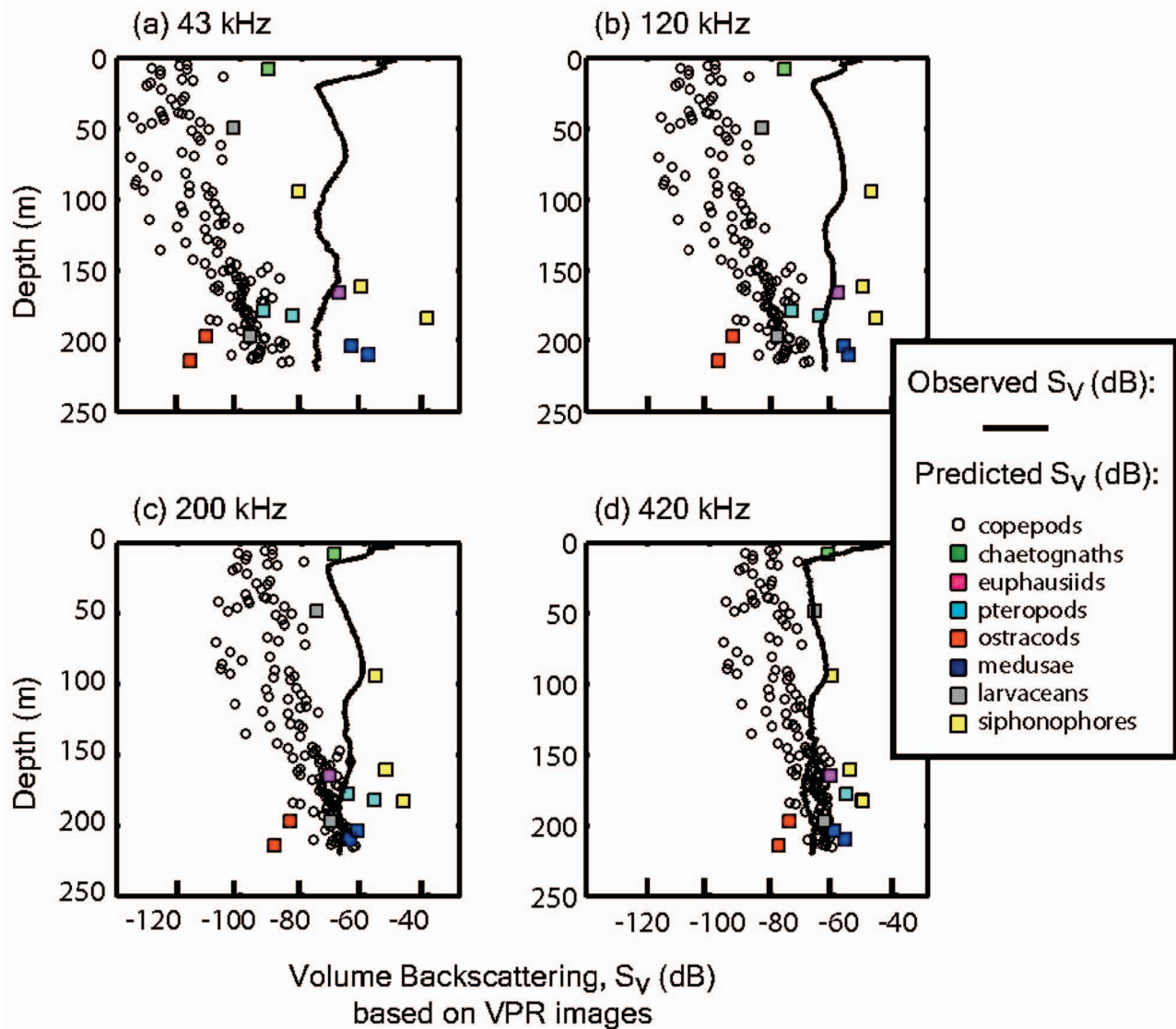


FIG. 18. Predicted volume backscattering as a function of depth based the VPR images for the first tow in Jordan Basin, GoM, on 6 December 1999 at the four BIOMAPER-II frequencies: a) 43 kHz, b) 120 kHz, c) 200 kHz, and d) 420 kHz. The solid lines correspond to the mean measured volume backscattering. The symbols correspond to the predicted volume backscattering, averaged in 1 m depth bins, for the different zooplankton taxa observed in the VPR images. As it was not possible to measure the pneumatophore sizes for the siphonophores observed with the VPR, the pneumatophore sizes used were based on the median pneumatophore size measured in the nearby MOCNESS samples at the corresponding depths.

the nearby MOCNESS data, the numerical abundance has been averaged into depth bins that correspond to the depths of nets 2 and 3 of the nearby MOCNESS. The acoustically inferred numerical abundance of copepods in the 175–200 m depth bin was $355/\text{m}^3$, compared to the VPR-based estimate of $930/\text{m}^3$, and the acoustically inferred numerical abundance of copepods in the 150–175 m depth bin was $370/\text{m}^3$, compared to the VPR-based estimate of $645/\text{m}^3$.

VI. DISCUSSION AND CONCLUSIONS

It has been found that in the mixed zooplankton populations investigated in this study, a single type of scatterer does not often dominate the scattering at all frequencies (Tables IV and V). Instead, low abundances of strong scatterers that do not contribute significantly to overall biomass dominated the scattering over a sub-range of frequencies. For

example, siphonophores with small gas-filled pneumatophores dominated the scattering at the lower frequencies, particularly at 43 kHz, at many locations and depths. Elastic-shelled pteropods that also made a very small contribution to biomass were important scatterers in the near-surface waters at higher frequencies, particularly at 420 kHz. Millimeter-sized weakly scattering copepods were highly abundant and made up the majority of the biomass at many locations, however, their contribution to scattering was not apparent except at the highest frequencies. Most other fluid-like zooplankton, including euphausiids, never dominated the scattering at any frequency.

The accuracy of acoustic scattering models for interpreting the measured volume backscattering data is critical. The scattering model employed in this study for most fluid-like zooplankton was based on the DWBA smoothly-tapered, uniformly-bent cylinder model, an extensively tested formu-

lation. The scattering from siphonophores with gas inclusions was described using a hybrid model that accounts for damping near the bubble resonances and used the actual distribution of measured bubble sizes. Siphonophore body parts were modeled as fluid spheres. Given the uncertainty in the elastic-shell scattering model used for pteropods, several models were used ranging from a semi empirical high-pass dense fluid-sphere model, through to an averaged exact modal series solution for an elastic shell, which has not been applied previously to pteropods. Since there were no regions in which the scattering from pteropods dominated the scattering at all frequencies based on any of these models, it was difficult to assess which scattering model was most accurate. However, at the few locations and frequencies where the scattering was dominated by pteropods, the high-pass models most accurately predicted the observed scattering. The scattering models for all types of zooplankton were challenged by the need for precise model parameters, in particular, the *in situ* characteristics (shape and size) of the siphonophore gas inclusions and the material properties of elastic-shelled pteropods.

In contrast to earlier studies in this region (Warren *et al.*, 2003), there was little evidence that turbulent oceanic microstructure was a significant contributor to volume backscattering. Warren *et al.* were probably able to observe scattering from turbulent microstructure during the passage of an internal wave for a combination of reasons: 1) Their measurements were conducted at a time and location when the stratification was slightly stronger, 2) their measurements were performed in the far eastern part of Wilkinson Basin, a region that may also have been more susceptible to internal waves with elevated dissipation rates of turbulent kinetic energy, and 3) during their measurements, the contribution to scattering of biological origin, at least in certain portions of the internal wave, was sufficiently low that it did not mask the scattering from turbulent microstructure.

The measured scattering at most locations investigated in this study generally increased monotonically with increasing frequency, across the frequency range pertinent to this study, though the scattering at 120 kHz was typically largest for a pervasive scattering layer observed at the seasonal thermocline, consistent with the presence of gas-bearing siphonophores, a conclusion also supported by net-based forward predictions of expected backscattering. An occasional deep scattering layer, most often observed during the daytime, had an almost flat frequency response. There is evidence in the literature that scattering from myctophids, small fish, some with and some without swimbladders, can give rise to a relatively flat, or even decreasing, backscattering spectrum between 38 and 200 kHz (Kloser *et al.*, 2002; Mair *et al.*, 2005). Myctophids, and fish more generally, are known to effectively avoid net systems such as the 1-m² MOCNESS, and very few individuals were captured in the nets. Thus it was not possible to quantitatively account for the scattering from fish in this study. However, the measured scattering observed throughout most of this study was generally consistent with scattering from zooplankton, that is, the scattering generally increased with increasing frequency across the frequency range of interest.

The scattering levels predicted by use of the models and net tow information were approximately 10 dB lower than the observed scattering at most frequencies and depths if the siphonophore pneumatophores were not included. Inclusion of the pneumatophores reduced, but did not completely remove, the discrepancies between the predicted and observed scattering levels. Inclusion of the pneumatophores resulted in a general tendency to over predict the observed scattering at 43 kHz, and most likely is a result of uncertainty in the hybrid scattering model and the high sensitivity of the resonance frequencies to small errors in the measured pneumatophore size, as most resonance frequencies occurred below 120 kHz. Net avoidance and the destruction of fragile organisms are likely to be significant contributors to the general tendency of under predicting the observed scattering.

Though the correlation between observed and predicted scattering was not as good as some other studies (Flagg and Smith, 1989; Wiebe *et al.*, 1997; Ressler, 2002; Fielding *et al.*, 2004), the key difference is that this study includes data from multiple locations, collected at different times of the day and night, at different times of the year, at multiple frequencies, and in heterogeneous zooplankton populations in which one particular type of scatterer does not often dominate at all frequencies. It is also of interest to note that the daily vertical migration of zooplankton, in which some zooplankton taxa migrate but others do not, changes the relative day/night abundances of different zooplankton at any given depth, potentially affecting the balance of dominant scatterers. However, no systematic differences between the predictions based on the day and night tows were observed.

Measured volume backscattering was also compared to model predictions that used coincident video images. Though the predicted scattering based on the video images could not explain all of the vertical layering observed acoustically, there was relatively good agreement between the predictions and the acoustic observations at high frequencies at depths where copepods dominated the biomass. The video-image-based estimates of abundance and depth distribution for copepods were in good agreement with the net tow observations, though the comparison was poor for all other zooplankton taxa. Based on the restricted analysis completed thus far it is concluded that the video imaging technique is suited to providing the ground-truthing information needed to compare predicted to observed scattering only in locations where rare scatterers do not dominate the scattering. More extensive quantitative analysis of the existing video images should allow larger regions of the acoustic data to be interpreted in terms of the abundance and size of copepods.

Two types of inversions of the volume backscattering data were performed: 1) At locations where the predicted scattering was dominated at all frequencies by gas-bearing siphonophores (Table IV), least-squares inversions of the multi-frequency data were performed to obtain both numerical abundance and size (Table V), as has been done previously in the literature (Warren *et al.*, 2003; Trevorrow *et al.*, 2005). Though this multi-frequency inversion produced accurate estimates of size, there were significant discrepancies between the acoustically inferred and directly measured net tow abundance. This difference, involving relatively large

and more mobile zooplankton, is believed to be largely due to net avoidance (Wiebe *et al.*, 2004 and references therein) and destruction of these fragile zooplankton, although there is also error associated with the scattering model and input parameters. 2) At locations where the predicted scattering was dominated at a single frequency by a single zooplankton taxon, namely copepods or pteropods (Table IV), simple single-frequency calculations of abundance, assuming a single known size, were performed (Table V). The acoustically inferred abundances compared quite favorably to net or video samples.

The discrepancy between the acoustically inferred and directly measured net abundance found in this study has not been observed in some other multi-frequency studies involving acoustic sampling volumes many orders of magnitude smaller than the sampling volume employed in the current study (e.g., 0.01 m³ in Pieper *et al.* (1990) versus 21–86 m³ in the current study) and in which large and less abundant zooplankton are less likely to be sampled. As a consequence the primary contributors to the measured scattering are often small and abundant zooplankton, such as copepods, resulting in more accurate estimates of abundance (e.g. Costello *et al.*, 1989; Pieper *et al.*, 1990; Napp *et al.*, 1993). However, larger sampling volumes are better suited to synoptic surveys of large regions and will contain many different taxa that potentially span multiple trophic levels. Under certain restricted conditions, abundance estimates of the smaller, more abundant, and less mobile zooplankton using the large sampling volume synoptic system were also in relatively good agreement (Table V).

In conclusion, one of the main goals of high-frequency acoustic scattering techniques that make use of instruments like the BIOMAPER-II, is to rapidly sample broad areas acoustically for the purpose of making inferences of biological quantities, such as the abundance of zooplankton. It is well known that multi-frequency acoustic scattering techniques can increase the amount of information regarding the distribution of heterogeneous zooplankton populations over relevant spatial and temporal scales, relative to more traditional sampling methods alone. However, in order to infer biological parameters from the acoustic data it is necessary to first determine the dominant scatterers (Table IV), for which direct ground-truthing measurements and accurate scattering models are currently essential. The larger the number of frequencies over which a single type of scatterer dominates, the more biological parameters that can be inferred (Table V). When one type of scatterer only dominates at a single frequency, the mean size of the dominant scatterer obtained by the ground-truthing measurements can be used to acoustically infer the abundance. However, small errors in the measured size, in addition to a broad distribution of sizes of the dominant scatterer, can lead to large errors in the acoustically inferred abundance, particularly when the dominant scatterer is a gas-bearing organism with a relatively strong and narrow resonance frequency, and the frequencies available are close to the resonance frequency (as was the case in this study).

A major contribution of this study has been to show that in mixed zooplankton populations, one type of scatterer does

not often dominate the scattering at all frequencies. In fact, at many locations, no single type of scatterer dominates at any frequency.

Looking to the future, acoustic systems that take advantage of emerging broadband technology will be better placed to address this problem. For example, a broadband system continuously spanning 30–150 kHz might have allowed the pneumatophore resonance frequencies (hence size) to be determined. However, due to the complexities associated with interpreting scattering from mixed zooplankton populations as outlined in this study, even with unlimited bandwidth, well-parameterized and accurate scattering models, and robust *in situ* information, accurate interpretation of volume scattering data will most likely remain restricted to limited conditions.

ACKNOWLEDGMENTS

The authors would like to thank Dezhong Chu for generously providing the computer code for calculating volume backscattering from some zooplankton. The authors would also like to thank Joe Warren for his assistance and participation during the GoM field surveys, and for providing length-to-width measurements of euphausiids specific to the GoM. Finally, comments from two anonymous reviewers led to significant improvement of this manuscript. This research was supported in part by the U.S. GLOBEC program, NOAA (Grant nos. NA17RJ1223 and NA67RJ0148), the James S. Cole and Cecily C. Selby Endowed Funds, the Penzance Endowed Fund for Support of Assistant Scientists, and the Adams Chair at the Woods Hole Oceanographic Institution. A selected number of focused experiments were also funded by the ONR (Grant No. N00014-98-1-0362). This is GLOBEC contribution number 541.

- Anderson, V. C. (1950). "Sound scattering from a fluid sphere," *J. Acoust. Soc. Am.* **22**, 426–431.
- Benfield, M. C., Davis, C. S., Wiebe, P. H., Gallagher, S. M., Lough, R. G., and Copley, N. J. (1996). "Video plankton recorder estimates of copepod, pteropod and larvacean distributions from a stratified region of Georges Bank with comparative measurements from a MOCNESS sampler," *Deep-Sea Res., Part II* **43**(7–8), 1925–1945.
- Benfield, M. C., Wiebe, P. H., Stanton, T. K., Davis, C. S., Gallagher, S. M., and Greene, C. H. (1998). "Estimating the spatial distribution of zooplankton biomass by combining Video Plankton Recorder and single-frequency acoustic data," *Deep-Sea Res., Part II* **45**(7), 1175–1199.
- Benfield, M. C., Davis, C. S., and Gallagher, S. M. (2000). "Estimating the *in situ* orientation of *Calanus finmarchicus* on Georges Bank using the Video Plankton Recorder," *Plankton Biol. Ecol.* **47**(1), 69–72.
- Benfield, M. C., Lavery, A. C., Wiebe, P. H., Greene, C. H., Stanton, T. K., and Copley, N. J. (2003). "Distribution of physonect siphonulae in the Gulf of Maine and their potential as important sources of acoustic scattering," *Can. J. Fish. Aquat. Sci.* **60**, 759–772.
- Brierley, A. S., Ward, P., Watkins, J. L., and Goss, C. (1998). "Acoustic discrimination of southern ocean zooplankton," *Deep-Sea Res., Part II* **45**, 1155–1173.
- Brierley, A. S., Alexsen, B. E., Boyer, D. C., Lyman, C. P., Didcock, C. A., Boyer, H. J., Sparks, C. A. J., Purcell, J. E., and Gibbons, M. J. (2004). "Single target detections of jellyfish," *ICES J. Mar. Sci.* **61**, 383–393.
- Brooks, D. A. (1996). "Physical oceanography of the shelf and slope seas from Cape Hatteras to Georges Bank: A brief overview," *The Northeast Shelf Ecosystem: Assessment, Sustainability, and Management*, edited by K. Sherman, N. A. Jaworski, and T. J. Smayda (Blackwell Science, Cambridge, MA), Chap. 4.
- Burgett, R. L., Hebert, D., and Oakey, N. S. (2001). "Vertical structure of turbulence on the southern flank of Georges Bank," *J. Geophys. Res.*

- 106(C10), 22545–22558.
- Chu, D., Foote, K. G., and Stanton, T. K. (1993). "Further analysis of target strength measurements of Antarctic krill at 38 and 120 kHz: Comparison with deformed cylinder model and inference of orientation distribution," *J. Acoust. Soc. Am.* **93**, 2985–2988.
- Chu, D., and Stanton, T. K. (1998). "Application of pulse compression techniques to broadband acoustic scattering by live individual zooplankton," *J. Acoust. Soc. Am.* **104**(1), 39–55.
- Chu, D., Wiebe, P. H., and Copley, N. J. (2000). "Inference of material properties of zooplankton from acoustic and resistivity measurements," *ICES J. Mar. Sci.* **57**, 1128–1142.
- Chu, D., Wiebe, P. H., Copley, N. J., Lawson, G. L., and Puvanendran, V. (2003). "Material properties of North Atlantic cod eggs and early-stage larvae and their influence on acoustic scattering," *ICES J. Mar. Sci.* **60**(3), 508–515.
- Chu, D., and Wiebe, P. H. (2005). "Measurement of sound-speed and density contrasts of zooplankton in Antarctic waters," *ICES J. Mar. Sci.* **62**, 818–831.
- Costello, J. H., Pieper, R. E., and Holliday, D. V. (1989). "Comparison of acoustic and pump sampling techniques for the analysis of zooplankton distributions," *J. Plankton Res.* **4**(11), 703–709.
- David, P., Guerin-Ancey, O., Oudot, G., and Van Cuyck, J. P. (2001). "Acoustic backscattering from salp and target strength estimation," *Oceanol. Acta* **24**(5), 443–451.
- Davis, C. S., and Wiebe, P. H. (1985). "Macrozooplankton biomass in a warm-core Gulf Stream ring: Time series changes in size structure, taxonomic composition, and vertical distribution," *J. Geophys. Res.* **90**, 8871–8882.
- Davis, D. S., Gallager, S. M., Berman, M. S., Haury, L. R., and Strickler, J. R. (1992). "The video plankton recorder (VPR): Design and initial results," *Adv. Limnol.* **36**, 67–81.
- Diachok, O. (2001). "Interpretation of the spectra of energy scattered by dispersed anchovies," *J. Acoust. Soc. Am.* **110**(6), 2917–2923.
- Endo, Y. (1993). "Orientation of Antarctic krill in an aquarium," *Nippon Suisan Gakkaishi* **59**, 465–468.
- Fielding, S., Griffiths, G., and Roe, H. S. J. (2004). "The biological validation of ADCP acoustic backscatter through direct comparison with net samples and model predictions based on acoustic-scattering models," *ICES J. Mar. Sci.* **60**(2), 184–200.
- Flagg, C. M., and Smith, S. L. (1989). "On the use of the Acoustic Doppler Current profiler to measure zooplankton abundance," *Deep-Sea Res.* **36**, 455–474.
- Foote, K. G., Knudsen, H. P., Vestnes, G., MacLennan, D. N., and Simmonds, E. J. (1987). "Calibration of acoustic instruments for fish-density estimation: A practical guide," *ICES Cooperative Research Report No.* 144.
- Foote, K. G. (1990). "Speed of sound in *Euphausia superba*," *J. Acoust. Soc. Am.* **87**, 1405–1408.
- Gallager, S. M., Davis, C. S., Epstein, A. W., Solow, A., and Beardsley, R. C. (1996). "High-resolution observations of plankton spatial distribution correlated with hydrography in the Great South Channel, Georges Bank," *Deep-Sea Res., Part I* **43**(7-8), 1627–1663.
- Gauthier, G., and Rose, G. A. (2002). "In situ target strength studies on Atlantic redfish (*Sebastes* spp.)," *ICES J. Mar. Sci.* **59**, 805–815.
- Goodman, L. (1990). "Acoustic scattering from ocean microstructure," *J. Geophys. Res.* **95**, 11557–11573.
- Goodman, R. R., and Stern, R. (1962). "Reflection and transmission of sound by elastic spherical shells," *J. Acoust. Soc. Am.* **34**(3), 338–344.
- Gregg, M. (1987). "Diapycnal mixing in the thermocline: A review," *J. Geophys. Res.* **92**, 5249–5286.
- Haury, L. R., McGowan, J. A., and Wiebe, P. H. (1978). "Patterns and processes in the time-space scales of plankton distribution," in *Spatial Pattern in Plankton Communities*, edited by J. H. Steele (Plenum, New York), pp. 277–327.
- Holliday, D. V., and Pieper, R. E. (1980). "Volume scattering strengths and zooplankton distributions at acoustic frequencies between 0.5 and 3 MHz," *J. Acoust. Soc. Am.* **67**(1), 135–146.
- Holliday, D. V., and Pieper, R. E. (1995). "Bioacoustical oceanography at high frequencies," *ICES J. Mar. Sci.* **52**, 279–296.
- Iida, K., Takahashi, R., Tang, Y., Mikua, T., and Sato, M. (2006). "Observations of marine animals using an underwater acoustic camera," *Jpn. J. Appl. Phys., Part I* **45**, 4875–4881.
- Kils, U. (1981). "Swimming behavior, swimming performance and energy balance of Antarctic krill *Euphausia superba*," *BIOMASS Sci. Series* **3**, 1–22.
- Kloser, R. J. (1996). "Improved precision of acoustic surveys of benthopelagic fish by means of a deep-towed transducer," *ICES J. Mar. Sci.* **54**, 407–413.
- Kloser, R. J., Ryan, T., Sakov, P., and Koslow, J. A. (2002). "Species identification in deep water using multiple acoustic frequencies," *Can. J. Fish. Aquat. Sci.* **59**, 1065–1077.
- Korneliussen, R. J. (2000). "Measurement and removal of echo integration noise," *ICES J. Mar. Sci.* **57**, 1204–1217.
- Korneliussen, R. J., and Ona E. (2002). "An operational system for processing and visualizing multi-frequency acoustic data," *ICES J. Mar. Sci.* **59**, 293–313.
- Lalli, C. M., and Gilmer, R. W. (1989). *Pelagic Snails: The Biology of Holoplanktonic Gastropod Mollusks* (Stanford University Press, Stanford, CA).
- Lascara, C. M., Hofmann, E. E., Ross, R. M., and Quetin, L. B. (1999). "Seasonal variability in the distribution of Antarctic krill, *Euphausia superba*, west of the Antarctic Peninsula," *Deep-Sea Res., Part I* **46**, 951–984.
- Lavery, A. C., Stanton, T. K., McGehee, D., and Chu, D. (2001). "Three-dimensional modeling of acoustic backscattering from fluid-like zooplankton," *J. Acoust. Soc. Am.* **111**(3), 1197–1210.
- Lavery, A. C., Schmitt, R. W., and Stanton, T. K. (2003). "High-frequency acoustic scattering from turbulent oceanic microstructure: The importance of density fluctuations," *J. Acoust. Soc. Am.* **114**(5), 2685–2697.
- Lawson, G. L., Wiebe, P. H., Ashjian, C. J., Gallager, S. M., Davis, C. S., and Warren, J. D. (2004). "Acoustically-inferred zooplankton distribution in relation to hydrography west of the antarctic peninsula," *Deep-Sea Res., Part II* **51**, 2041–2072.
- Lawson, G. L., Wiebe, P. H., Ashjian, C. J., Chu, D., and Stanton, T. K. (2006). "Improved parameterization of Antarctic krill target strength models," *J. Acoust. Soc. Am.* **119**(1), 232–242.
- Mair, A. M., Fernandes, P. G., Lebourges-Dhaussy, A., and Brierley, A. S. (2005). "An investigation into the zooplankton composition of a prominent 38 kHz scattering layer in the North Sea," *J. Plankton Res.* **27**(7), 623–633.
- McGehee, D. E., O'Driscoll, R. L., and Traykovski, L. V. M. (1998). "Effects of orientation on acoustic scattering from Antarctic krill at 120 kHz," *Deep-Sea Res., Part II* **45**(7), 1273–1294.
- Miyashita, K., Aoki, I., and Inagaki, T. (1996). "Swimming behaviours and target strength of isada krill (*Euphausia pacifica*)," *ICES J. Mar. Sci.* **53**, 303–308.
- Monger, B. C., Chinniah-Chandy, S., Meir, E., Billings, S., Greene, C. H., and Wiebe, P. H. (1998). "Sound scattering by the gelatinous zooplankters *Aequorea victoria* and *Pleurobrachia bachei*," *Deep-Sea Res., Part II* **45**(7), 1255–1271.
- Mutlu, E. (1996). "Target strength of the common jellyfish (*Aurelia aurita*): A preliminary experimental study with a dual-beam acoustic system," *ICES J. Mar. Sci.* **53**(2), 309–311.
- Napp, J. M., Ortner, P. B., Pieper, R. E., and Holliday, D. V. (1993). "Biovolume-size spectra of epipelagic zooplankton using a multi-frequency Acoustic Profiling System (MAPS)," *Deep-Sea Res., Part I* **40**(3), 445–459.
- Oakey, N. S. (1982). "Determination of the rate of turbulent energy from simultaneous temperature and shear microstructure measurements," *J. Phys. Oceanogr.* **12**, 256–271.
- Ortner, P. B., Cummings, S. R., Afring, R. P., and Edgerton, H. E. (1979). "Silhouette photography of oceanic zooplankton," *Nature (London)* **277**, 50–51.
- Pershing, A. J., Wiebe, P. H., Manning, J. P., and Copley, N. J. (2001). "Evidence for vertical circulation cells in the well-mixed area of Georges Bank and their biological implication," *Deep-Sea Res., Part II* **48**(1-3), 283–310.
- Pieper, R. E., Holliday, D. V., and Kleppel, G. S. (1990). "Quantitative zooplankton distributions from multifrequency acoustics," *J. Plankton Res.* **12**(2), 443–441.
- Pieper, R. E., McGehee, D. E., Greenlaw, C. F., and Holliday, D. V. (2001). "Acoustically measured seasonal patterns of zooplankton in the Arabian sea," *Deep-Sea Res., Part II* **48**(6-7), 1325–1343.
- Ressler, P. (2002). "Acoustic backscatter measurements with a 153 kHz ADCP in the northeastern Gulf of Mexico: Determination of dominant zooplankton and micronekton scatterers," *Deep-Sea Res., Part I* **49**, 2035–2051.
- Ross, T., and Lueck, R. (2003). "Sound scattering from oceanic turbulence,"

- Geophys. Res. Lett. **30**(6), 1344.
- Ross, T., Garrett, C., and Lueck, R. (2004). "On the turbulent co-spectrum of two scalars and its effect on the acoustic scattering from oceanic turbulence," *J. Fluid Mech.* **514**, 107–119.
- Rothschild, B. J., and Osborn, T. R. (1988). "Small-scale turbulence and plankton contact rates," *J. Plankton Res.* **10**, 465–474.
- Seim, H. E., Gregg, M. C., and Miyamoto, R. T. (1995). "Acoustic backscatter from turbulent microstructure," *J. Atmos. Ocean. Technol.* **12**(2), 367–380.
- Seim, H. E. (1999). "Acoustic backscatter from salinity microstructure," *J. Atmos. Ocean. Technol.* **16**(11), 1491–1498.
- Seuront, L., Schmitt, F., and Lagadeuc, Y. (2001). "Turbulence intermittency, small-scale phytoplankton patchiness and encounter rates in plankton: Where do we go from here?" *Deep-Sea Res., Part I* **48**(5), 1199–1215.
- Stanton, T. K. (1989). "Simple approximate formulas for backscattering of sound by spherical and elongated objects," *J. Acoust. Soc. Am.* **86**(4), 1499–1510.
- Stanton, T. K., Chu, D., Wiebe, P. H., and Clay, C. S. (1993). "Average echoes from randomly oriented random-length finite cylinders: Zooplankton models," *J. Acoust. Soc. Am.* **94**, 3463–3472.
- Stanton, T. K., Wiebe, P. H., Chu, D., Benfield, M. C., Scanlon, L., Martin, L., and Eastwood, R. L. (1994). "On acoustic estimates of zooplankton biomass," *ICES J. Mar. Sci.* **51**, 505–512.
- Stanton, T. K., Chu, D., Wiebe, P. H., Martin, L. V., and Eastwood, R. L. (1998a). "Sound scattering by several zooplankton groups. I. Experimental determination of dominant scattering mechanisms," *J. Acoust. Soc. Am.* **103**(1), 225–235.
- Stanton, T. K., Chu, D., and Wiebe, P. H. (1998b). "Sound scattering by several zooplankton groups. II. Scattering models," *J. Acoust. Soc. Am.* **103**(1), 236–253.
- Stanton, T. K., and Chu, D. (2000). "Review and recommendations for the modeling of acoustic scattering by fluid-like elongated zooplankton: Euphausiids and copepods," *ICES J. Mar. Sci.* **57**, 793–807.
- Stanton, T. K., Chu, D., Wiebe, P. H., Eastwood, R. L., and Warren, J. D. (2000). "Acoustic scattering by benthic and planktonic shelled animals," *J. Acoust. Soc. Am.* **108**(2), 535–550.
- Thorpe, S. A. (1977). "Turbulence and mixing in a Scottish loch," *Philos. Trans. R. Soc. London, Ser. A* **286**, 125–181.
- Thorpe, S. A., and Brubaker, J. M. (1983). "Observations of sound reflection by temperature microstructure," *Limnol. Oceanogr.* **28**, 601–613.
- Trevorrow, M. V., and Tanaka, Y. (1997). "Acoustic and *in situ* measurements of freshwater amphipods (*Jesogammarus annandalei*) in Lake Biwa, Japan," *Limnol. Oceanogr.* **42**(1), 121–132.
- Trevorrow, M. V., Mackas, D. L., and Benfield, M. C. (2005). "Comparison of multifrequency acoustic and *in situ* measurements of zooplankton abundance in Knight Inlet, British Columbia," *J. Acoust. Soc. Am.* **117**(6), 3574–3588.
- Warren, J. D., Stanton, T. K., Benfield, M. C., Wiebe, P. H., Chu, D., and Sutor, M. (2001). "In situ measurements of acoustic target strength of gas-bearing siphonophores," *ICES J. Mar. Sci.* **58**, 740–749.
- Warren, J. D. (2001). "Estimating Gulf of Maine zooplankton distributions using multiple frequency acoustics, video, and environmental data," Ph.D. thesis, MIT/WHOI, Cambridge, MA, 2001–2003.
- Warren, J. D., Stanton, T. K., Wiebe, P. H., and Seim, H. E. (2003). "Inference of biological and physical parameters in an internal wave using multiple-frequency, acoustic-scattering data," *ICES J. Mar. Sci.* **60**(5), 1033–1046.
- Weston, D. E. (1967). "Sound propagation in the presence of bladder fish," in *Underwater Acoustics*, edited by V. Albers (Plenum, New York), Vol. 2, Chap. 5, pp. 55–88.
- Wiebe, P. H., Morton, A. W., Bradley, A. M., Backus, R. H., Craddock, J. E., Barber, V., Cowles, T. J., and Flierl, G. R. (1985). "New developments in the MOCNESS, an apparatus for sampling zooplankton and micronekton," *Mar. Biol. (Berlin)* **87**(3), 313–323.
- Wiebe, P. H., Mountain, D. G., Stanton, T. K., Greene, C. H., Lough, G., Kaartvedt, S., Dawson, J., and Copley, N. (1996). "Acoustical study of the spatial distribution of plankton on Georges Bank and the relationship between volume backscattering strength and the taxonomic composition of the plankton," *Deep-Sea Res., Part II* **43**(7-8), 1971–2001.
- Wiebe, P. H., Stanton, T. K., Benfield, M. C., Mountain, D. G., and Greene, C. H. (1997). "High-frequency acoustic volume backscattering in the Georges Bank coastal region and its interpretation using scattering models," *IEEE J. Ocean. Eng.* **22**(3), 445–464.
- Wiebe, P. H., Stanton, T. K., Greene, C. H., Benfield, M. C., Sosik, H. M., Austin, T., Warren, J. A., and Hammar, T. (2002). "BIOMAPER II: An integrated instrument platform for coupled biological and physical measurements in coastal and oceanic regimes," *IEEE J. Ocean. Eng.* **27**, 700–716.
- Wiebe, P. H., Ashjian, C. J., Gallager, S. M., Davis, C. S., Lawson, G. L., and Copley, N. J. (2004). "Using a high-power strobe light to increase the catch of Antarctic krill," *Mar. Biol. (Berlin)* **144**, 493–502.

Full wave-field reflection coefficient inversion

Jan Dettmer^{a)} and Stan E. Dosso

School of Earth and Ocean Sciences, University of Victoria, Victoria, BC V8W 3P6, Canada

Charles W. Holland

The Pennsylvania State University, Applied Research Laboratory, State College, Pennsylvania 16804-0030

(Received 18 April 2007; revised 4 September 2007; accepted 6 September 2007)

This paper develops a Bayesian inversion for recovering multilayer geoacoustic (velocity, density, attenuation) profiles from a full wave-field (spherical-wave) seabed reflection response. The reflection data originate from acoustic time series windowed for a single bottom interaction, which are processed to yield reflection coefficient data as a function of frequency and angle. Replica data for inversion are computed using a wave number-integration model to calculate the full complex acoustic pressure field, which is processed to produce a commensurate seabed response function. To address the high computational cost of calculating short range acoustic fields, the inversion algorithms are parallelized and frequency averaging is replaced by range averaging in the forward model. The posterior probability density is interpreted in terms of optimal parameter estimates, marginal distributions, and credibility intervals. Inversion results for the full wave-field seabed response are compared to those obtained using plane-wave reflection coefficients. A realistic synthetic study indicates that the plane-wave assumption can fail, producing erroneous results with misleading uncertainty bounds, whereas excellent results are obtained with the full-wave reflection inversion. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2793609]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT]

Pages: 3327–3337

I. INTRODUCTION

Geoacoustic inversion provides estimates of physical seabed parameters without the need of direct measurements (e.g., coring). Knowledge of these parameters is of interest to a wide range of applications in underwater acoustics, including acoustic propagation modeling and source localization. Single-bounce reflectivity measurements have been developed to resolve local sediment structure using plane wave approximations.^{1–3} Holland and Osler⁴ developed an experimental technique to measure high-resolution seabed reflectivity in shallow water, using a moving impulsive source (e.g., seismic boomer). For impulsive sources, Stickler⁵ and Schmidt and Jensen⁶ showed that the plane wave approximation is often violated, and suggested taking spherical wave effects into account when calculating reflection coefficients. Holland and Osler⁴ analyzed their data including spherical wave effects in the ray theory approximation, i.e., accounting for the various angles at subbottom interfaces and refractions that are geometry dependent.

This paper develops a reflection coefficient inversion for geoacoustic parameters including spherical wave effects in a full-wave-field approach. Quantitative inversion results for the full wave-field forward model are compared to inversion results obtained with the plane wave approximation for a realistic simulated experiment. The full wave-field forward model uses wave number integration (OASES⁷) and two point ray tracing to compute the reflectivity for complex, multilayered seabed models and arbitrarily layered water col-

umns. The geoacoustic model parameters consist of layer thickness, sound velocity, density, and attenuation for multiple (homogeneous) layers.

The inverse problem is solved using a Bayesian approach that samples the posterior probability density (PPD), incorporating both data and prior information. The solution is quantified in terms of properties of the PPD representing parameter estimates and parameter uncertainties (e.g., marginal distributions, credibility intervals), which can be computed using numerical methods for nonlinear inverse problems.^{8–11}

The remainder of this paper is organized as follows. Section II considers the single-bounce reflectivity experiment and the data processing to yield reflection coefficients versus frequency and angle. Section III considers the Bayesian formulation of the inverse problem, data misfit functions, and methods for treating errors. Section IV develops the Bayesian reflection inversion, including data processing and averaging schemes, and the full wave-field forward model. Section V A presents results for a simulated experiment using plane-wave inversion. Section V B presents inversion results for the same data, but using the full wave-field inversion developed here. Finally, Sec. VI summarizes and discusses the results.

II. EXPERIMENT AND DATA PROCESSING

The experiment considered here (Fig. 1) involves the configuration of Holland and Osler,⁴ and consists of a single, fixed hydrophone \mathbf{r} and an impulsive broadband moving source \mathbf{s} . The source is towed near the water surface at constant speed and transmits at uniform time intervals. Seismoa-

^{a)}Electronic mail: jand@uvic.ca

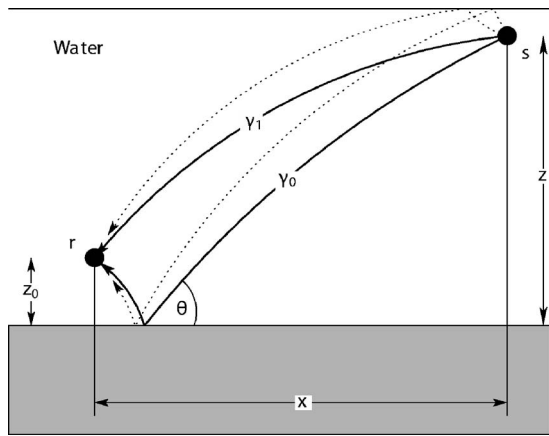


FIG. 1. Experiment geometry to measure reflection coefficients vs grazing angle θ . The direct (γ_1) and the bottom bounce path (γ_0) are shown as solid lines, the surface reflected paths as dotted lines. Source s moves away from receiver r which is mounted at height z_0 above the seabed and transmits at ranges x .

oustic time series (traces) are recorded for uniform increments in source–receiver range x . The area where the acoustic rays (γ_0) interact with the seabed defines the lateral footprint of the experiment. The radius of this footprint is typically of the order of 10^2 m, and the seabed is assumed to be laterally homogeneous over this area.⁴ Hence, the method averages over a seabed area in which no smaller lateral variations can be resolved. Because of the local experiment scale, effects of spatial and temporal variability in the water column and seabed are generally small.

For the inversions, the experiment was simulated by calculating full wave-field synthetic seismograms using OASES.⁷ The underlying geoaoustic model consists of six sediment layers over a sediment half-space (Fig. 2). Each layer is defined by layer thickness h , sound velocity c , den-

sity ρ , and attenuation coefficient α . A single receiver is located at a depth of 122 m in a 150 m water column. The source was simulated at 0.35 m depth using the pulse shown in Fig. 2, transmitting at ranges of 20–560 m with a range spacing of 4.7 m. The source function was taken from experimental data by averaging the direct arrival of several traces recorded at small ranges from a boomer source¹² and then low-pass filtering with a 3000 Hz cutoff. Green’s functions for all ranges were calculated for the seabed model (Fig. 2) and then convolved with the source function. Note that the source pulse length is of the order of the layer thicknesses of the geoaoustic model, limiting the resolution of the data for seabed structure. The simulated seismoacoustic traces are shown in Fig. 3 as a function of two way travel time (TWT) and range. The direct arrival or water wave can be seen at 0.075 s and the shortest range in Fig. 3. Next, at 0.120 s, is the sediment–water interface reflection, followed by a number of subbottom reflections. The data contain no water-column multiples as the sea surface was omitted in the simulation as the source pulse included sea surface interaction for the shallow source. For the experiment geometry, water-column multiples would be well separated from the primary arrivals and would not be involved in the analysis.

The processing used to compute the reflection coefficient data from the simulated seismoacoustic traces follows Holland.¹³ The seismoacoustic traces are time windowed into one packet that contains the reflection effects of the subbottom reflectors and another packet that contains the direct wave. These packets are then processed while accounting for various experimental effects, such as source directivity, geometric spreading, refraction, absorption, and the first sea–surface interaction, to yield reflection coefficients as a function of grazing angle and frequency.¹³ In processing time series to obtain the measured reflection coefficients, the data

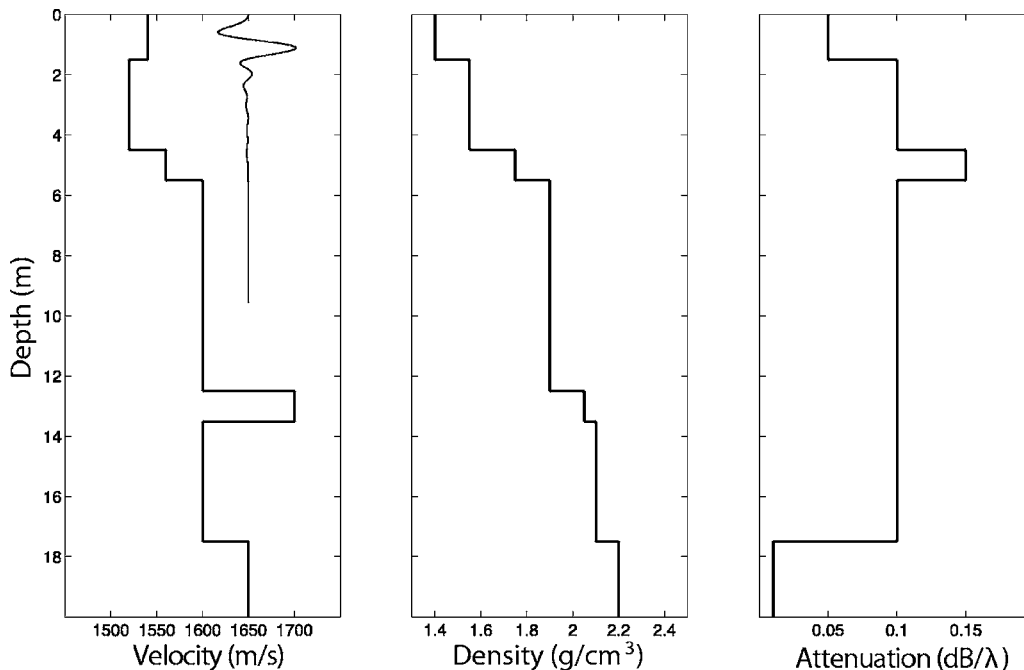


FIG. 2. Seabed profiles with six layers over a half-space for sound velocity, density, and attenuation. Also shown is the source pulse, converted to depth for a sound velocity of 1511 m/s. Note that the source pulse length is comparable to the layer thicknesses in the seabed model.

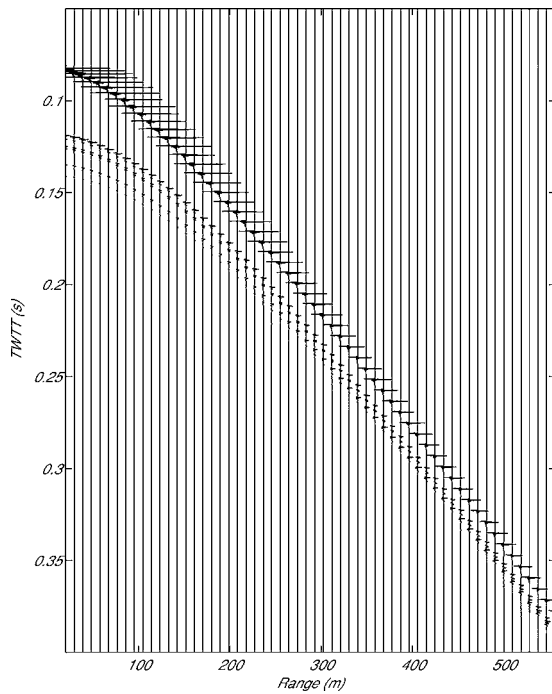


FIG. 3. Synthetic seismoacoustic traces for the seabed model and source pulse shown in Fig. 2. The direct or water wave and several seabed reflections are visible.

are available at many frequencies and a Gaussian frequency average is applied to improve the signal to noise ratio.¹⁴ The data are processed for five frequency bands with center frequencies between 400 and 1600 Hz and fractional bandwidth $\kappa=1/20$ (i.e., the bandwidth is 1/20 of the center frequency). Each frequency band contains data at grazing angles of θ

$=15-75^\circ$. The experiment results in a nonuniform angle spacing, but the data are resampled to a uniform spacing. Random Gaussian noise of standard deviation 0.05 is added to the reflection coefficient data (representative of error levels on measured reflection-coefficient data¹⁰). The reflection coefficient data derived from the seismoacoustic traces are shown in Fig. 4.

III. INVERSION TECHNIQUE

A. Bayesian inversion

This section provides a brief overview of the Bayesian formulation used here for the geoacoustic inversion,⁸⁻¹² more general treatments of Bayesian theory can be found elsewhere.^{15,16} Bayes' rule can be written

$$P(\mathbf{m}|\mathbf{d}) = \frac{\mathcal{L}(\mathbf{d}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}, \quad (1)$$

where $\mathbf{m} \in \mathbb{R}^M$ and $\mathbf{d} \in \mathbb{R}^N$ are random variables that represent the seabed model parameters and data, respectively. $P(\mathbf{m}|\mathbf{d})$ is the PPD, $\mathcal{L}(\mathbf{d}|\mathbf{m})$ is the likelihood function, $P(\mathbf{m})$ is the model prior distribution, and $P(\mathbf{d})$ is the data prior distribution (a constant factor once the data are measured). The likelihood function can generally be expressed as $\mathcal{L}(\mathbf{m}) \propto \exp(-E(\mathbf{m}))$ where $E(\mathbf{m})$ is an appropriate data error function (considered later). Equation (1) then becomes

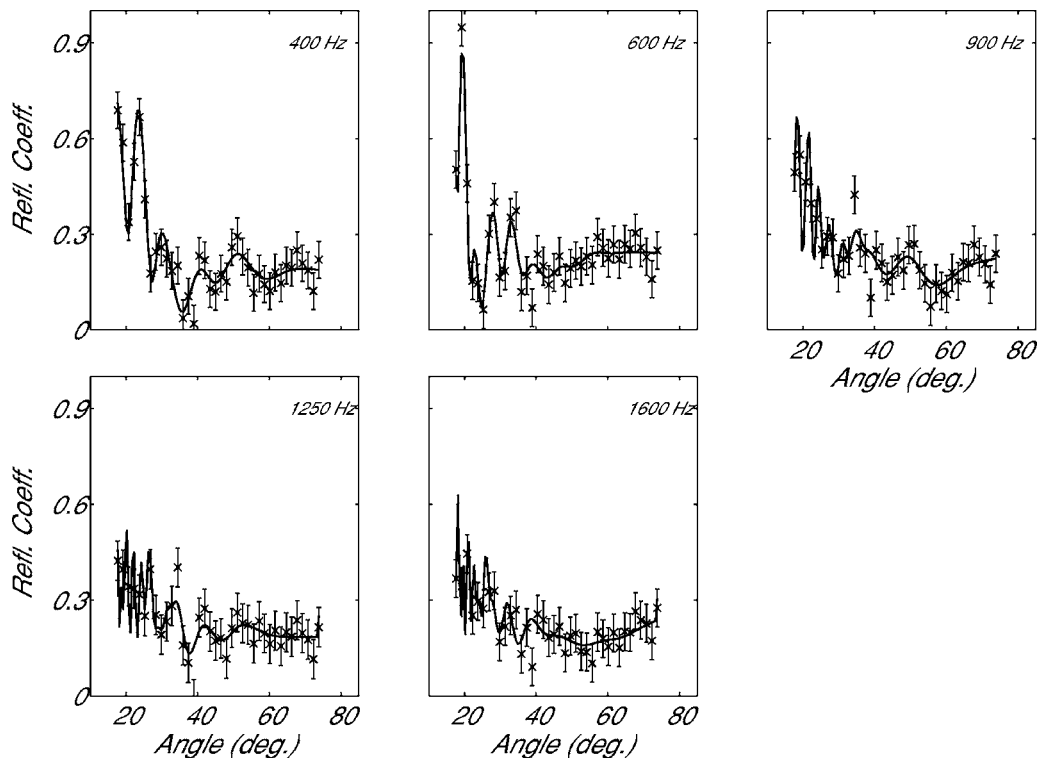


FIG. 4. Reflection coefficient data (symbols) with one standard deviation error bars as derived from the seismoacoustic traces in Fig. 3. For clarity, only every third datum is shown. Solid lines are the replica data generated for the optimal seabed model from full wave-field inversion.

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp(-\phi(\mathbf{m}))}{\int_{\mathcal{M}} \exp(-\phi(\mathbf{m}'))d\mathbf{m}'}, \quad (2)$$

where the integration is over the model space $\mathcal{M} \subset \mathbb{R}^M$ and $\phi(\mathbf{m}) = E(\mathbf{m}) - \log_e P(\mathbf{m})$ is the generalized misfit. The PPD represents the full solution to the inverse problem in the Bayesian formulation. However, due to the PPD's multidimensional nature, interpretation is non-trivial and properties such as the maximum *a posteriori* (MAP) model $\hat{\mathbf{m}}$, the mean model $\bar{\mathbf{m}}$, the model covariance matrix $\mathbf{C}^{(\mathbf{m})}$, and marginal probability distributions $P(m_i|\mathbf{d})$, must be calculated to quantify parameter estimates, uncertainties and interrelationships:

$$\hat{\mathbf{m}} = \text{Arg}_{\text{max}} P(\mathbf{m}|\mathbf{d}) \quad (3)$$

$$\bar{\mathbf{m}} = \int_{\mathcal{M}} \mathbf{m}' P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}' \quad (4)$$

$$\mathbf{C}^{(\mathbf{m})} = \int_{\mathcal{M}} (\mathbf{m}' - \bar{\mathbf{m}})(\mathbf{m}' - \bar{\mathbf{m}})^T P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}' \quad (5)$$

$$P(m_i|\mathbf{d}) = \int_{\mathcal{M}} \delta(m'_i - m_i) P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}', \quad (6)$$

where δ denotes the Dirac delta function. Higher-dimensional marginal distributions can be defined similar to Eq. (6). Uncertainties of parameter estimates can also be quantified in terms of highest probability density credibility intervals. The $\beta\%$ HPD interval is defined as the interval of minimum width that contains $\beta\%$ of the area of the marginal probability distribution. Interrelations of model parameters can be quantified by the model correlation matrix $R_{ij} = C_{ij}/(C_{ii}C_{jj})^{1/2}$. Although analytic solutions to Eqs. (3)–(6) exist for linear inverse problems, nonlinear problems such as geoacoustic inversion must be solved numerically.

MAP estimates, Eq. (3), can be found by numerical minimization of $\phi(\mathbf{m})$, such as adaptive simplex simulated annealing (ASSA), an efficient hybrid optimization algorithm that combines the local downhill-simplex method with a very fast simulated annealing global search.⁸ The integrals of Eqs. (4)–(6) are computed here using the Markov-chain Monte Carlo method of fast Gibbs sampling^{9,10} (FGS) to sample $\phi(\mathbf{m})$. FGS applies an adaptive Metropolis Gibbs sampling scheme^{17,18} in a principal-component parameter space where the coordinate axes align with the dominant correlation directions. The rotation matrix is obtained during an initial *burn-in*¹⁹ phase where the samples are not used for the final PPD estimate. Convergence of the burn-in phase is monitored by running independent samples and intercomparing their correlation matrices. Once a satisfactory rotation matrix is obtained, sampling of the PPD begins using a number of independent samples with convergence monitored in terms of the difference between the marginal distributions computed for each sample. The final integral estimates are based on the union of all samples collected after burn-in.

B. Massively parallel inversion algorithms

ASSA and FGS have been optimized on scalar computers in the past to perform inversions efficiently.^{8,9} However, the speed of the algorithm depends strongly on the performance of the forward model. To conduct research using computationally demanding forward models more efficiently, the ASSA and FGS codes were implemented for massively parallel computers using the message-passing interface.²⁰ The performance gain for parallel implementations depends strongly on the granularity of the algorithm, i.e., how independently different processes in the algorithm can perform computations without the need of intercommunication. As ASSA optimization is essentially a serial algorithm, its granularity is small. The main performance gain for optimization was obtained by parallelizing the forward model by processing many independent acoustic frequencies simultaneously.

As FGS is a Monte Carlo method, it belongs to a group of algorithms that are commonly referred to as *embarrassingly parallel* and have large granularity. FGS was implemented for parallel computers by distributing the sampling of the PPD over a large number of central processing units (CPUs). For the burn-in phase, each CPU samples for a certain number of forward steps and then passes the sample to the master CPU to carry out a convergence test of the various model parameter correlation estimates. Once convergence of the rotation matrix is obtained, the actual sampling of the PPD begins in parallel on all CPUs. It was found that after the burn-in, the performance scaled almost linearly with the number of CPUs used. The algorithm was tested with up to 80 CPUs with a speedup factor of close to 80, reducing total run times from several days to less than 1 h.

C. Likelihood function

Formulating the likelihood function $\mathcal{L}(\mathbf{m})$ requires specifying the data uncertainty distribution, including both measurement errors and theory errors. In general, the Bayesian inversion outlined earlier can be applied with arbitrary uncertainty distributions. In practice, however, the lack of specific knowledge about error distributions often suggests a mathematically simple distribution be assumed. In particular, Gaussian distributions are commonly considered, with their statistical parameters estimated from the data.

Consider reflection coefficient data as a function of grazing angle θ (Fig. 1), given at several independent frequencies. For N observed data \mathbf{d}_i at each of F frequencies with unbiased Gaussian-distributed random errors, the likelihood function is given by

$$\mathcal{L}(\mathbf{m}) \propto \prod_{i=1}^F \exp\left(-\frac{1}{2}(\mathbf{d}_i - \mathbf{d}_i(\mathbf{m}))^T (\mathbf{C}_i^{(\mathbf{d})})^{-1} (\mathbf{d}_i - \mathbf{d}_i(\mathbf{m}))\right), \quad (7)$$

where $\mathbf{d}_i(\mathbf{m})$ are the replica data computed for model \mathbf{m} and $\mathbf{C}_i^{(\mathbf{d})}$ are the $F \times N \times N$ data covariance matrices.^{11,21,22} The data misfit function is given by the negative log likelihood

$$E(\mathbf{m}) = \sum_{i=1}^F \frac{1}{2} (\mathbf{d}_i - \mathbf{d}_i(\mathbf{m}))^T (\mathbf{C}_i^{(d)})^{-1} (\mathbf{d}_i - \mathbf{d}_i(\mathbf{m})). \quad (8)$$

In practical applications, the covariance matrices $\mathbf{C}_i^{(d)}$ are not known. Depending on the data error distribution, different approaches can be applied to estimate the covariance matrices. If the data do not have significantly correlated errors (as in this simulation), the covariance matrices can be approximated as diagonal, $\mathbf{C}_i^{(d)} = \sigma_i^2 \mathbf{I}$, where \mathbf{I} is the identity matrix and σ_i is the standard deviation at the i th frequency. The likelihood function then can be written

$$\mathcal{L}(\mathbf{m}) \propto \prod_{i=1}^F \exp\left(-\sum_{j=1}^N \frac{(d_{ij} - d_{ij}(\mathbf{m}))^2}{2\sigma_i^2}\right), \quad (9)$$

where d_{ij} represents the j th datum at the i th frequency. A maximum likelihood (ML) estimate for the σ_i can be found by maximizing $\mathcal{L}(\mathbf{m})$ over σ_i and \mathbf{m} to yield

$$\tilde{\sigma}_i = \left(\frac{1}{N} \sum_{j=1}^N (d_{ij} - d_{ij}(\tilde{\mathbf{m}}))^2\right)^{1/2}, \quad (10)$$

where the ML estimate $\tilde{\mathbf{m}}$ is found by minimizing the misfit obtained by substituting Eq. (10) into Eq. (9) to yield

$$E(\mathbf{m}) = \sum_{i=1}^F \left(\frac{N}{2} \log_e \left[\sum_{j=1}^N (d_{ij} - d_{ij}(\mathbf{m}))^2\right]\right). \quad (11)$$

The ML estimate $\tilde{\sigma}_i$ can then be used in sampling the PPD by applying the data misfit function

$$E(\mathbf{m}) = \sum_{i=1}^F \sum_{j=1}^N \frac{(d_{ij} - d_{ij}(\mathbf{m}))^2}{2\tilde{\sigma}_i^2}. \quad (12)$$

IV. FORWARD MODELING

A. Frequency and range averaging

The full wave-field forward model applied here to compute replica data is computationally intensive. This, combined with the number of forward computations needed for a typical ASSA or FGS run (in the order of 100 000 seabed models), can make it slow to carry out inversions based on reflection coefficients at multiple frequencies, even on sophisticated supercomputers. Further, the measured data are frequency averaged to improve the signal-to-noise ratio. Due to the strong frequency dependence of the reflection coefficient, the forward model must also include similar frequency averaging for the replica data. Depending on the bandwidth and the particular environmental model, the number of frequencies required for averaging varies. In most cases, 8 or 9 frequencies have proven to provide sufficient averaging.

Figure 5 shows the reflection coefficient over range and frequency for the seabed model in Fig. 2 consisting of six layers over a semiinfinite half-space. The cross shown in Fig. 5 suggests that averaging over a certain range interval could have a similar effect to averaging over a certain band of frequencies. The advantage of averaging over range rather than frequency is that the wave number integration used to calculate the replica acoustic fields automatically provides

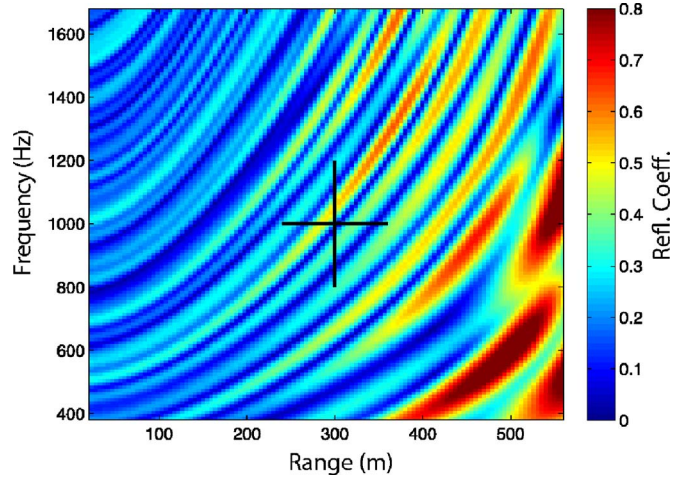


FIG. 5. (Color online) Simulated reflection coefficients over range and frequency for the geoacoustic model of Fig. 2. The cross indicates a fractional bandwidth of $\kappa=5$ in frequency and range.

the field at a dense range spacing. Thus, spacial averaging can be carried out at almost no extra computational cost, whereas the computational cost of frequency averaging scales with the number of frequencies used in the average.

Harrison and Harrison¹⁴ illustrated a simple relationship between frequency and range averaging for broadband sonar. Consider a Gaussian frequency average defined as

$$I_f = \frac{\int \Psi(f, r_0) \exp(-(f - f_0)^2 / (\kappa f_0)^2) df}{\int \exp(-(f - f_0)^2 / (\kappa f_0)^2) df}, \quad (13)$$

where f_0 denotes the center frequency of the average, κ is the fractional bandwidth, and Ψ is the acoustic field which is explicitly given as a function of frequency and range. A range average can also be defined as

$$I_r = \frac{\int \Psi(f_0, r) \exp(-(r - r_0)^2 / (\kappa r_0)^2) dr}{\int \exp(-(r - r_0)^2 / (\kappa r_0)^2) dr}, \quad (14)$$

where r_0 is the center range and κ is the same fraction as above. Equation (14) uses a sliding Gaussian window of width proportional to the center range. Harrison and Harrison¹⁴ showed for Lloyd's mirror (i.e., a point source in a semiinfinite half-space with perfectly reflecting interface) that the averages in Eqs. (13) and (14) are identical. In other examples they illustrated that the two are very similar. Depending on the choice of κ , the Gaussian frequency window can vary from monochromatic to broadband. The cross in Fig. 5 indicates a fractional bandwidth of $\kappa=5$ for a center of $f=1000$ Hz and $r=250$ m. Qualitative examination of the features along either lines suggests a very similar averaging effect for this example.

Figure 6 compares frequency and range averages for a simulated experiment. The seabed consists of six sediment layers over a sediment halfspace (Fig. 2); data were generated as time series (synthetic seismoacoustic traces from

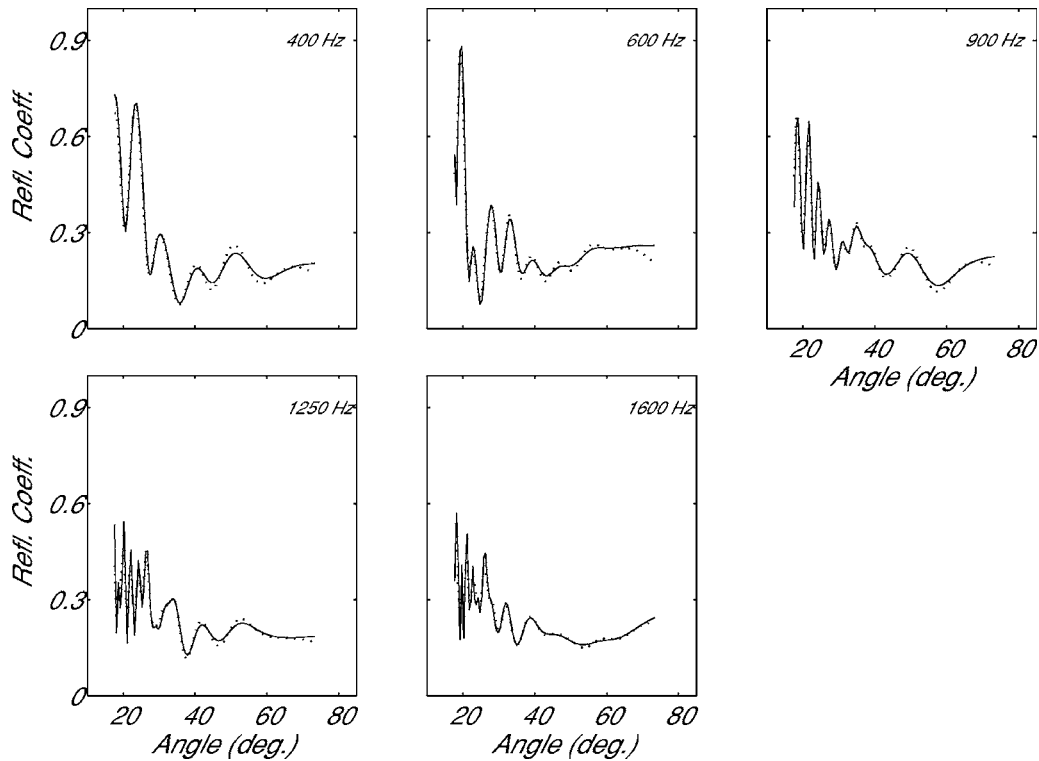


FIG. 6. Comparison of frequency (dotted line) and range (solid line) averages for simulated data derived from synthetic seismic traces for the seabed model in Fig. 2.

Green's functions convolved with the source pulse in Fig. 2) and then processed to resemble reflection coefficients. The agreement is excellent for all frequencies and angles. Other simulations showed that the agreement degrades for frequencies much below those shown in Fig. 6. However, for the frequency band of interest in this work, the range average reduces computation time by an order of magnitude at no significant loss in accuracy. Further, the computational expense for lower frequencies is considerably lower, making it less important to replace frequency averaging with range averaging for low frequencies.

B. Plane wave reflection coefficient forward model

The forward models used in this paper approximate the seabed as a layered lossy fluid. This is justified, since shear velocities in fine grained sediments are low,²³ and earlier inversion studies¹⁰ that treated the seabed as a halfspace showed that the reflection coefficient is relatively insensitive to shear properties and that ignored shear properties do not result in a bias for the other physical properties.

As indicated in Fig. 1, the experiment uses a point source generating spherical waves, and, therefore, the measured reflectivity contains spherical wave effects. Depending on the geometry of the experiment and the environment at the measurement site, these effects may be negligible, or may be important to take into account to recover meaningful estimates of the geoacoustic parameters.

For shallow water, low sound velocities, small penetration depth of interest, and certain frequency ranges, a plane wave assumption can be reasonable.¹¹ In that case, the forward model consists of a well established recursive calcula-

tion (see Chapter 1.6 in Ref. 24), referred to here as the plane wave forward model. For cases where the plane wave assumption is not sufficient, the following sections develop a full wave-field forward model.

C. Reflection of spherical waves from interfaces

For finite distance between source and interface or receiver (i.e., distances less than many wavelengths), full wave-field (spherical wave) effects must be taken into account^{5,6} to completely describe the reflection problem. The classic work on reflection of spherical waves is by Sommerfeld;²⁵ this section follows the derivations in Brekhovskikh and Godin.²⁶

For two homogeneous media, the total acoustic field of a spherical wave at a point that includes both direct and reflected components can be written as

$$\Psi = \frac{\exp(ikR)}{R} + \Psi_{\text{ref}}, \quad (15)$$

where Ψ_{ref} is the reflected wave, k is the wave number, and R is the distance from the source. Ψ_{ref} can be written as a superposition of reflected plane waves

$$\Psi_{\text{ref}} = \frac{ik}{2\pi} \int_0^{\pi/2-i\infty} \int_0^{2\pi} \exp(ik(x \cos \theta \cos \phi + y \cos \theta \sin \phi + (z+z_0)\sin \theta)) V(\theta) \cos \theta d\theta d\phi, \quad (16)$$

where x , y , and z are spatial coordinates, $z+z_0$ is the source-receiver vertical separation with respect to the interface, ϕ is

the azimuth, and $V(\theta)$ is the plane wave reflection coefficient.

It can be shown that the integral over ϕ in Eq. (16) reduces to a Bessel function of zeroth order.²⁶ By rearranging the integral limits in Eq. (16), the Bessel function can be expressed in terms of Hankel functions

$$\Psi_{\text{ref}} = \frac{ik}{2} \int_{-\pi/2+i\infty}^{\pi/2-i\infty} H_0^{(1)}(u) \times \exp(ik(z+z_0)\sin\theta)V(\theta)\cos\theta d\theta, \quad (17)$$

where $u=kr \cos\theta$. Equation (17) can be used to calculate the up-going spherical wave and thus the up-going energy (including reflected and lateral waves) for an arbitrary set of layers, given the plane wave reflection coefficient $V(\theta)$. The spherical reflection coefficient V_s can then be defined²⁷ as to include all effects but spherical spreading

$$\begin{aligned} \Psi_{\text{ref}} &= V_s \frac{\exp(ikR_1)}{R_1} \\ &= \frac{ik}{2} \int_{-\pi/2+i\infty}^{\pi/2-i\infty} H_0^{(1)}(u) \exp(ik(z+z_0)\sin\theta)V(\theta)\cos\theta d\theta. \end{aligned} \quad (18)$$

A method to yield V_s [Eq. (18)], based on calculating the total acoustic field is discussed in Sec. IV D.

For finite frequencies, the reflection is not from a single point but rather from a finite volume (a Fresnel volume) around the specular point. In essence, a hydrophone registers energy from a range of angles for a planar reflector. Spherical wave effects are important when the Fresnel zone spans an angular range over which the reflection coefficient changes significantly. For instance, the reflection coefficient can change rapidly with angle, particularly around the critical angle. The steepest descent solution²⁶ for the reflected wave is given by

$$\Psi_{\text{ref}} = \frac{\exp(ikR)}{R} \left(V(\theta_0) - \frac{i}{2kR} (V''(\theta_0) + V'(\theta_0)\tan(\theta_0)) \right), \quad (19)$$

where V' and V'' are the first and second derivative of the plane wave reflection coefficient with respect to angle and θ_0 is the specular angle. The dependence of the reflected field on the derivatives of the reflection coefficient with respect to angle illustrates the importance of full wave-field effects when the reflection coefficient changes rapidly with angle. Hence, a plane wave solution is expected to work well when the total sediment thickness is much smaller than the water depth and when the reflection coefficient changes slowly with angle, but may break down for increased depth of interest and cases involving rapid changes with angle.

The breakdown of the plane-wave approximation for plane reflectors at large depth below the water sediment interface can also be illustrated using ray theory. In the plane wave case, every sediment interface contributes to the reflection coefficient at the same angle. However, at finite distance from the source, as illustrated in Fig. 7, the energy for each sediment interface originates at a different angle. This effect

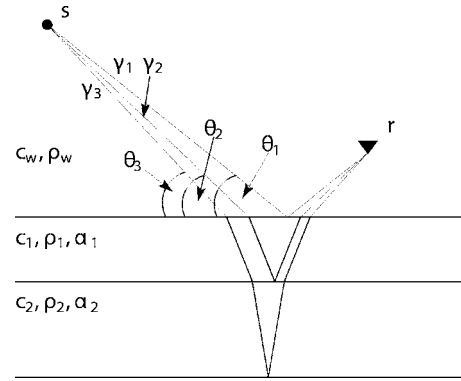


FIG. 7. Reflection of a spherical wave (represented by rays γ_1 , γ_2 , and γ_3) from multiple interfaces (represented by three layers with sound velocities c_i , densities ρ_i , and attenuations α_i); the point r registers energy contributions from different angles θ_i .

becomes more significant with larger reflection depth and means that spherical wave effects are more important for deeper layers in an environmental model.

To illustrate the above points, Fig. 8 shows a simulation of plane wave and spherical wave reflections for the case of a single interface between two halfspaces. The simulation was carried out at a frequency of 500 Hz for two different environmental models that differ only in the sound velocity of the lower halfspace. The source-receiver vertical separation was 178 m and one seabed model was given by $c_w = 1511$ m/s, $\rho_w = 1.029$ g/cm³, $\alpha_w = 0$ dB/ λ , $c_1 = 1700$ m/s, $\rho_1 = 1.4$ g/cm³, and $\alpha_1 = 0.06$ dB/ λ . For the second seabed model, c_1 was set to 1670 m/s. It can be seen that the greatest differences occur near the critical angle (25°) and that V_s is consistently lower than V around the critical angle. As Stickler⁵ has shown, the amount of displacement of the critical angle depends on the source receiver height above the interface as well as on the frequency. With increasing frequency or source-receiver vertical separation, the displacement of the critical angle decreases. Considering that the Fresnel zone is not symmetric around the specular point, this results in an effective critical angle that is shifted towards lower angles. The shift toward lower angles will generally

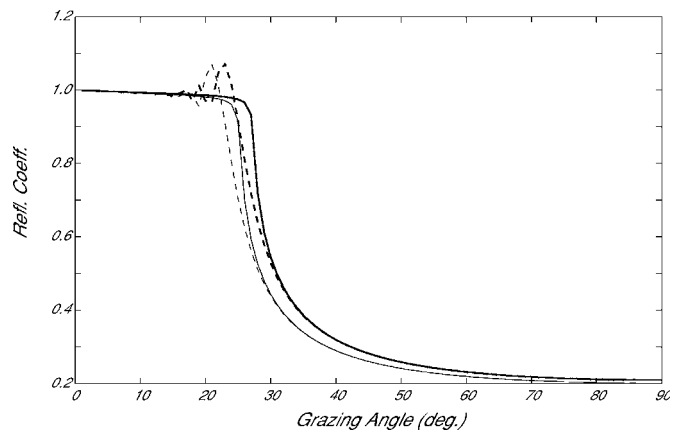


FIG. 8. Comparison of plane wave (V) and spherical wave (V_s) reflection coefficients (solid and dashed lines respectively) for an interface between two half-spaces at 500 Hz. Two different sets of parameters are considered (see the text). They only differ in the sound speed specified for the lower halfspace, 1670 m/s (thin lines) and 1700 m/s (thick lines).

result in a negative bias of sound velocities in inversions carried under the plane wave assumption. The phenomenon of reflection coefficient values greater than one, evident in Fig. 8, results from lateral waves.

For a rigorous examination of the reflection phenomena, the intuitive effects described above are not sufficient, and contributions from complex angles or inhomogeneous waves must be considered as well. All of these effects are included in the spherical wave reflection coefficient forward model described in the following section.

D. Spherical wave reflectivity forward model

For reflection coefficients that change rapidly with angle and for complex environments (many layers, many wavelengths depth below the seafloor), spherical wave effects are often significant. Harrison and Nielsen^{27,28} showed that plane wave and spherical wave reflection coefficients can be significantly different even for large source-receiver vertical separation (e.g., 210 m source-receiver vertical separation at 500 Hz). In this paper, a forward model is developed that derives a reflection coefficient, including spherical wave effects, from complex acoustic fields. The full acoustic fields are computed by wavenumber integration²⁹ and then processed using ray tracing to resemble reflection coefficients that take full spherical wave effects into account. Several numerical optimization techniques are applied to run the computationally intensive forward model as efficiently as possible.

Calculation of the full acoustic field Ψ is based on wave number integration (OASES⁷). For single-bounce reflection problems, the field can be split into two parts, a direct wave and a seabed response term

$$\Psi = \frac{\exp(i \int_{R_1} k(u) du)}{R_1} + V_s \frac{\exp(i \int_{R_2} k(u) du)}{R_2}, \quad (20)$$

where k is the wave number and R_1 and R_2 are the path length of the direct and bottom bounce paths, respectively. V_s is a measure for the seabed response which can be recovered by rearranging Eq. (20) to yield

$$V_s = \frac{\Psi - \exp\left(i \int_{R_1} k(u) du\right)/R_1}{\exp\left(i \int_{R_2} k(u) du\right)/R_2}. \quad (21)$$

The path lengths R_1 and R_2 and wave number integrals along the rays are determined using ray tracing in the water column, employing a depth dependent sound velocity profile. Hence, V_s quantifies the seabed response by correcting the field Ψ for spreading loss of direct and bottom reflected paths. All spherical wave effects (including possible lateral and inhomogeneous waves) are contained in V_s and will be accounted for in the inversion. It should be noted that Eq. (20) is a convenient way to process the acoustic field replica data to resemble the processing of the measured data, and is therefore suitable for the inversion. For convenience, V_s is referred to as the spherical reflection coefficient, although

TABLE I. Prior bounds for all layers.

Layer	h (m)	c (m/s)	ρ (g/cm ³)	α (dB/ λ)
1	0.5–3.5	1500–1600	1.2–1.6	0.01–0.10
2	1.0–4.0	1500–1600	1.3–1.8	0.05–0.15
3	0.5–3.5	1500–1600	1.3–1.8	0.10–0.20
4	6.0–9.0	1500–1600	1.4–2.0	0.05–0.15
5	0.5–3.5	1550–1650	1.6–2.2	0.05–0.15
6	2.0–6.0	1550–1650	1.8–2.2	0.01–0.15
7		1600–1700	1.8–2.4	0.00–0.05

more than just reflection effects enter the coefficient.

As reflection coefficients show a strong frequency dependence, frequency averaging must be applied to be consistent with the measured data. Here, the frequency average is replaced by an equivalent range average, as outlined in Sec. IV A. With the shortest range being in the order of 10^1 m, the wave number integration model is run with full Hankel transforms, eliminating any far field approximation. This ensures precise modeling of the acoustic field even at small ranges, at the price of higher computational effort.

This forward model is powerful and general and can calculate the spherical wave reflection coefficient for arbitrary sound velocity profiles in the water and arbitrary layering in the seabed. It can also account for interface roughness, and shear and gradient layers can be built into the seabed model parameterization, although these features are not used here.

V. INVERSION RESULTS

A. Plane wave inversion

To compare the performance of the plane wave forward model to the full wave-field forward model, both forward models are applied in inversion. First, the Bayesian geoaoustic inversion with the plane wave forward model is applied to the reflectivity data extracted from the synthetic seismograms (Figs. 3 and 4). The prior bounds, given in Table I, are chosen to be representative of the information that would

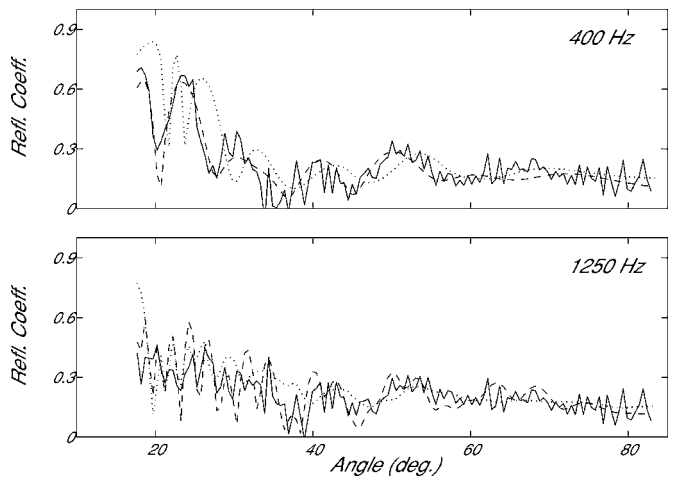


FIG. 9. Fit of the MAP replica data (solid line) to the simulated experimental data (dotted line) for plane-wave inversion. The dashed line shows data generated with the plane wave forward model for the true parameters.

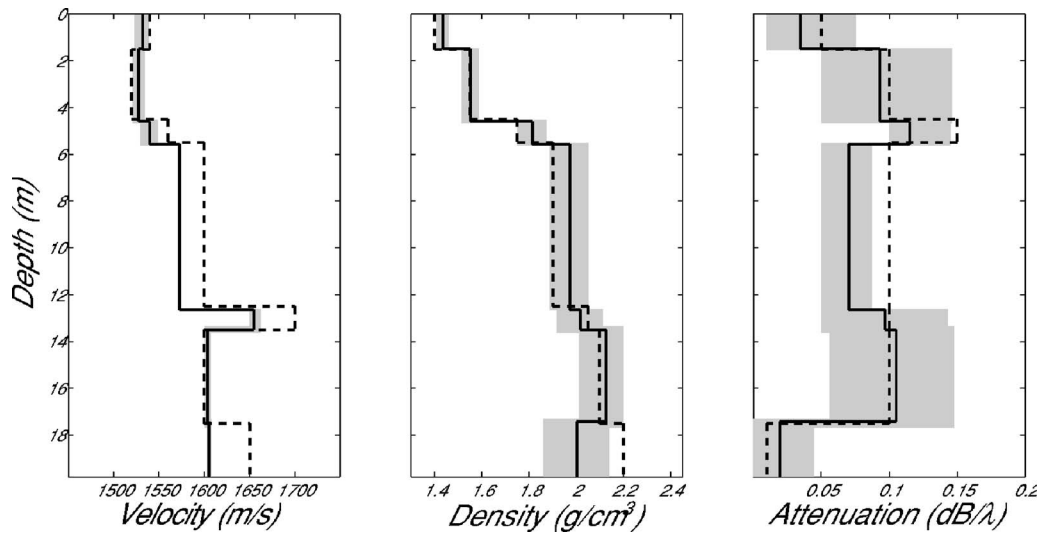


FIG. 10. Bayesian inversion results using the plane wave forward model. The solid line indicates the *a posteriori* mean seabed model; shaded areas are the 95% HPD credibility intervals. The dashed line represents the true seabed model.

typically be available, and are wide enough so that the data rather than the prior primarily determine the inversion results. The data are first inverted using ASSA under the assumption of unknown error magnitudes at each frequency to minimize misfit function Eq. (11). The resulting ML seabed model was used to find a ML estimate of the error standard deviation at each frequency according to Eq. (10). These standard deviations were then used in Eq. (12) and FGS was applied to sample the PPD as a function of the 27 unknown parameters. Figure 9 shows that the data fit of the MAP seabed model is reasonably good (for clarity, two out of five frequencies are shown; the results are representative of all frequencies). Figure 9 also shows data that were generated with the plane wave forward model using the true geoaoustic parameters. There is a noticeable difference between the simulated experiment data and plane wave data for the true seabed model, especially at low frequencies.

Figure 10 shows the plane-wave inversion results in terms of the mean seabed model and 95% HPD credibility intervals (see Sec. III A). It can be seen that, although velocities and densities of the upper three layers are reasonably well determined, deeper layers cannot be recovered with the plane wave forward model. The velocity profile shows strong negative biases at depth, as expected from the analysis in

Sec. IV D, and does not represent the true profile well. Biases also occur for the density profile; however, the 95% HPD credibility intervals are wide enough to include the true seabed model with the exception of the deepest layer. Further, some layer thicknesses show significant discrepancies from the true seabed model and the 95% HPD credibility intervals do not include the true seabed model (see also Table II). The credibility intervals for the attenuation profile generally extend over much of the search bounds (Table I), indicating little ability to resolve this parameter. Numerical values for the inversion results are given in Table II.

The inversion results in Fig. 10 show that the plane wave forward model is not sufficient to recover the complex structure in this case. Even though the data can be fit reasonably well, this fit is obtained by introducing a substantial bias in the parameter estimates to account for the theory error. Not only is the data fit misleading about the quality of the results, but the HPD credibility intervals also provide false confidence. In this case, a full wave-field forward model must be considered.

B. Spherical wave inversion

This section applies the Bayesian inversion using the full wave-field forward model to the reflection data derived

TABLE II. Numerical values for true parameters and inversion results (True credibility intervals) for the plane wave inversion.

Layer true	h (m)		c (m/s)		ρ (g/cm ³)		α (dB/ λ)	
	True	True	True	True	True	True	True	True
1	1.5	1.44–1.53	1540	1523–1538	1.40	1.41–1.46	0.05	0.01–0.08
2	3.0	3.05–3.16	1520	1521–1534	1.55	1.51–1.59	0.10	0.05–0.14
3	1.0	0.90–1.02	1560	1529–1548	1.75	1.75–1.87	0.15	0.01–0.14
4	7.0	7.03–7.07	1600	1571–1573	1.90	1.88–2.05	0.10	0.05–0.08
5	1.0	0.69–0.99	1700	1650–1662	2.05	1.91–2.11	0.10	0.05–0.14
6	4.0	3.82–4.19	1600	1600–1607	2.10	2.01–2.19	0.10	0.06–0.14
7			1650	1603–1607	2.20	1.86–2.14	0.01	0.00–0.04

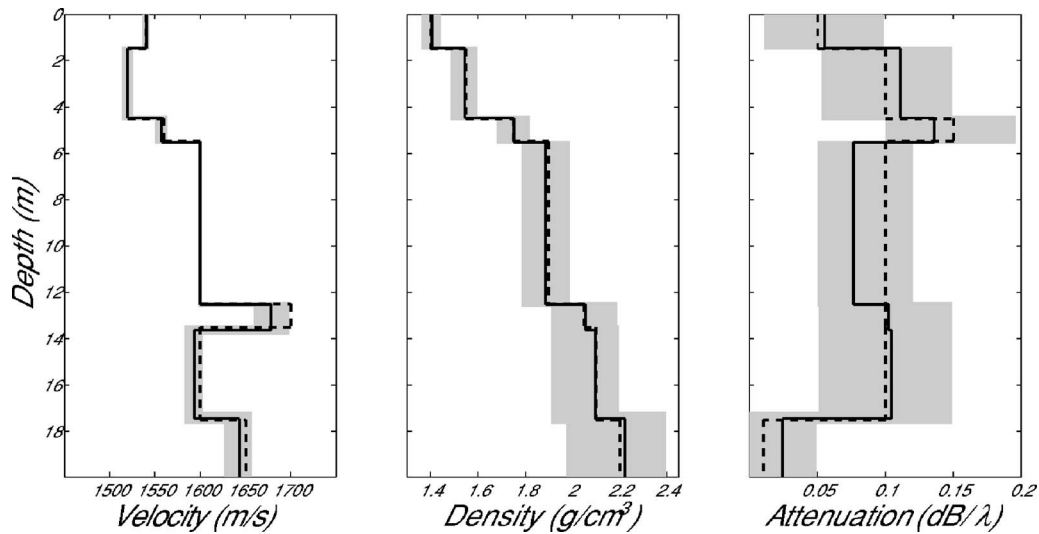


FIG. 11. Bayesian inversion results for the spherical wave forward model. The solid line indicates the *a posteriori* mean seabed model; shaded areas are the 95% HPD intervals. The dashed line represents the true seabed model. The inversion results and their associated uncertainties agree well with the true seabed model.

from seismo-acoustic traces (Figs. 3 and 4). As before, ASSA optimization is used to obtain a maximum likelihood estimate of the error standard deviations at each frequency, which are then used in FGS to sample the PPD. The fit to the data is shown in Fig. 4. Figure 11 shows the sound-velocity, density, and attenuation profiles and their associated uncertainties in terms of 95% HPD credibility intervals. The *a posteriori* mean seabed model is very close to the true seabed model for both the velocity and density profiles. Layer thicknesses are also matched well. The velocity tends to be resolved better (smaller uncertainty) in thicker layers. In general, the uncertainties in the parameter estimates grow with depth, particularly for the density profile. Attenuation is not resolved particularly well, with credibility intervals covering much of the prior bounds given in Table I. The best resolution for attenuation was obtained in the thickest layers (see Table III). Comparing the full wave-field results in Fig. 11 to the plane wave approximation results in Fig. 10 indicates that a substantial improvement is obtained using the full wave-field forward model.

VI. DISCUSSION AND SUMMARY

This paper developed a full wave-field Bayesian inversion for reflection coefficient data derived from a single

bounce experiment. The inversion was applied to data from a simulated experiment, and results were compared to those obtained with an inversion using the standard plane wave approximation. The full wave-field inversion algorithm is based on a forward model that use complex acoustic fields, obtained through wavenumber integration, to calculate the replica reflection coefficients. Hence, the replica reflection coefficients contain full wave-field effects which are not included in a plane wave approximation. Several techniques were applied to carry out the computational intensive inversions efficiently, including replacing frequency with range averaging and parallel processing.

The data considered here were generated by computing synthetic seismograms for a seabed model including six layers and a halfspace. Green's functions were computed for the seabed model and convolved with a source pulse from measured data. The synthetic seismograms were then processed to yield reflection coefficients as a function of angle and frequency, with Gaussian noise added. Both inversions (plane wave and full wave-field forward models) were applied to this data set. The presented results depend on experiment geometry and full wave-field effects increase with depth of a layer beneath the sea floor.

TABLE III. Numerical values for true parameters and inversion results (True credibility intervals) for the spherical wave inversion.

Layer true	h (m)		c (m/s)		ρ (g/cm ³)		α (dB/ λ)	
	True	True	True	True	True	True	True	True
1	1.5	1.46–1.51	1540	1539–1543	1.40	1.35–1.39	0.05	0.02–0.10
2	3.0	2.98–3.05	1520	1515–1520	1.55	1.51–1.57	0.10	0.05–0.12
3	1.0	0.95–1.02	1560	1552–1559	1.75	1.68–1.78	0.15	0.10–0.20
4	7.0	6.91–7.07	1600	1597–1599	1.90	1.79–1.96	0.10	0.08–0.13
5	1.0	0.86–1.12	1700	1684–1716	2.05	1.92–2.17	0.10	0.05–0.15
6	4.0	3.80–4.17	1600	1589–1602	2.10	1.94–2.19	0.10	0.05–0.15
7			1650	1635–1655	2.20	2.01–2.39	0.01	0.00–0.05

Comparing the inversion results for the two forward models indicated the plane wave assumption was insufficient for this case. The environmental model could not be resolved at depth using the plane wave forward model, resulting in biased estimates and misleading credibility intervals. The full wave-field inversion provided substantial improvements over the plane wave inversion. In particular, the sediment velocity profile was estimated to high accuracy, with smaller uncertainties associated with thicker layers. Density was also well resolved with reasonable credibility intervals. Uncertainties for velocity and density generally increase with depth although smaller velocity uncertainties are associated with thicker layers.

ACKNOWLEDGMENTS

The first author was supported by the Natural Sciences and Engineering Research Council of Canada and the University of Victoria during Ph.D. studies. The authors gratefully acknowledge the support of the Office of Naval Research, in particular the Office of Naval Research postdoctoral fellowship.

- ¹O. F. Hastrup, "Digital analysis of acoustic reflectivity in the Tyrrhenian Abyssal Plain," *J. Acoust. Soc. Am.* **42**, 181–190 (1970).
- ²R. E. Christensen, J. A. Frank, and W. H. Geddes, "Low-frequency propagation via shallow refracted paths through deep ocean unconsolidated sediments," *J. Acoust. Soc. Am.* **57**, 1421–1426 (1975).
- ³N. R. Chapman, "Modelling ocean-bottom reflection loss measurements with the plane-wave reflection coefficient," *J. Acoust. Soc. Am.* **73**, 1601–1607 (1983).
- ⁴C. W. Holland and J. Osler, "High-resolution geoacoustic inversion in shallow water: A joint time- and frequency-domain technique," *J. Acoust. Soc. Am.* **107**, 1263–1279 (2000).
- ⁵D. C. Stickler, "Negative bottom loss, critical-angle shift, and the interpretation of the bottom reflection coefficient," *J. Acoust. Soc. Am.* **61**, 707–710 (1977).
- ⁶F. B. Jensen and H. Schmidt, "Evaluation of experimental techniques for determining the plane wave reflection coefficient at the sea floor," in *Ocean Seismo-Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York, 1987), pp. 721–730.
- ⁷H. Schmidt, "OASES version 2.2: User guide and reference manual," Technical Report, Massachusetts Institute of Technology, Boston, 1999.
- ⁸S. E. Dosso, M. J. Wilmut, and A.-L. S. Lapinski, "An adaptive-hybrid algorithm for geoacoustic inversion," *IEEE J. Ocean. Eng.* **26**, 324–336 (2001).
- ⁹S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," *J. Acoust. Soc. Am.* **111**, 129–142 (2002).

- ¹⁰S. E. Dosso and C. W. Holland, "Geoacoustic uncertainties from visco-elastic inversion of seabed reflection data," *IEEE J. Ocean. Eng.* **31**, 657–671 (2006).
- ¹¹C. W. Holland, J. Dettmer, and S. E. Dosso, "Remote sensing of sediment density and velocity gradients in the transition layer," *J. Acoust. Soc. Am.* **118**, 163–177 (2005).
- ¹²J. Dettmer, "Geoacoustic reflectivity inversion: A Bayesian approach," Ph.D. thesis, School of Earth and Ocean Sciences, University of Victoria, Victoria, 2006.
- ¹³C. W. Holland, "Seabed reflection measurement uncertainty," *J. Acoust. Soc. Am.* **114**, 1861–1873 (2003).
- ¹⁴C. H. Harrison and J. A. Harrison, "A simple relationship between frequency and range averages for broadband sonar," *J. Acoust. Soc. Am.* **97**, 1314–1317 (1995).
- ¹⁵A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation* (Elsevier, Amsterdam, 1987).
- ¹⁶K. Mosegaard and M. Sambridge, "Monte Carlo analysis of inverse problems," *Inverse Probl.* **18**, R29–R54 (2002).
- ¹⁷D. H. Rothman, "Automatic estimation of large residual statistics corrections," *Geophysics* **51**, 337–346 (1986).
- ¹⁸S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721–741 (1984).
- ¹⁹*Markov Chain Monte Carlo In Practice, Interdisciplinary Statistics*, edited by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Chapman & Hall/CRC, Boca Raton, FL, 1996).
- ²⁰W. Gropp, E. Lusk, and A. Skjellum, *Using MPI, Portable Parallel Programming with the Message-Passing Interface* (MIT Press, Cambridge, 1999).
- ²¹J. Dettmer, S. E. Dosso, and C. W. Holland, "Uncertainty estimation in seismo-acoustic reflection travel-time inversion," *J. Acoust. Soc. Am.* **122**, 161–176 (2007).
- ²²S. E. Dosso, P. L. Nielsen, and M. J. Wilmut, "Data error covariance in matched-field geoacoustic inversion," *J. Acoust. Soc. Am.* **119**, 208–219 (2006).
- ²³M. D. Richardson, *Shallow Water Acoustics* (China Ocean Press, Beijing, 1997).
- ²⁴F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, Series in Modern Acoustic and Signal Processing (AIP Press, New York, 1993).
- ²⁵A. Sommerfeld, *Partial Differential Equations in Physics* (Academic, New York, 1949).
- ²⁶L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media II: Point Source and Bounded Beams*, Wave Phenomena, 1st ed. (Springer, New York, 1992).
- ²⁷C. H. Harrison and P. Nielsen, "Plane wave reflection coefficient from near field measurements," *J. Acoust. Soc. Am.* **116**, 1355–1361 (2004).
- ²⁸C. H. Harrison and P. Nielsen, "Plane wave reflection coefficient derived from spherical wave measurements," *Proceedings of the Seventh European Conference on Underwater Acoustics*, 2004, pp. 3–8.
- ²⁹H. Schmidt, "SAFARI: Seismo-acoustic fast field algorithm for range-independent environments," Technical Report SR-113, SACLANT ASW Research Centre, La Spezia, Italy, 1988.

Subbottom profiling using a ship towed line array and geoacoustic inversion

T. C. Yang,^{a)} Kwang Yoo, and L. T. Fialkowski
Naval Research Laboratory, 4555 Overlook Ave., Washington, D.C. 20375

(Received 21 May 2007; revised 4 September 2007; accepted 18 September 2007)

Bottom profiling traditionally uses broadband signals received on a line array at long ranges to estimate the bottom layer structure and thickness. In this paper, a subbottom profiling method is developed and applied to a ship-towed line array using the same ship towed source to estimate the subbottom layer structure and thickness. A ship towed line-array system can be used to estimate bottom properties using geoacoustic inversion and can cover a wide area in a short time. It needs some prior information about the subbottom structure and layer thickness, without which the solution can be ambiguous and even erratic when resolving parameters over a wide area. It is shown that the required subbottom information can be obtained from the time-angle relation by beamforming the *same* acoustic signal data used for geoacoustic inversion. The time-angle analysis is used to expose the prevalent physics intrinsic to geoacoustic inversion. One finds that the tau- p relation of the bottom and the bottom reflection coefficients, sampled at discrete angles associated with bottom and multiple surface-bottom returns, are often adequate, for many practical applications, to uniquely determine the geoacoustic bottom at low (≤ 1 kHz) frequencies.
[DOI: 10.1121/1.2799495]

PACS number(s): 43.30.Pc, 43.60.Pt, 43.30.Wi [AIT]

Pages: 3338–3352

I. INTRODUCTION

In shallow water, acoustic signal propagation is significantly affected by the bottom as acoustic wave interaction with the bottom is unavoidable. Bottom properties become a critical component in the modeling and prediction of transmission loss (TL) in shallow water, making a data base of the bottom parameters highly desirable. There are various ways to collect bottom property information, ranging from the estimation of sound speed near the surface from core samples to determination of deep layer structure and sound speed based on travel time data from seismic reflection/refraction. The latter measurement requires a dedicated ship with specialized equipment including a powerful source (e.g., air-gun, boomer) and a multichannel streamer (a line array deployed near the surface).

Recently, there has been great interest in assessing the bottom properties using existing sonar arrays. This through-the-sensor approach, if successful, could provide a capability for rapid environment (bottom) assessment (REA) and is very attractive from the operational perspective as the REA result could be used for real time performance assessment. For long-range propagation in shallow water, one notes that sound interaction with the bottom is limited mostly to less than, say, 50 m in depth, as the deep penetrating acoustic rays are quickly attenuated with increasing range. It has been shown recently that the subbottom parameters can be inverted from acoustic data received on a vertical or horizontal array.¹ This method, referred to as geoacoustic inversion, uses the signal amplitude and phase over a band of frequencies. Seismic reflection/refraction normally uses only the signal amplitude as a function of time.

The motivation for geoacoustic inversion lies in the (potential) ability to estimate bottom properties from the received acoustic data without requiring specialized equipment as used in the seismic community. The method has, however, some limitations. A “critical” problem is that the solution is often not unique and not repeatable (between different data snapshots). Several factors contribute to the non-uniqueness. One problem is the fluctuation of the acoustic signal between different snapshots of data. Siderius *et al.*² found that geoacoustic inversion results from data on a vertical array at a source-receiver range of 5–10 km had high fluctuations due to the ocean random media which are not included in the propagation model. The results at 2 km are much more stable as the random media effect on signal propagation is limited, so that a range independent geoacoustic inversion method can be applied.

Another factor contributing to the nonuniqueness of inversion solutions is the lack of information about the (sub-) bottom layer structure and layer thickness. In the absence of such information, one is forced to make some assumptions; the results are often influenced by the assumptions made. It is natural to ask, “Is it possible to determine the subbottom structure from the acoustic data?” To this end, Battle *et al.*³ calculated the probability distribution for different sediment layers that best match the data. As is well known, the Bayesian approach requires a prior probability distribution on the number of layers. They found that, if a uniform distribution is assumed, the probability can be skewed toward a large number of layers. Without using the Bayesian approach, one could also (by brute force) search for the number of sediment layers in the geoacoustic inversion. Such an approach is found impractical as it requires intensive computations, and

^{a)}Electronic-mail: yang@wave.nrl.navy.mil

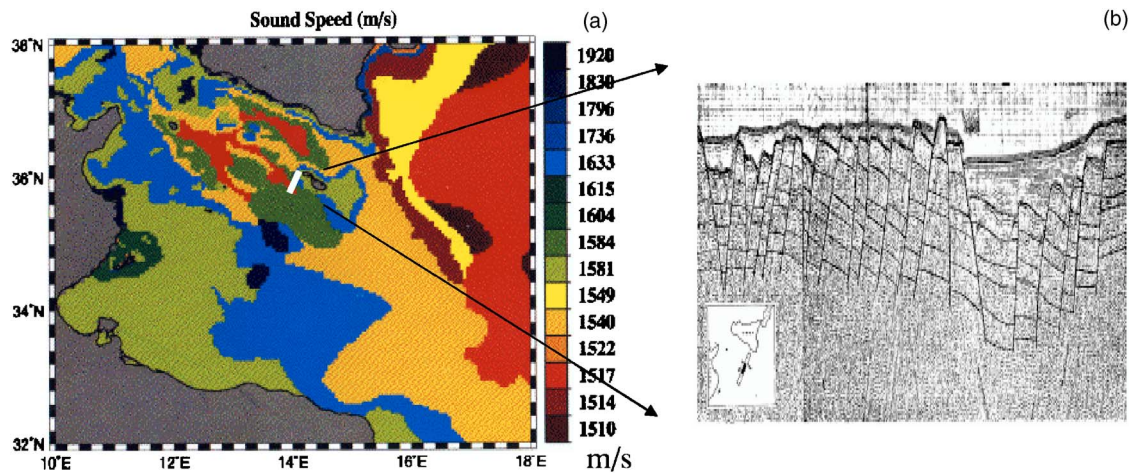


FIG. 1. (Color online) (a) Surface sound speed and (b) vertical layer structure.

also not useful as multiple solutions are probable. In practice, one assumes one or two sediment layers over a hard bottom in most inversion applications.

It is clear that if the bottom consists of three layers (e.g., two layers of sediments over a hard bottom) and the inversion model assumes only two layers, the inversion result only yields, at best, an average property of the bottom. Conversely, if the bottom consists of two layers (one sediment layer over a hard bottom), and the inversion results assume multilayers, geoacoustic inversion should show that several layers are of the same kind.⁴ However, this is not guaranteed. Note also that if the sediment layer has a thickness of 30 m, but the model assumes a thickness of less than 20 m, the inversion method will yield an incorrect answer. Without prior knowledge of the subbottom layering structure, one has no way of knowing whether the answer is correct. On the other hand, if one has approximate knowledge of layer thicknesses, then the inversion method could yield a more accurate estimate of the sediment thicknesses, as well as other bottom parameters, which are coupled to the thicknesses.

In this paper, a subbottom profiling method is applied to a ship-towed line array to yield a crude estimate of the bottom sediment thickness. The method uses the same acoustic data which is used for geoacoustic inversion, and is based on the travel time and arrival angle information which is often ignored in geoacoustic inversion. The subbottom profiling result is intended for use as the prior knowledge for geoacoustic inversion on a towed sonar array. It should be noted that, although the bottom properties could be determined directly from the subprofiling approach (see the following), the analysis requires identification of acoustic arrivals by an expert and therefore cannot be easily automated as the geoacoustic inversion method can.

The subbottom profiling method determines the bottom reflection loss and bottom impulse response at angles associated with various bottom-bounced arrivals. The delay-time and angle relationship of the bottom impulse response is shown to be identical to the τ - p relationship used in seismic profiling.⁵ However, unlike seismic reflection/refraction measurement, which is designed so that the bottom-bounced arrivals are clearly identified in the time domain data, the

sonar array data contains multiple surface and bottom bounced arrivals which are mixed in the time domain. Consequently, a beam domain approach is proposed to determine the subbottom layer structure and properties.

The motivation for using a towed line array for REA is that shallow water environmental properties are very location dependent, with a scale of a few to tens of kilometers. Figure 1(a) shows the surface sound speed in the Mediterranean⁶ and Fig. 1(b) shows how the bottom structure varies substantially over approximately a 40 km range.⁷ One finds that the geographical provinces are small and many, and the bottom layer structure changes significantly with range. Often the bottom properties deduced from certain TL data in one area yield incorrect prediction of TL for an adjacent area. The error is due in part to the spatial variability of the bottom. Note that measuring bottom properties using (seismic) reflection/refraction is generally expensive and time consuming. A towed sonar array is mobile and capable of assessing the bottom properties over a wide area (in a reasonable amount of time) using through-the-sensor data (collected either during the run or from previous runs). Siderius *et al.*,⁸ Fallat *et al.*,⁹ Fialkowski *et al.*,¹⁰ and Battle *et al.*¹¹ have recently demonstrated the usefulness of geoacoustic inversion using a ship-towed line array. Siderius *et al.*⁸ inverted for the bottom parameters using a cost function summing coherently over frequencies and incoherently over receivers. The data are linear frequency modulated (LFM) signals transmitted from a towed source received on the horizontal line array (HLA). A ping provides a local measurement of the bottom. Fallat *et al.*⁹ applied this method to data over a range of 11 km. Fialkowski *et al.*¹⁰ used a cost function summing incoherently over frequencies but coherently over receivers. This method is insensitive to source-receiver range but more sensitive to array tilt and bow. Battle *et al.* used the ship noise of a trailing ship as the source for geoacoustic inversion¹¹; one could also use own ship noise as the source. It is noted that to achieve consistency and reliability of the geoacoustic inversion results as a function of range, some prior knowledge of the bottom layering and thickness is required.¹⁰ In these (previous) works, certain assumptions were made about the subbottom layer structure based on

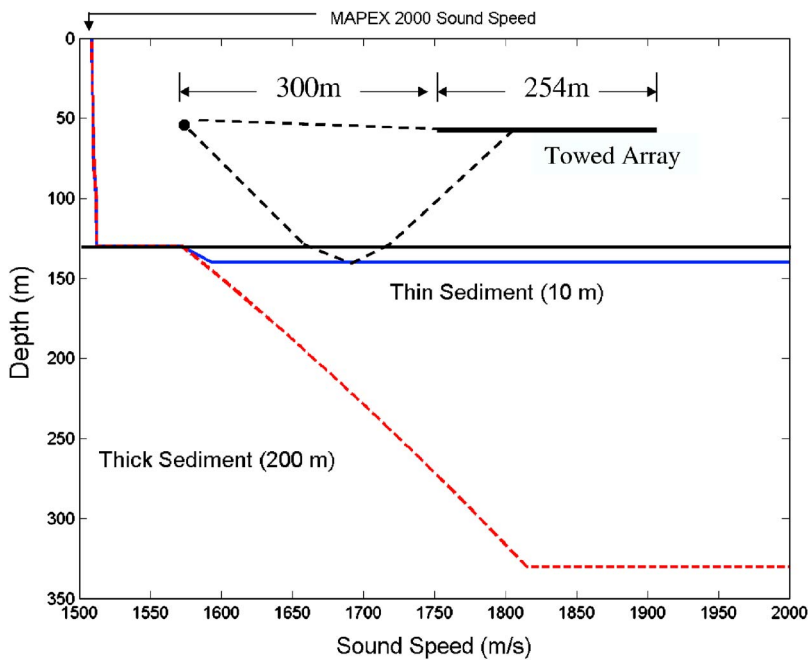


FIG. 2. (Color online) MAPEX2000 Speed, source receiver configuration and two hypothetical bottoms.

prior seismic surveys. This paper will show how the subbottom layer structure can be determined (directly) from the acoustic data using the sub-profiling method without requiring a prior seismic survey.

This paper is organized as follows. The basic properties of the bottom reflection coefficient, bottom impulse response function and the tau- p relation of the bottom are reviewed in Sec. II. The MAPEX2000 data is analyzed for bottom profiling in Sec. III. The physics behind bottom inversion is discussed in Sec. IV. In Sec. V, the bottom profiling method is applied to OREX05 data as another example. A short summary is given in Sec. VI.

II. BOTTOM IMPULSE RESPONSE AND TAU- p RELATION

One way to characterize sound interaction with the bottom is in terms of the sound reflection coefficients of the bottom. If the reflection coefficient is known as a function of frequency and incidence angle, then the bottom is well understood for sound propagation. It is generally accepted that no two bottoms can produce the same reflection coefficient. This principle will be used to guide the analysis of the geoaoustic inversion method.

To illustrate how the bottom reflection coefficients uniquely determine the sediment layer structure and thickness, consider two bottom types as shown in Fig. 2. One bottom has a sediment layer of 10 m thickness over a hard bottom and the other has a sediment layer of 200 m thickness over a hard bottom. The sound speed in the sediment layer increases with depth. The two sediments have approximately the same density (1.76 g/cm^3), but different attenuations (0.02 and 0.06 dB/m/kHz for the thin and thick sediments, respectively). The hard bottom is the same for both cases; it has a sound speed of 3200 m/s and is totally reflective for grazing angles below 60° . Given the bottom parameters, one can calculate the reflection coefficients as a function of frequency and incident (grazing) angle. The cal-

culated reflection loss (dB) is shown in Figs. 3(a) and 4(a) as a function of frequency and incident grazing angle, for the thin and thick sediments, respectively. Taking the Fourier transform of the reflection coefficients at a fixed angle, one obtains the bottom impulse response function as a function of time for that angle. The delay time at vertical incidence (grazing angle at 90°) is related to the layer thickness.¹² To illustrate this, the impulse response is shown in Fig. 3(b) as a function of the depth using the mean sound speed of 1582 m/s in the thin layer (depth=delay time \times 1582 m/s). At the normal incidence angle, it yields the correct layer thickness of 10 m for the thin sediment case. Using the same sound speed, the estimate of the sediment thickness as shown in Fig. 4(b) is somewhat shorter compared with the true layer thickness; the reason is that the mean sound speed in the thick layer is higher than 1582 m/s. Note that often the sediment sound speed is not known, therefore a nominal value is assumed for the sound speed, and the so estimated depth is called the effective depth. This rough estimate of the sediment thickness is found to be adequate for many applications.

Bottom structure and layer thickness are routinely surveyed by the seismic community and oil industry. It is worth reviewing their method to show its relationship to the reflectivity discussed previously. The seismic community uses the tau- p relation for seismic reflection/refraction data analysis. It is based on the acoustic ray, as illustrated in Fig. 5(a) for two layers. For sound propagating through n layers of constant sound speeds, tau is defined as⁵

$$\tau \equiv t - px = \sum_{i=1}^n q_i z_i, \quad (1)$$

where t is the ray travel time, x is the horizontal range along the ray path, and p is the horizontal slowness, i.e., the inverse of the sound speed times the cosine of the grazing angle. By Snell's law, p is the same in all layers. The z_i on the right-

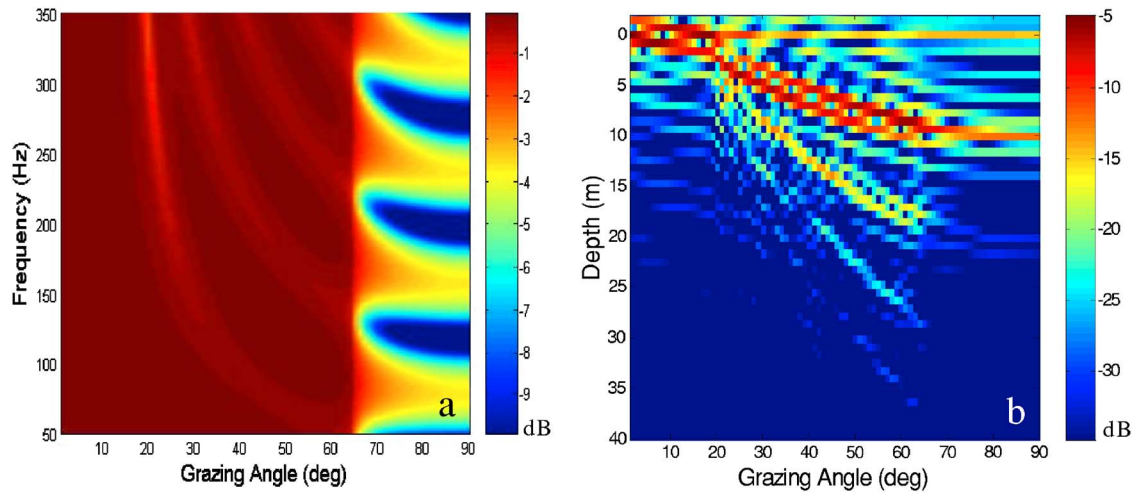


FIG. 3. (Color online) (a) Reflection loss vs angle for frequencies for the thin sediment bottom. (b) The corresponding impulse response function, depth = time \times 1582 m/s.

hand side is the depth span of the ray in the i th layer, and q_i is the vertical slowness in the i th layer. They are given by

$$u_i = 1/c_i, \quad p_i = u_i \cos \theta_i = p, \quad q_i = u_i \sin \theta_i, \quad (2)$$

for each layer i . Figure 5(a) shows the tau- p relation for a bottom bounced ray. For a ray completing a ray cycle, $\tau = 2\sum_{i=1}^n q_i z_i$, where z_i is the i th layer depth.

Tau (τ) is often called the reduced travel time. The advantage of using τ (instead of t) is that the tau- p relationship is independent of the horizontal range and thus can be applied to data at different ranges (with the same common middle point), or at long ranges where the acoustic ray has traveled many ray cycles. In seismic reflection/refraction, by stacking the ray arrival time series at different source-receiver ranges, one identifies different ray arrivals. From the ray travel time and arrival angle, one determines the tau- p relationship. By comparing the measured tau- p relation with the model predictions, one can determine the bottom structure in different areas.

The tau- p relation for a bottom reflected ray can be determined from the reflection coefficient or the bottom impulse response. Figure 5(b) shows the tau- p relation for a

three layer bottom, each with a 10 m layer thickness with a sound speed increasing with the layer number (depth). The top solid curve is $\tau(p)$ that for the first layer. As the slowness (p) decreases (when the grazing angle increases beyond the critical grazing angle of the second layer), the sound penetrates into the second layer and the $\tau(p)$ is taken over by that of the second layer, until the sound penetrates into the third layer and then the $\tau(p)$ is controlled by the third layer. The result is that the overall $\tau(p)$ is that of the envelope of the three curves.

Figure 5(c) plots the same bottom impulse response previously shown in Fig. 3(c) now in terms of the horizontal slowness using the relation between θ and p in Eq. (2). By overlaying the solid line from Fig. 5(b), one observes that the two plots are identical; tau is, in this case, the same as the delay time of the bottom impulse response. This indicates that the sediment layer information is contained in the $\tau(p)$ relation and can be deduced from the bottom reflection impulse response. Comparing Fig. 5(c) with Fig. 5(b), one notes that the bottom impulse response has sidelobes in the time domain which can be (easily) mistaken as $\tau(p)$ from

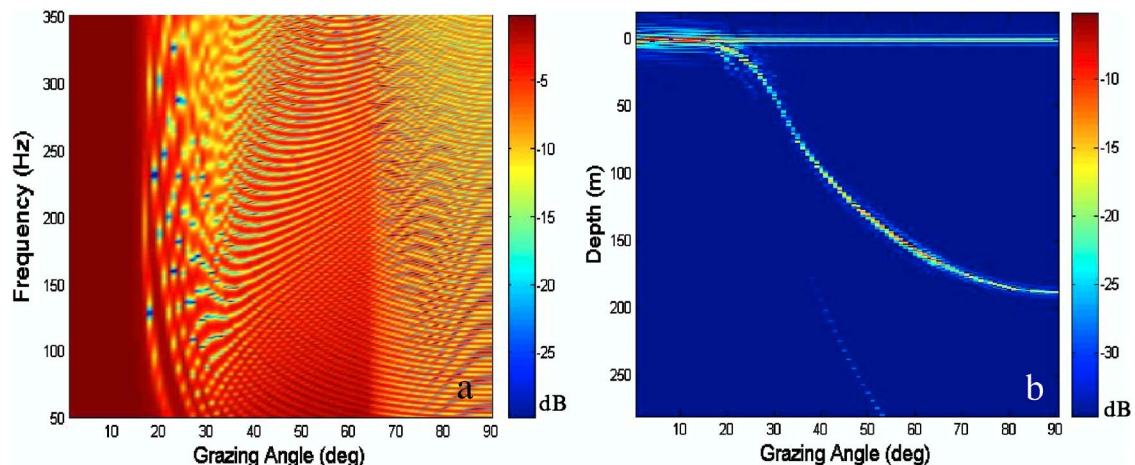


FIG. 4. (Color online) (a) Reflection loss vs angle for frequencies for the thick sediment bottom. (b) The corresponding impulse response function, depth = time \times 1582 m/s.

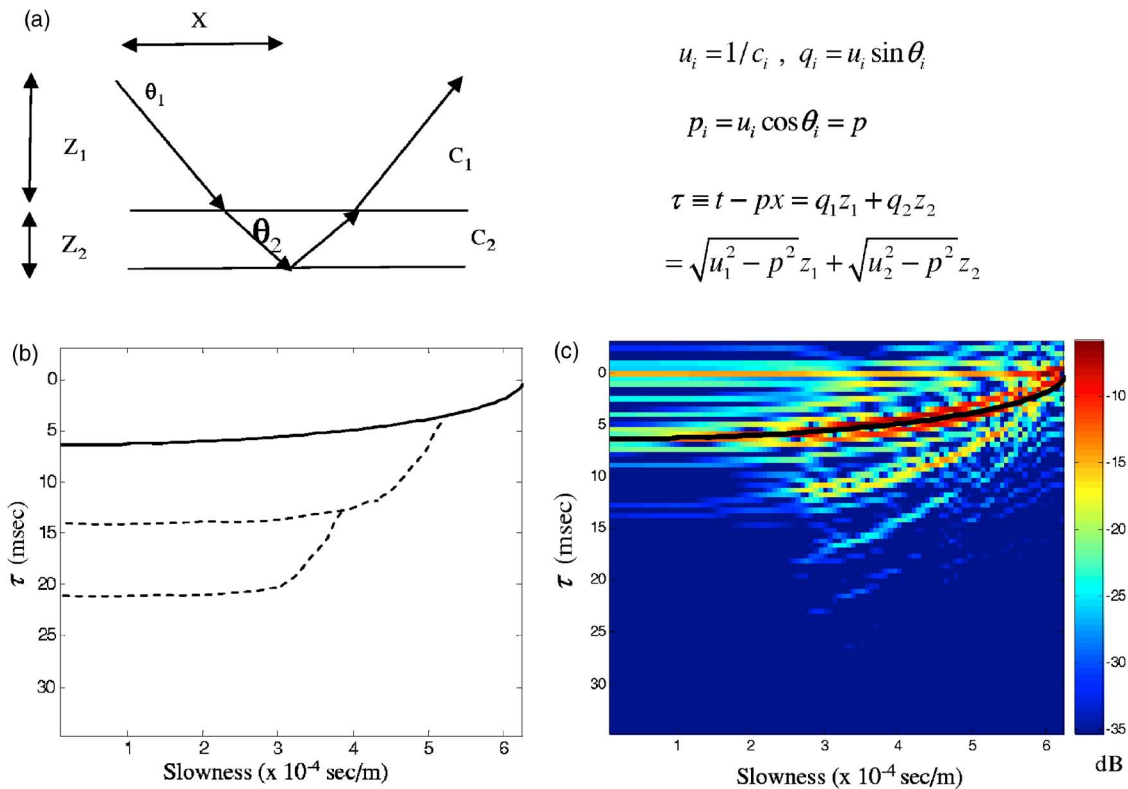


FIG. 5. (Color online) (a) Tau- p relation of the acoustic ray penetrating tow layers of sediment; x is the horizontal distance from the source/receiver to the common middle point. (b) A tau- p plot for three layers of sediment, each of 10 m thickness. The solid curve will be that for one layer. The envelope of the top two curves will be for two layers and the envelope of the three curves will be for three layers. (c) Display of the bottom impulse response shown in Fig. 3(b) in terms of slowness (p). The black solid curve is the same as the top curve of (b).

multiple layers. Although there are significant differences between Fig. 5(b) and 5(c), the sidelobes appearing in real data (see the following section) are not so obvious and need to be examined carefully to avoid incorrect conclusions.

Figure 5(c) shows that the bottom impulse response is uniquely characterized by its time dispersion (the tau- p relation) and its amplitude (the reflection loss). As no two different linear systems can have the same impulse response (the linear system theory), the bottom is uniquely determined if the corresponding tau- p relation and reflection loss are known. In practice, there are system limitations (small, discrete measurement points) and data uncertainties. How well can the bottom parameters be determined is the subject of the data analysis in the following sections.

It should be noted that the physical bottom may contain many fine layers with different densities and sound speed. At low frequencies, the acoustic response of the bottom can be characterized by a much simpler “acoustic bottom” with one or two layers with constant sound speeds. How well can the bottom layer structure be resolved with the towed array data will be studied with real data.

III. SUBBOTTOM PROFILING USING THE TOWED LINE ARRAY: THE MAPEX 2000 EXPERIMENT

The MAPEX 2000 experiment was carried out by NURC in the Mediterranean.⁸ The data analyzed here were taken over a track north of the Malta Plateau as shown by the line in Fig. 6, extending from site 4 to site 1, covering approximately 24 km. An acoustic projector and a HLA were

towed behind the ship with a separation of 300 m between them. The source and the HLA were both at a depth of approximately 60 m. The source broadcasted a LFM signal approximately every minute, covering a frequency band of 150–850 Hz. The receiver array has 128 phones spaced at 2 m. During this run, more than 100 LFM pings were transmitted and collected on the towed line array. The data were previously analyzed for geoacoustic inversion.^{8–10} A powerful vertical seismic profiler, called the uniboomer, was also used over this track. From the two-way travel time of the bottom bounced returns, the water depth and sediment thickness (obtained from the travel time using an assumed layer sound speed) were deduced. This provides the ground truth of the sediment layer structure over the track discussed here.

A seismic reflection/refraction experiment was conducted at site 4 using a bottom mounted source and a towed receiver near the surface at different ranges. Like multichannel seismic profiling methods, which use a source and a streamer array near the surface, the surface-bottom or bottom-surface bounced arrival lags behind the bottom-bounced arrivals by twice the water depth divided by the source speed, allowing a clean measurement of the arrival time of bottom reflected and sediment-penetrating refracted ray arrivals. From the arrival time versus the ray angle, determined from the source-receiver range and depth, one can estimate the layer structure based on extrapolation to the normal incidence.

This simple method is usually not applicable to sonar array data because the source and array are normally de-

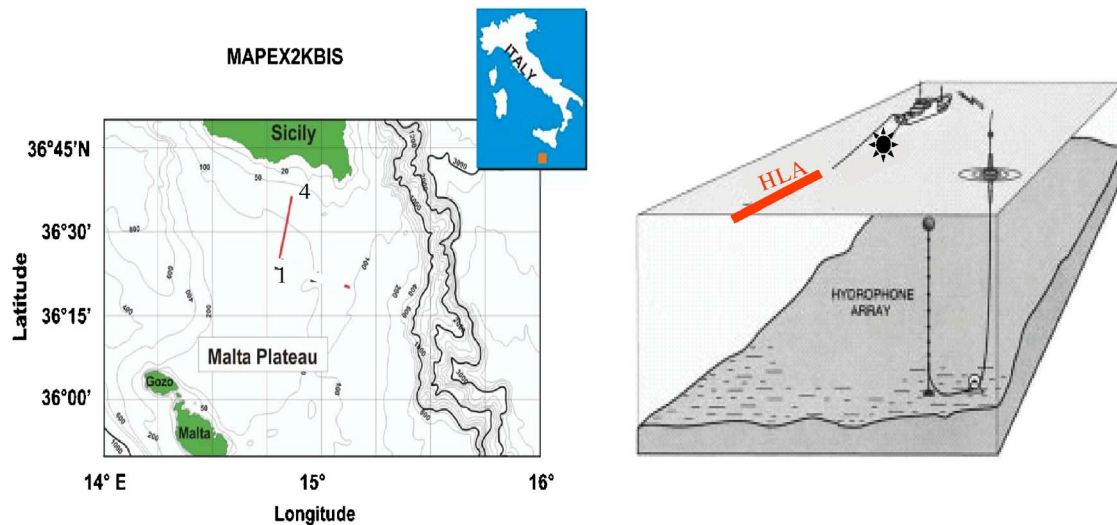


FIG. 6. (Color online) Location of the MAPEX2000 experiment and the assets used in the experiment.

ployed in the water column, and consequently, the surface and bottom bounced arrivals are interleaved with each other and not well separated in time. To examine the ray arrival as a function of the range to the source, the LFM signals received on the array elements are processed using the transmitted LFM signal as the matched filter to obtain the channel impulse response for each receiver. A snapshot of the channel impulse response function is shown in Fig. 7(a) as function of the receiver number, counting from the end closest to the ship. One observes many arrivals in the time series data, and it is not obvious which arrival corresponds to which ray path without a detailed acoustic ray analysis. The surface-bottom arrivals can overlap in time with the sediment penetrating bottom-bounced arrivals, and are not clearly identified in the raw time series data. To separate the bottom arrivals from the surface and surface-bottom arrivals, a beamforming approach is used in this section. The angle and time plot will be shown to provide a tool for ray identification and analysis. Beamforming has the advantage that it

improves the signal-to-noise ratio (SNR) for a weak sediment penetrating return so that a less powerful source can be used.

A. Beam time series

The deduced channel impulse time-series data [Fig. 7(a)] are beamformed to obtain the beam time series for a given beam angle. Beamforming is done in the frequency domain. The data are Fourier transformed and beamformed for each frequency component within the signal band. The beamformed data are inverse Fourier transformed to obtain the beam time series. The resulting beam intensity (dB) is shown in Fig. 7(b) as a function of time for each beam angle. The y-axis denotes the delay time between the late arrivals and the direct path.

One observes in Fig. 7(b) several arrivals at grazing angles around 8° , 19° , 25° , 31° , and 43° . These can be interpreted using ray acoustics. Based on ray tracing from the

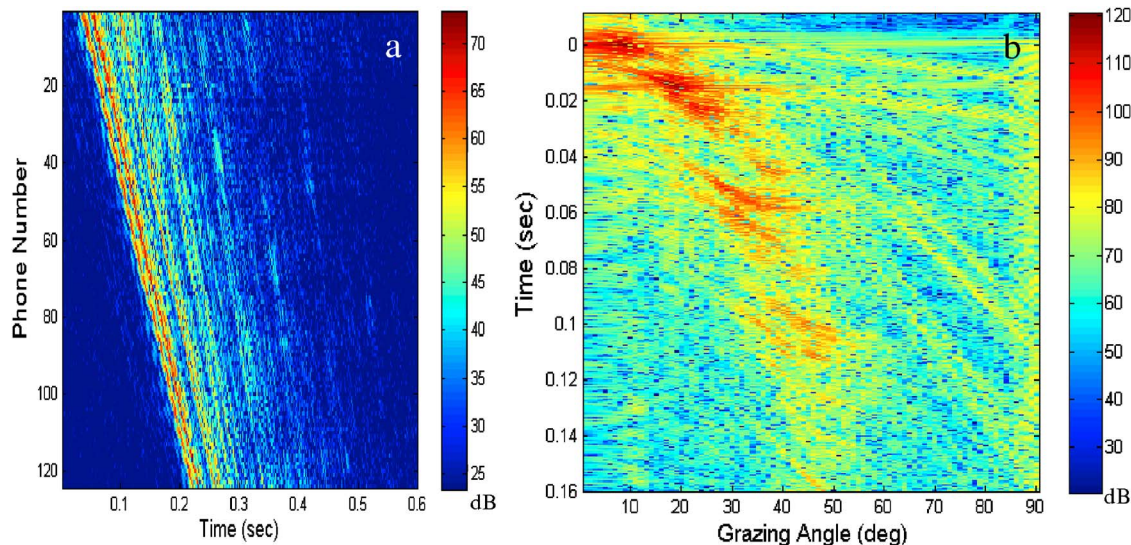


FIG. 7. (Color online) (a) Channel impulse response measured from one LFM signal for individual receivers on the towed array data. (b) The beam time series as a function of beam angle counting from forward endfire.

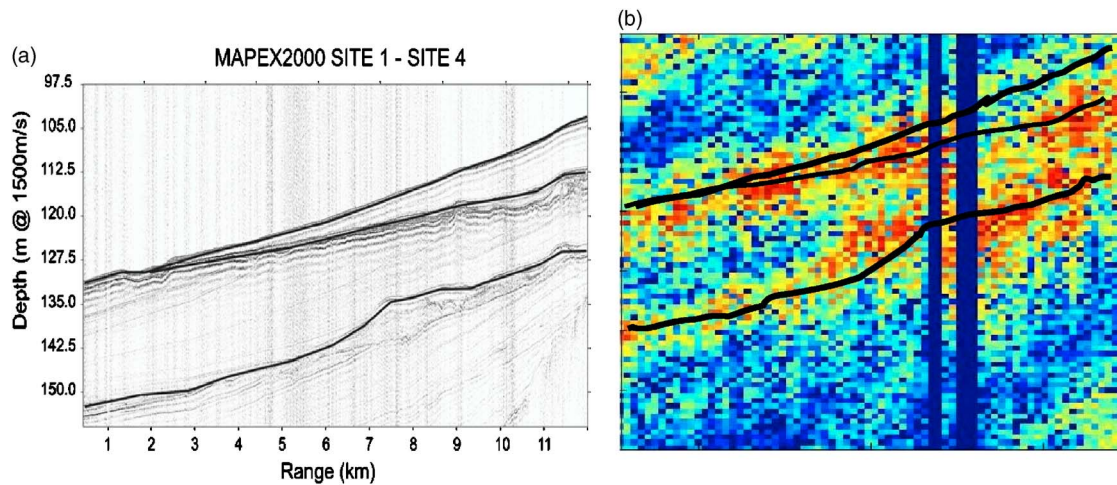


FIG. 8. (Color online) (a) Effective depth (travel time \times 1500 m/s) as a function of range determined from the uniboomer. (b) Effective depth over the same track estimated from the towed array data. The axes in (b) is scaled to that of (a) and the color bar is the same as given in Fig. 7(b) (see Sec. III B for details).

source to the center of the array, using the nominal source and receiver range depth, one finds a direct path arrival at 10° , a surface (S) reflected arrival at 15° , a bottom (B) bounced arrival at 18.5° , a BS arrival at 32° , and SBS/BSB arrivals at 41° – 43° ; the arrival at 25° was not explicitly found and is interpreted as the SB arrival. One notes that for each arrival, there are two beam “wave fronts.” Considering the direct path arrival at $\sim 8^\circ$, one observes a “wave front” at $t=0$. These are the sidelobes in other-than-the-signal arrival angle, the well-known beam sidelobes. One also sees a curved wave front with a longer delay time at higher grazing angles. These are the sidelobes in other-than-the-signal arrival time, the time domain sidelobes. This sidelobe pattern is mainly a function of the signal arrival angle and can be modeled using a simple signal model.

It should be noted that the sidelobe level using plane wave beamforming is higher than that using spherical wave beamforming; the latter is more appropriate when the array aperture is of the order of the source-receiver range. Plane wave beamforming is used here as a tool to identify the arrival time of main lobe; the sidelobes are of no interest, so long as they do not create any interference in the measurement.

It should also be noted that the time-angle display of the beam intensity is not the same as the tau- p plot (or time-angle plot) shown in Figs. 3(b) and 4(b), although there is a relationship between the two. If a ray arrival is clearly identified in the received time series, one can stack the ray travel time at different ranges versus the ray arrival angle, following the same ray, to determine the tau- p relation. For the towed HLA, as remarked above, the time series data consists of multiple ray arrivals, including both the waterborne and bottom penetrating rays, which are not always distinguishable in the time series. Beamforming is used in this paper and it will be shown that one can extract the tau- p relation of the sediment layer from the beam intensity plot. The difference is that tau- p is sparsely sampled at a few points using a (short) towed line array at a fixed range from the towed source, versus dense sampling afforded by a long HLA at different ranges in seismic reflection/refraction measurement.

The principle of tau- p estimation is as follows: A ray bounced from the water-sediment interface and a ray bounced from the sediment-bottom interface are expected to show up as two arrivals in approximately the same beam. The relative delay time between the two can be used to estimate the sediment thickness and is illustrated next using the MAPEX2000 data. The relative delay time at different grazing angles yields the tau- p relation.

B. Subbottom profiling over the ship track

Note in Fig. 7(b), there are two arrivals in the beams centered around 19° , 31° , and 43° . These are the bottom and surface-bottom arrivals. The early arrival in each pair corresponds to a ray reflected from the water-sediment interface, whereas the later (second) arrival is from a ray that penetrates the sediment and reflects from the sediment-bottom interface. The delay time between the two can be used to estimate the sediment thickness. From Eqs. (1) and (2), one finds the two-way travel time for a ray penetrating the sediment is given by $\Delta t = 2z_1 \sin \theta / c_1$, where θ is the ray grazing angle. Therefore

$$z_1 = c_1 \Delta t / 2 \sin \theta. \quad (3)$$

This relation can be used to estimate the effective depth of the sediment layer given an assumed sound speed.

Beamforming is applied to 80 LFM signals received on the HLA, whereas the ship traversed over the track shown in Fig. 6. For each ping, the beam intensity at $\theta=31^\circ$ for a fixed window of delay time (approximately 0.04–0.07 s) is saved and concatenated and plotted in Fig. 8(b). It is customary to display the result by converting the delay time into effective depth using Eq. (3); this is done in Fig. 8(b), assuming $c_1 = 1500$ m/s. The method yields the sediment thickness as a function of range.

To assess the quality of the measurement, one compares the sediment thickness ensemble using this method with that using the uniboomer. The uniboomer (echo sounder) was deployed along the track to measure the water and sediment layer depth. The measurement result is shown in Fig. 8(a),

assuming a sound speed of 1500 m/s. These measured water and sediment depths are overlaid on top of the beam color plot [Fig. 8(b)] as the dark lines. One finds that the effective layer depth measured using the subbottom profiling method is generally in good agreement with that using the uni-boomer. The uni-boomer measurement has a higher resolution because it uses a much higher frequency band with a large bandwidth. However, it requires a dedicated instrument and uses a high source level in order for the sound to penetrate the bottom. With the subbottom profiling method based on through-the-sensor data, one can get an approximate estimate of the layer structure and layer thickness without the need to deploy an echo sounder or other specialized equipment.

It was demonstrated in Ref. 10 that geoaoustic inversion with a towed line array can get consistent and reliable estimates of the bottom parameters if the parameter search is limited to a reasonable window around the true parameter values; otherwise the solution may become erratic. Earlier geoaoustic inversion work applied to the MAPEX2000 data constrained the sediment thickness parameter search window based on prior information gathered from the uni-boomer data.^{8–10} This section shows that, even without the uni-boomer data to guide the search, one can determine the number of sediment layers and the sediment thickness from the same acoustic data used for geoaoustic inversion by using the subbottom profiling method.

IV. INTERPRETATION OF THE DATA: THE QUESTION OF UNIQUENESS

Geoaoustic inversion is a numerical method used to estimate bottom parameters by matching the modeled acoustic field with the received data. As with any numerical method, the physics governing the inversion results is not explicit from the numerical outputs and should be understood for proper implementation of the method and interpretation of the results. The questions of interest are: What physical information is actually contained in the data (what parameters can one determine from the data) and how well it is determined (i.e., the sensitivity issue). These questions are addressed in this section. Also addressed in this section is the question of uniqueness in geoaoustic inversion using a towed line array. As geoaoustic inversion uses commonly a global search algorithm, the question of uniqueness can be posed as whether the geoaoustic inversion will always converge to the same peak in the ambiguity surface. This concerns the consistence and robustness issue which was addressed in Ref. 10. The contribution of the subbottom profiling method is (additional) information about the bottom layer structure and thickness which confines/narrows the parameter search windows so that consistent and robust solutions can be obtained.

A more specific question is whether the bottom parameters are uniquely determined. As discussed earlier, if the reflection loss or the bottom impulse response is known as a function of angle frequency or angle time, then the bottom is in principle uniquely determined. With respect to a towed line array, one finds that the reflection coefficients or the impulse responses are only sampled at a number of discrete angles. Does this lead to a unique inversion result? Although

theoretically this does not prove the uniqueness of the bottom, in practice it is difficult to find two or more physical bottoms that will fit the data if the towed array covers a decent span of reflection angles. In that case, the bottom can be considered, for all practical purposes, determined by the towed array data. However, the “uniqueness” of the result is only valid in the context of the given data uncertainty and measurement limitation as discussed at the end of Sec. IV D.

A. Geometric parameters

For a constant sound speed profile, the direct, surface-reflected (S), bottom reflected (B), bottom-surface-reflected (BS) and surface-bottom reflected (SB) rays arrive at a receiver with grazing angles θ_1 , θ_2 , θ_3 , θ_4 , and θ_5 , respectively, where

$$\begin{aligned}\tan \vartheta_1 &= \frac{|z_s - z_r|}{r}, & \tan \vartheta_2 &= \frac{|z_s + z_r|}{r}, \\ \tan \vartheta_3 &= \frac{|2H - z_s - z_r|}{r}, \\ \tan \vartheta_4 &= \frac{|2H - z_s + z_r|}{r}, & \tan \vartheta_5 &= \frac{|2H + z_s - z_r|}{r}.\end{aligned}\quad (4)$$

Here r is the source–receiver range, z_s and z_r are the source and receiver depth and H is the water depth. In principle, one could determine the source receiver range, depths, and water depth from the previous five equations by simply measuring the arrival angle. Similarly, one could determine the five geometric parameters using the ray arrival time. Because the ray arrival angle and time can be read off from the beam output data, these parameters are uniquely determined by geoaoustic inversion. However, as indicated by Fig. 7(b), the angle and time resolution are insufficient to determine the range and depths precisely, unless high frequencies are used. Therefore, for geoaoustic inversion at low frequencies (<1 kHz), one still needs direct measurements of the source–receiver range and depths (e.g., via engineering sensors) as well as direct measurement of the water depth (e.g., via a bottom sounder).

B. Measurement of the tau– p relation

Recall that the tau– p relation, discussed previously in Secs. II and III, is a display of the ray travel time through the sediment as a function of the ray slowness or ray grazing angle. The ray travel time through the sediment can be measured from the beam intensity-versus-time plot, based on the two-way ray travel time through the sediment layer at angles associated with the bottom returns. To illustrate this method, another LFM signal collected at a site close to the left-hand edge of Fig. 8 (range <1 km) is analyzed in this section. The LFM time series after the matched filter processing is shown in Fig. 9(a). It shows a longer multipath spread than Fig. 7(a), indicating more bottom bounced returns. Figure 9(b) is the beam time series as function of beam angle, and shows 2 or 3 returns from the bottom; the number of returns and the time separations between the returns are more clearly seen at higher grazing (e.g., >40°) angles. As discussed for Fig.

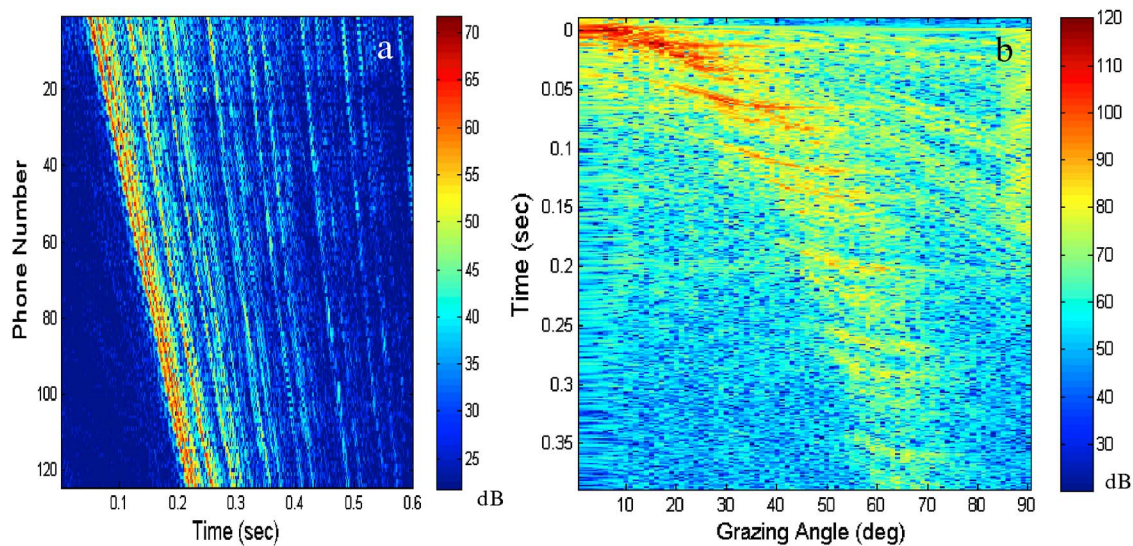


FIG. 9. (Color online) (a) Channel impulse response measured from another LFM signal with deeper bottom penetration. (b) The beam time series as a function of beam angle counting from forward endfire.

7(b), the first return is the reflection from the water–sediment interface, and the second return is from the bottom of the sediment layer. Does the third return suggest another layer below the first layer? To determine this, one examines the travel time differences between the arrivals and compares them with the modeled impulse response. First, the travel time differences between the later and first bottom returns are measured at selected grazing angles and shown in Fig. 10(a), where the y-axis is conventionally converted into an effective depth using the mean sediment sound speed. Figure 10(b) shows the modeled bottom impulse response using the bottom parameters obtained via inversion of acoustic data from this site.^{10,13,14} The impulse response (in pass band) has positive and negative pressures which show up as two peaks

(in absolute value) in the main lobe. The impulse response has a sidelobe in delay time which shows up in the depth plot as a later arrival. Comparing Fig. 10(a) with Fig. 10(b), one concludes that the third bottom return is a sidelobe and does not suggest the existence of a second layer under the first layer.

The previous analysis shows that typical towed line array data samples the tau– p relationship only at a few discrete angles as compared with the almost continuous sampling of $\tau(p)$ in a seismic study. Despite the small number of samples, the tau– p relationship reveals important information about the bottom. In this example, one finds, first, the bottom has effectively one layer of sediment (geoaoustically speaking, for the frequency band of interest). Second, the layer thick-

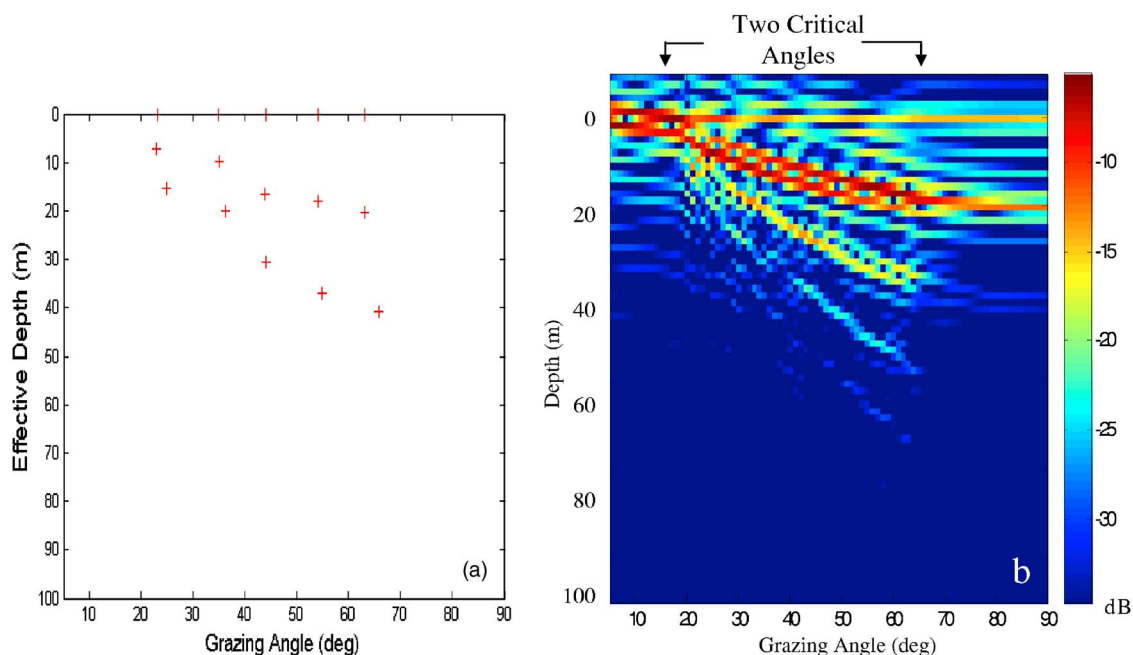


FIG. 10. (Color online) (a) Measured bottom impulse response at discrete angles plotted in terms of the effective depth. (b) Modeled bottom impulse response plotted in terms of the effective depth.

ness is approximately 20 m. Note that although it is theoretically possible to fit more than one curve through the data points in Fig. 10(b), it is physically impossible to find two or more types of bottoms that will yield the same τ - p relation. Applying the same analysis to other LFM signals along the track (even though the number of τ - p samples may vary between pings), the deduced sediment thicknesses are stacked in time, revealing the bottom structure along the track as shown in Fig. 8(b).

C. Bottom reflection loss

For the reflection loss analysis, the last LFM signal analyzed previously and reported in Ref. 14 will be used. The geoacoustic inversion result is shown by the left most data point in Fig. 5 of Ref. 14, located at the x -coordinate of 14.715; the x -axis labels the longitude where the data is taken. Although other parameters remain approximately unchanged, this last ping shows a significantly lower bottom sound speed (1700 m/s) when compared with the bottom sound speed determined from previous pings (1780–1800 m/s at longitudes of 14.73°–14.74°). This inversion result was not reported in Ref. 10, as there were some questions about the inversion result. (Note that over the track, the bottom sound speed was changing on a scale of approximately 5 km. If a change in bottom speed is supported by multiple pings, the results are generally believable. This was the last ping processed and it showed a significant change in bottom sound speed, so it is natural to question this geoacoustic inversion result). Like any numerical inversion method, the problem one faces is a lack of confidence in accepting or rejecting a particular inversion result. Ideally, the value of the cost function can be an indicator/measure of the acceptability of a solution. In practice, the cost function value is often not discriminating enough to be useful. On the other hand, if geoacoustic inversion is to become a practical tool, one needs to ensure that a high percentage of the inversion results is reliable. This requires an additional analysis tool. Toward this goal, we turn to a physical analysis based on the reflection loss measurement. The analysis will show that a bottom sound speed of 1800 m/s does not fit the reflection loss data well, while a speed of 1700 m/s provides a better fit to the data. This suggests that the geoacoustic inversion result indeed has some validity.

Applying the same processing as shown in Sec. III B, the beam time series deduced from this LFM signal data is shown in Fig. 11(a). Figure 11(b) shows the beam time series calculated using the bottom properties inverted from the last data point as reported in Ref. 14, and summarized in Table I. One finds that the modeled arrival angles and arrive time are in good agreement with data. The beam intensities at high grazing angles relative to that of the direct arrival seems a little lower in the modeled beam plot than in the data beam plot. Note that at high angles, a small difference in the reflection loss (per bounce) between the model and data is amplified in the data by the multiple numbers of bottom reflections. Although the high angle data can in principle be used to estimate the high angle bottom reflection loss, the high angle data have a weak SNR making the estimation less

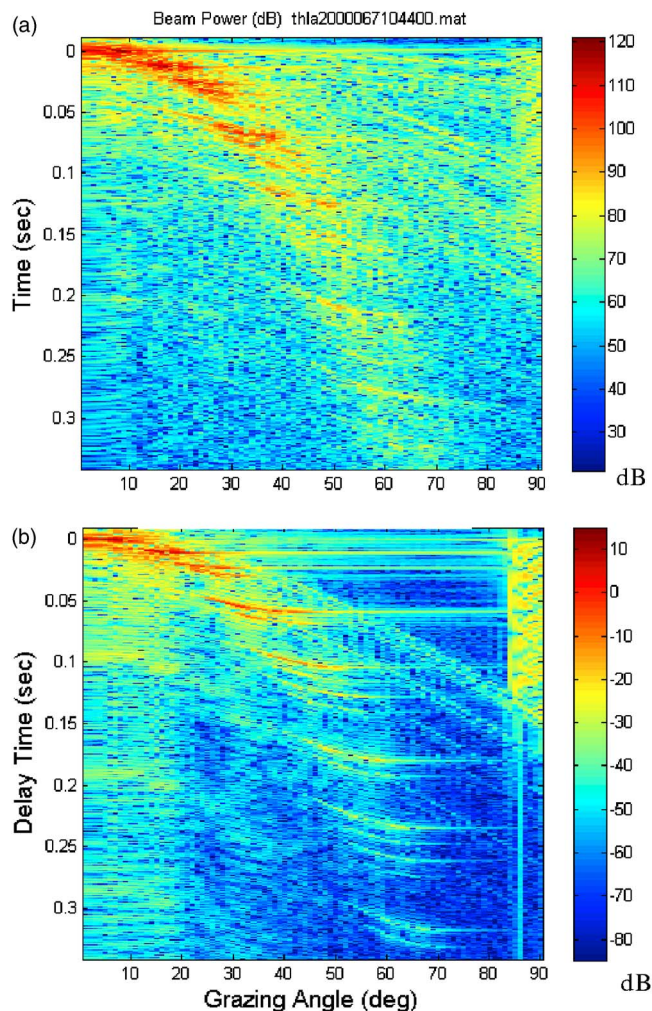


FIG. 11. (Color online) (a) Measured beam time series compared with (b) the modeled time series using the previously inverted bottom parameters.

reliable. To estimate the bottom reflection loss, the geometrical spreading loss needs to be estimated and removed from the data, which is another source of uncertainty as discussed in the following section.

One observes in Fig. 11(a) that the first arrival covers a beam angle from 0° to 15° with a peak around 8°. This beam contains the direct and reflected arrivals, based on the arrival angles estimated using ray tracing. The next arrival, centered at 15°–18°, is a bottom bounced arrival. The arrivals at approximately 25° and 33° are surface-bottom and bottom-surface reflected arrivals. One can also identify arrivals at 40° and 50°, which are multiple bounced ray arrivals based on the arrival time estimates.

Recall in Sec. III B, the beam time series was obtained by inverse Fourier transform of the beamformed frequency

TABLE I. Bottom parameters inverted from acoustic data; ρ is density and α represents attenuation.

	Thickness (m)	Sound speed (m/s)	ρ (g/cm ³)	α (dB/ λ)
Sediment	13.5	1573.6	1.8	0.22
Bottom		1710	1.8	0.55

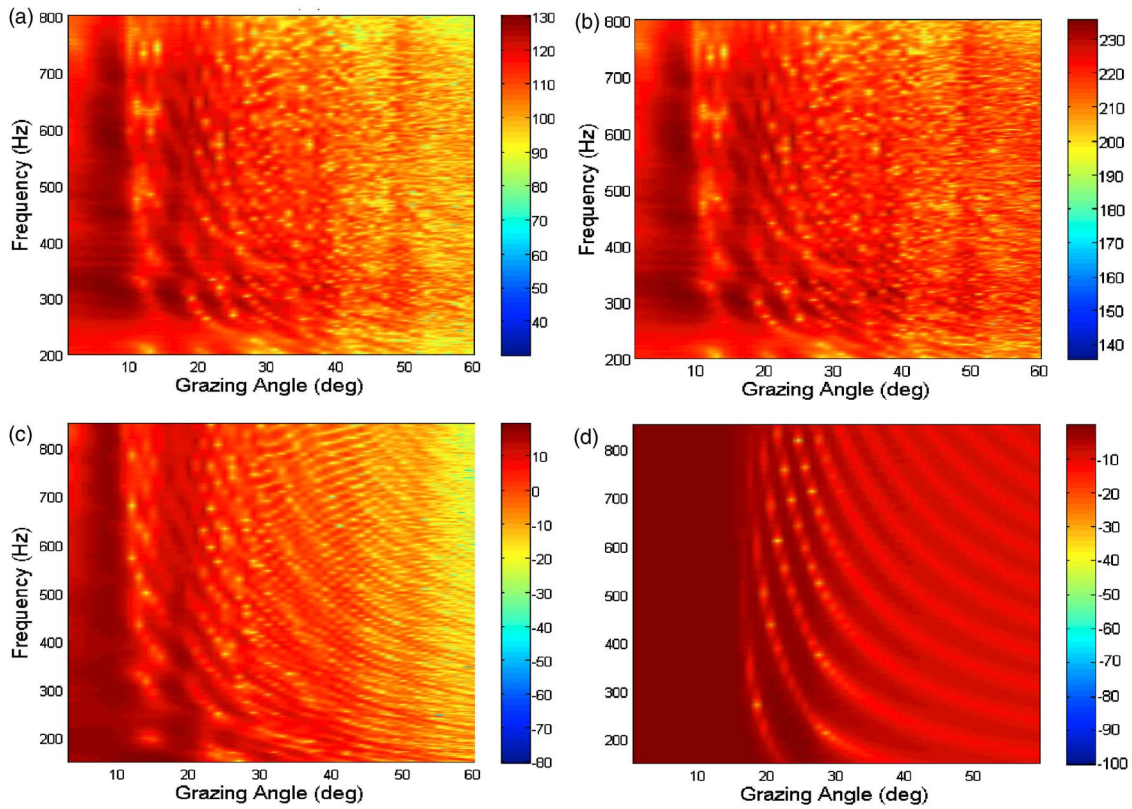


FIG. 12. (Color online) (a) Measured beam spectral level as a function of angle and frequency. (b) Beam spectral level compensated for the geometrical spreading loss. (c) Modeled beam spectral level and (d) reflection loss as a function of angle and frequency using the inverted bottom. Color scale in decibels.

data. Here, the beamformed frequency data will be examined as a function of beam angle and frequency, which is shown in Fig. 12(a) for this event.

The beam spectral level can be expressed theoretically by

$$B_{\text{pwr}}(\vartheta, f) = \sum_i \Psi_i(\vartheta_i, f) B(\vartheta - \vartheta_i, f), \quad (5)$$

where f is the acoustic frequency, Ψ_i denotes the acoustic field for the direct, surface, bottom surface-bottom arrivals, and

$$B(\vartheta - \vartheta_i, f) = \left[\frac{\sin(\pi N d (\cos \vartheta - \cos \vartheta_i) / \lambda)}{N \sin(\pi d (\cos \vartheta - \cos \vartheta_i) / \lambda)} \right]^2 \quad (6)$$

is the beam pattern for an HLA of N elements spaced at d . Based on Fig. 11(a), the signal arrivals are at approximately 8° , $15\text{--}18^\circ$, 30° , 40° , 45° , etc. The energy between these angles is due to the nonzero beam width, i.e., the beam pattern, of the HLA. For example, the null in the beam intensity at around 13° in Fig. 12(a) is due to the fact that the adjacent two beams are far apart. Above 30° , the beam intensities are a smooth function of angle because there are many arrivals with multiple surface and bottom reflections arriving as well as closely spaced arrival angles.

The next step is to estimate the bottom reflection loss from the beam intensities. To do so, one needs to correct the spreading loss for the various paths. One can estimate the spreading loss based on the ray path length (using ray tracing), or using a propagation model with a totally reflective bottom; here the spreading loss is estimated for simplicity

assuming a constant sound speed profile (the error is estimated to be <1 dB for angles $\leq 30^\circ$). The spreading loss is calculated using the center of the array as the receiver and then applied to the beam data. The beam pattern after such correction is shown in Fig. 12(b). The spectral level averaged over frequency is shown in Fig. 13 in a relative scale for bottom bounced arrivals.

Figure 12(d) shows the reflection loss (dB) as a function of frequency using the bottom parameters determined from geoacoustic inversion (Table I). Reflection loss is calculated

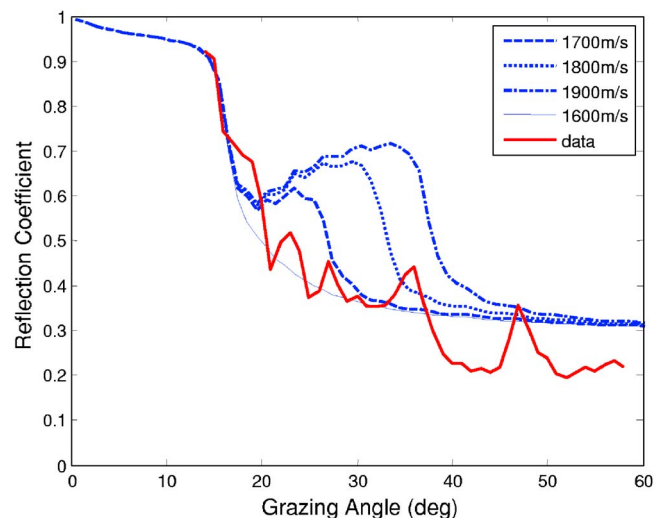


FIG. 13. (Color online) Measured and modeled reflection coefficient averaged over frequencies.

for all grazing angles using the OASES model. Figure 12(c) shows the modeled beam spectral level (dB) as a function of frequency and angles (using the same model) for a 0 dB source level using the inverted bottom parameters. It shows a qualitative agreement with data [Fig. 12(a)] except at high grazing angles. Figure 12(d) shows the reflection loss (dB scale) for all grazing angles. In comparison, Fig. 12(b) shows the reflection loss (relative dB scale) measured at discrete angles corresponding to the surface and bottom bounced arrivals on the towed line array.

In Fig. 13 the calculated reflection loss averaged over the frequency band is plotted using four different bottom sound speeds from 1600 to 1900 m/s, keeping the rest of the bottom parameters fixed. The data are normalized so that the levels of the bottom bounced return at 15°–18° agree with the model. Comparing the modeled reflection loss with the data, one finds that bottom with sound speed equal to or greater than 1800 m/s is not favored by the data; the data favor a bottom sound speed somewhere in between 1600 and 1700 m/s. This comparison suggests that geoacoustic inversion which yields 1710 m/s at this location is probably correct even though the inversion result at previous locations showed a bottom sound speed of 1800 m/s. Note that the above reflection coefficient estimate is subject to the assumption of a constant sound speed profile; the error increases with the number of bottom reflections. The geoacoustic inversion includes the sound speed variation with depth and consequently, its result should be more reliable.

D. Interpretation of the data: The question of uniqueness

The previous analysis can be summarized as follows. The tau- p relation determines uniquely the number of the bottom layers and layer thickness. Given the layer structure, one can determine the bottom parameters by fitting the reflection loss (amplitude) and phase with the data.^{15,16} The problem is that the reflection coefficient is measured only at a few discrete angles on the towed line array. With only discrete measurements, is there more than one kind of bottom that will produce the same reflection coefficients at these angles? For a simple one layer case, the answer is likely to be “no.” The reason is that for different bottoms found in the oceans, the parameters of the bottom are constrained by the physical properties of the material, and are not free to change arbitrarily; they are confined to a certain range of values. As remarked earlier, the towed array inversion is intended for low (<1 kHz) frequencies; at low frequencies, most of the bottoms can be treated geoacoustically as one or two layers of sediment over a hard bottom. At high frequencies, the bottom may need to be treated as multiple layers, even then the parameters between layers are closely related.

To determine if there is more than one kind of bottom which can fit the reflection coefficients data, one considers twelve typical sediment types commonly found in shallow water,¹⁷ whose properties are summarized in Table II ranging from silty clay to sand and gravel. It is assumed that the hard bottom below the sediment layer has the properties determined in the previous section, the reason being that the same

TABLE II. Shallow water sediment types and properties; ρ is density and α represents attenuation. Sediment sound speed is given by the speed ratio times the sound speed at the bottom of the ocean (1512.2 m/s).

No.	Sediment type	Speed ratio	Speed (m/s)	ρ (g/cm ³)	α (dB/ λ)
1	Silty clay	0.994	1503.1	1.40	0.03
2	Clay-silt	1.004	1518.2	1.41	0.03
3	Clayey silt	1.010	1527.3	1.49	0.03
4	Sand-silt-clay	1.020	1542.4	1.44	0.15
5	Silty sand	1.030	1557.6	1.61	0.11
6	Sandy silt (fine sand)	1.040	1572.7	1.81	0.20
7	Silt (medium fine sand)	1.060	1602.9	1.82	1.89
8	Medium sand	1.080	1633.2	1.95	3.00
9	Medium coarse sand	1.100	1663.4	1.97	2.70
10	Coarse sand	1.120	1693.7	2.00	2.55
11	Silt-sand-gravel	1.200	1814.6	2.00	2.40
12	Gravel	1.250	1890.3	2.20	2.40

hard bottom will be favored when the analysis of the previous section is repeated for each sediment type. For each sediment layer over the hard bottom, the reflection coefficient is calculated as a function of grazing angle and averaged over the frequency band of the data, as was done in Fig. 13. The results are shown in Fig. 14 and compared with the reflection loss coefficient measured from the data (solid line).

One finds that for types 7–12 (medium sand to gravel), the reflection coefficient decreases slowly with an increasing grazing angle, and the rate of decrease does not agree with data. For types 1–3, the “interference” pattern (between the returns from the water-sediment and sediment-bottom interfaces) of the reflection coefficient does not agree with the data. The modeled reflection coefficients for types 4–6 are close to the data, with type 6 yielding a better fit. Note that the modeled reflection for type 6 is almost identical to that shown in Fig. 13 for the bottom sound of 1600–1700 m/s. Indeed, the sediment parameters obtained from geoacoustic inversion (Table I) are very close to that of the fine sand (type 6) given in Table II. This suggests that a simple mea-

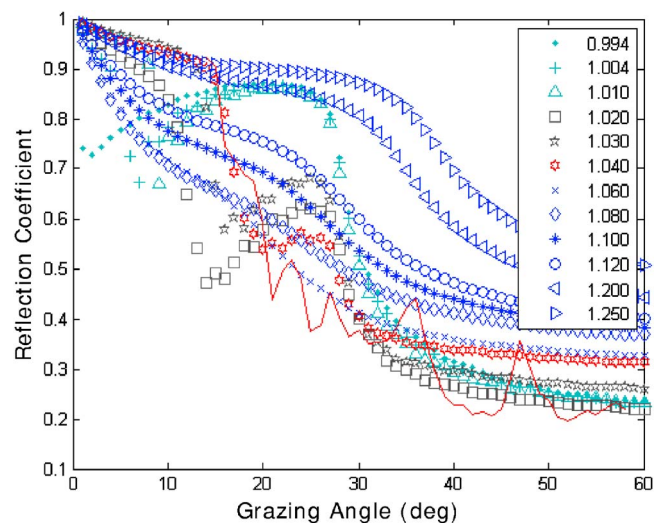


FIG. 14. (Color online) Reflection coefficient modeled for 12 different sediment types averaged over frequencies. The data shown in Fig. 13 is repeated.

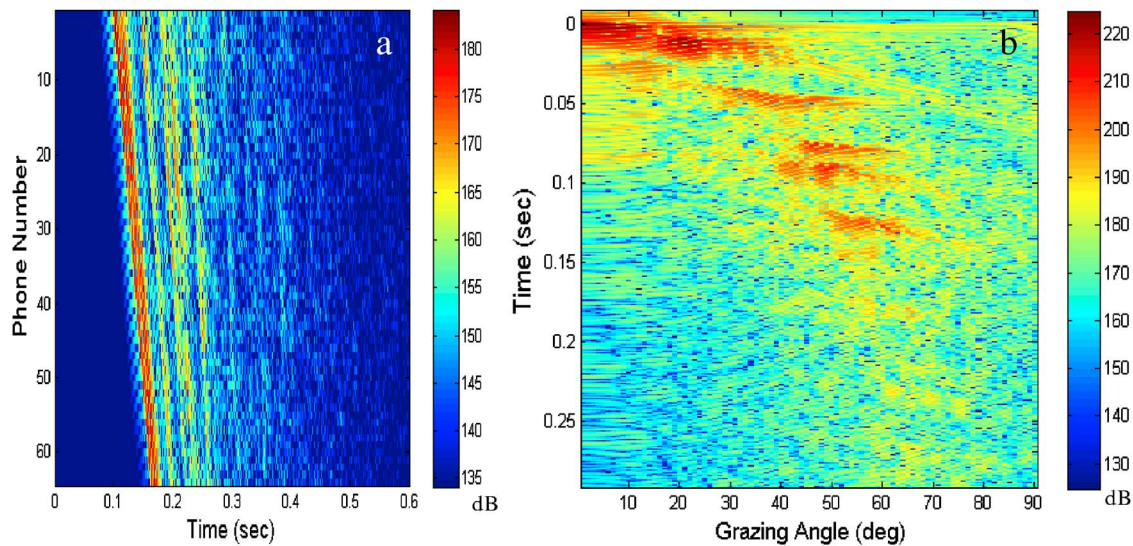


FIG. 15. (Color online) (a) Channel impulse response measured from one LFM signal for individual receivers on the towed array data. (b) The beam time series as a function of beam angle counting from forward endfire. OREX05 data.

surement of the reflection loss could discriminate the different bottom types, and the preferred bottom type could then be used to predict the TL.

On the other hand, given measurements of the reflection coefficients covering a wide range of angles of bottom reflections, as when a long towed line array is available, one may be able to determine the bottom parameters as the reflection loss and the τ - p relation uniquely determine the bottom impulse response. The result is a more refined measurement of the bottom parameters than that suggested by the 12 bottom types used previously.

The earlier reflection loss analysis describes the physics intrinsic to geoacoustic inversion. In practice, geoacoustic inversion may be preferred over the reflection loss analysis as a numerical search algorithm (used in geoacoustic inversion) can be more easily automated. Note that the objective of geoacoustic inversion, used in the context of REA, is not necessarily a precise determination of the bottom parameter but rather a capability to predict TL as a function of range. Unfortunately, to this date, geoacoustic inversion has not been put into practical use. The problem has been that inversion methods often have multiple solutions. Ref. 10 shows that if the search windows are narrow and confined around the true parameters, the algorithm should find the correct solution. For a towed line array, the search windows for the geometric parameters (source-receiver ranges and depths) are constrained by the measurements of the engineering sensors. Previously unknown is how to properly constrain the search window of the parameters of the sediment layers and bottom. The subbottom profiling method presented here provides this critically needed information, and it uses the same data as the geoacoustic inversion method. It enables geoacoustic inversion to obtain consistent and reliable solutions and is useful for REA and for building a data base; the uniqueness of the solution is of a lesser consequence from a practical point of view.

V. THE OREX05 DATA

The OREX05 experiment was conducted by the Naval Research Laboratory off the coast of Oregon in June 2005.

Although the purpose of the experiment was focused on reverberation measurement, a short segment of the experiment was devoted to geoacoustic inversion using LFM signals transmitted from a towed source received on a towed line array. This data collection took place southwest of Alsea Bay at $44^{\circ} 16.5135' N$ and $124^{\circ} 48.2663' W$. The water depth is 113 m.

The bottom on the shelf is complex and not very well defined.¹⁸ It consists generally of a thin layer of recent or relict sediment overlying either a well indurated sedimentary rock or a poorly indurated rock. The sediment is silt-sand for the most part, with some areas of sand and some areas of silt. The thickness of the sedimentary layer varies from 1.5 to 15 m. During OREX05, a chirp sonar was used to estimate sediment thickness and subbottom properties; in many places, the only return was from the water-sediment interface and there was no evidence of sound penetrating into the bottom. The problem is attributed to the high attenuation of the sediment layer at high frequencies.

With this lack of information about the area, the LFM data are analyzed to see what information can be revealed by the subbottom profiling method. Although originally the segment run was to cover tens of kilometers, only 39 pings of LFM signals collected over a range of 2 km were available. Nevertheless, the utility of the method can still be demonstrated.

The sound speed in this area is downward refractive, typical of a summer profile. The sound speed increases nearly quadratically from 1476 m/s at the bottom to 1494 m/s at a depth of 2.5 m, then to 1503 m/s at the surface. The source was towed directly behind the ship. The towed source depth was not accurately measured and was estimated by the length of cable paid out to be 25–30 m. The array was towed at a depth of 45 m. It has 64 elements spaced at 1.5 m. The range from the towed source to the first phone on the array is estimated to be approximately 121.5 m.

Figure 15(a) shows the channel impulse response determined from the first LFM signal as a function of the phone number on the array. The LFM signals covered a frequency

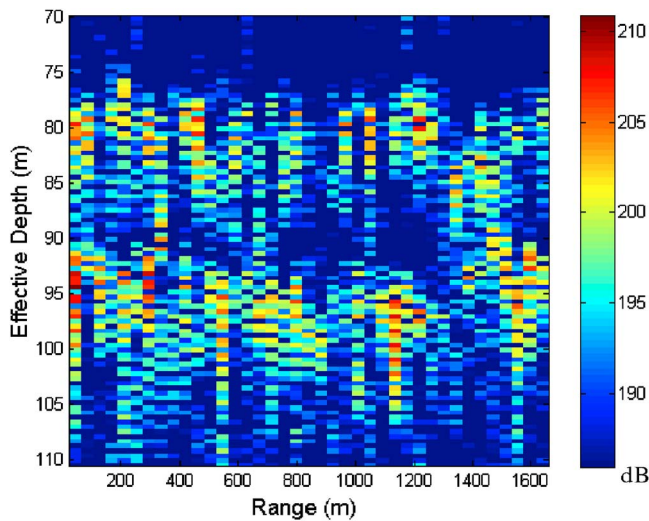


FIG. 16. (Color online) Sediment layer thickness as a function of range estimated from the towed array data. OREX05 data depth reference arbitrary.

band from 300 to 1200 Hz. Figure 15(b) shows beamformed time series versus angle. One notes that the multipath arrivals are clearly separated in the time domain [Fig. 15(a)]. The reason is the short distance between the source and the receiver array, which allows a large separation in time, and the large bandwidth of the signal (1.1 kHz), which yields a narrow pulse width. The multipath arrival angles are clearly seen in Fig. 15(b). One finds the direct path arrives at an angle somewhere between 0° and 6° . The surface reflected path arrives approximately at 21° . The bottom bounced path arrives at approximately 41° and the surface-bottom path arrives at approximately 50° . The surface-bottom shows clearly two arrivals which can be used to estimate the sediment thickness.

Using the method described in previous sections, one measures the travel time difference at 50° from each ping, and converts it to an effective depth using a sound speed of 1600 m/s. (The sediment sound speed in the general area has a sound speed of 1600 m/s at the top of the sediment.¹⁸) The effective depth is shown in Fig. 16 as a function of range. It

shows a mean layer thickness of approximately 15 m, with a variation of a few meters. Note that because the thickness measurement is based on the travel time difference, it is insensitive to the array motion. Note also that the energy of the bottom return varies significantly from ping to ping, sometimes by as much as 10 dB; the cause is unknown and may be due to bottom roughness.

The previous result is a first measurement of the sediment thickness in this area, and so there is no measurement to compare with. As remarked earlier, the chirp sonar used during the experiment failed to penetrate the sediment due to the high frequency of its source (showing that there is a value in using a low frequency signal as projected by the towed source).

Figure 17(a) shows the beam spectral level as a function of angle, corrected by the spreading loss associated with each path as described in Sec. IV C. The spectral level at angles associated with bottom bounced arrivals (40° and higher) are normalized by the spectral levels at 2° – 3° (the direct path arrival) giving the reflection loss due to the interaction with the bottom. The average of the reflection loss (over the signal frequency band) is shown in Fig. 17(b), plotted in terms of its amplitude. Note in Fig. 17(a), at frequencies above 500 Hz, the energy between 12° and 18° is due mainly to the sidelobes of the direct path arrival, while the arrival around 20° is due to surface reflection. The null between 25° and 35° is due to lack of signal arrivals at these angles. Therefore, the energy above 500 Hz, for grazing angles less than 35° , is not related to the bottom. Similarly, the data in Fig. 17(b) at angles less than 35° is not related to bottom; the bottom reflection loss is only measured at angles above 40° . The dashed line in Fig. 17(b) is a typical model prediction of the bottom reflection coefficient, showing that important information about the reflection loss is missing below 40° . For this experiment, the towed line array was too close to the tow ship to adequately measure the bottom reflection coefficient; a less-than-ideal configuration for geoacoustic inversion. A simple remedy will be to tow the array at least 300 m away from the source/ship, similar to the line array during the MAPEX2000 experiment.

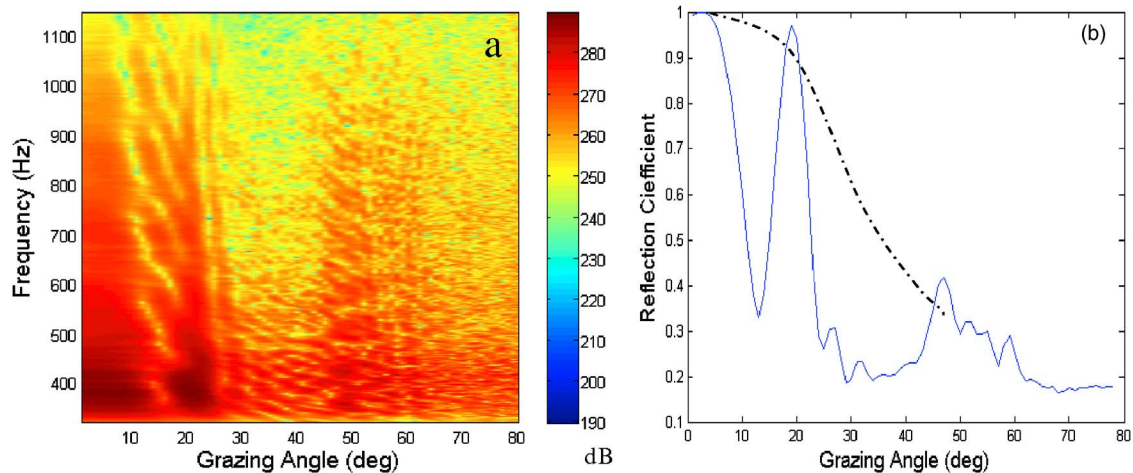


FIG. 17. (Color online) (a) Beam spectral level as a function of angle and frequency-OREX05 data. (b) The data above 40° are used to estimate the frequency-averaged bottom reflection coefficient. The dot-dashed line indicates a typical reflection loss where measurement was missing; data points around 20° are surface returns.

For a more general environment, what is the optimal towed array depth and separation from the ship for geoaoustic inversion? The answer is dependent on the water depth and source depth; to a less degree it also depends on the sound speed profile. The primary requirement, based on the previous analysis, is that the bottom bounced arrivals on the towed array should cover a grazing angle important for reflection loss measurement (e.g., 10° – 50°). If a long array is available, once can apply the same processing to several sub-arrays. This will increase the number of data point covering a wider span of grazing angles for a more complete measurement of the bottom reflection loss.

VI. SUMMARY

In this paper, a subbottom profiling method is applied to the acoustic data received on a towed line array to estimate the sediment layer structure and thickness. This method provides *a priori* information about the bottom, needed to improve the quality (consistency and reliability) of a geoaoustic inversion. Without this information, one is forced to make assumptions and randomly set the search windows defining the sediment structure and thickness. In such cases, the inversion results of measured data can be ambiguous and erratic over an area.

The subbottom profiling method uses the same acoustic data used for the geoaoustic inversion. The novelty is that it uses the time-angle information contained in the data, which was previously ignored in typical geoaoustic inversion approaches.

The inversion physics for a towed line array is discussed, i.e., which aspects of the physical parameters are really measured by the towed line array. It is noted that the data received on the towed line array only samples the bottom reflection coefficients at certain discrete angles associated with returns from bottom and multiple surface-bottom bounces. The towed line array measurements provide the τ - p relation at these angles, which defines the sediment structure and thickness. To compensate for the discrete measurements, it helps to have a wide span of arrival angles and a wide spectrum of frequencies in order to improve the quality of the bottom measurement.

At low frequencies (<500 Hz), the geoaoustic bottom in shallow water can often be approximated by one or two layers of sediment over a hard bottom. For these bottoms, it is likely that no more than one kind of bottom can produce the same τ - p relation and the same reflection coefficients over the band of frequencies at the measurement angles. Under the circumstances, the general properties of the geoaoustic bottom model are, for practical purposes, uniquely determined.

One notes that although the bottom properties can be deduced directly from the method discussed in this paper (Sec. IV), it is more practical to use a geoaoustic inversion method as the latter can be more easily automated. The key to achieving a unique inversion solution, as shown in a previous publication,¹⁰ is to have a narrow search window around the true values. The geometric parameters (the source, receiver depth and range) need to be precisely measured using engineering sensors. The sediment structure and

thickness can be determined to a reasonable accuracy using the subbottom profiling method presented here, providing the narrow search windows necessary for a successful inversion. As such, a towed sonar line array may be used for rapid through-the-sensor bottom assessment.

ACKNOWLEDGMENTS

This work is supported by the U.S. Office of Naval Research. The authors are thankful to NURC for the MAPEX2000 data, and M. Siderius, M. Fallat, P. Gerstoft, and P. Nielson for discussions on the MAPEX data inversion. They thank the NRL team for conducting the OREX05 experiment and D. Fromm, R. Gauss, R. Soukup, and A. Turgut for discussions.

- ¹M. Siderius, P. L. Nielsen, and P. Gerstoft, "Performance comparison between vertical and horizontal arrays for geoaoustic inversion," *IEEE J. Ocean. Eng.* **28**, 424–431 (2003), and references therein.
- ²M. Siderius, P. L. Nielsen, J. Sellschopp, M. Snellen, and D. Simons, "Experimental study of geo-acoustic inversion uncertainty due to ocean sound-speed fluctuations," *J. Acoust. Soc. Am.* **110**, 769–781 (2001).
- ³D. J. Battle, P. Gerstoft, W. S. Hodgkiss, W. A. Kuperman, and P. L. Nielsen, "Bayesian model selection applied to self-noise geoaoustic inversion," *J. Acoust. Soc. Am.* **116**, 2043–2056 (2004).
- ⁴M. R. Fallat, P. L. Nielsen, S. E. Dosso, and M. Siderius, "Geoacoustic characterization of a range-dependent ocean environment using towed array data," *IEEE J. Ocean. Eng.* **30**, 198–206 (2005).
- ⁵P. L. Stoffa, *Tau-p: A Plane Wave Approach to the Analysis of Seismic Data* (Kluwer, Dordrecht, The Netherlands, 1989).
- ⁶B. R. Gomes, R. A. Fisher, and J. K. Fulford, "Environmental characterization for the Littoral Warfare Advance Development 00-3 Experiment in the Strait of Sicily and the Adriatic Sea," Naval Research Laboratory, Bay St. Louis, MS, Report No. NRL/MR/7180-01-8254, 2001.
- ⁷I. Finetti and C. Morelli, "Geophysical exploration of the Mediterranean," *Boll. Geofis. Teor. Appl. (Bulletin of Geophysics: Theory and Applications)* **15**, 263–341 (1973).
- ⁸M. Siderius, P. L. Nielsen, and P. Gerstoft, "Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array," *J. Acoust. Soc. Am.* **112**, 1523–1535 (2002).
- ⁹M. R. Fallat, P. L. Nielsen, and M. Siderius, "The characterization of a range dependent environment using towed horizontal array data," MAPEX 2000 experiment NATO Undersea Research Centre Technical Memorandum, SM-402, 2002.
- ¹⁰L. T. Fialkowski, T. C. Yang, K. Yoo, D. K. Dacol, and E. Kim, "Consistency and reliability of geoaoustic inversions with a horizontal line array," *J. Acoust. Soc. Am.* **120**, 231–246 (2006).
- ¹¹D. J. Battle, P. Gerstoft, W. A. Kuperman, W. S. Hodgkiss, and M. Siderius, "Geoacoustic inversion of tow-ship noise via near-field-matched-field processing," *IEEE J. Ocean. Eng.* **28**, 454–467 (2003).
- ¹²C. H. Harrison, "Sub-bottom profiling using ambient noise," *J. Acoust. Soc. Am.* **115**, 1505–1515 (2004).
- ¹³C. W. Holland, "Coupled scattering and reflection measurements in shallow water," *IEEE J. Ocean. Eng.* **27**, 454–470 (2002).
- ¹⁴T. C. Yang, K. Yoo, and L. Fialkowski, "Sub-bottom profiling and geoaoustic inversion using a ship towed line," *Proceedings of the Underwater Measurement and Technology Conference*, Crete, Greece, June 28–July 2, 2005.
- ¹⁵T. C. Yang and T. Yates, "Geo-acoustic inversion of bottom reflection coefficients and sound speed profiles in shallow water," *IEEE J. Ocean. Eng.* **21**, 367–376 (1996).
- ¹⁶S. A. Stotts, D. P. Knobles, J. A. Keller, J. N. Piper, and L. A. Thompson, "Geoacoustic inversion of short range source data using a plane wave reflection coefficient approach," *J. Acoust. Soc. Am.* **120**, 3607–3626 (2006).
- ¹⁷P. J. Vidmar and W. F. Monet, "Development of the shallow water extension of the bottom loss upgrade," Science Applications Internal Corporation, Vienna, VA, Report No. SAIC-94/1013, 1994.
- ¹⁸B. R. Gomes, R. A. Fisher, and J. K. Fulford, "Environmental characterization for the Littoral Warfare Advance Development 99-3 Experiment," Naval Research Laboratory, Bay St. Louis, MS, Report No. NRL/MR/7180-00-8243, 2000.

Restless rays, steady wave fronts

Oleg A. Godin^{a)}

CIRES, University of Colorado and NOAA/Earth System Research Laboratory, DSRC, Mail Code R/PSD99, 325 Broadway, Boulder, Colorado 80305-3328

(Received 1 March 2007; revised 18 September 2007; accepted 18 September 2007)

Observations of underwater acoustic fields with vertical line arrays and numerical simulations of long-range sound propagation in an ocean perturbed by internal gravity waves indicate that acoustic wave fronts are much more stable than the rays comprising these wave fronts. This paper provides a theoretical explanation of the phenomenon of wave front stability in a medium with weak sound-speed perturbations. It is shown analytically that at propagation ranges that are large compared to the correlation length of the sound-speed perturbations but smaller than ranges at which ray chaos develops, end points of rays launched from a point source and having a given travel time are scattered primarily along the wave front corresponding to the same travel time in the unperturbed environment. The ratio of root mean square displacements of the ray end points along and across the unperturbed wave front increases with range as the ratio of ray length to correlation length of environmental perturbations. An intuitive physical explanation of the theoretical results is proposed. The relative stability of wave fronts compared to rays is shown to follow from Fermat's principle and dimensional considerations. © 2007 Acoustical Society of America.
[DOI: 10.1121/1.2799479]

PACS number(s): 43.30.Re, 43.20.Dk, 43.30.Cq [JAC]

Pages: 3353–3363

I. INTRODUCTION

It has been established reliably in field experiments^{1–4} and confirmed by extensive numerical simulations^{5,6} that early arrivals of acoustic waves at long-range propagation in a deep ocean are stable and can be identified with specific ray paths despite strong perturbations of the ray paths due to sound-speed fluctuations primarily induced by internal gravity waves. Numerical ray simulations^{5–9} of sound propagation in underwater acoustic waveguides show that acoustic wave fronts (and, consequently, time fronts) remain relatively stable and predictable at ranges where rays experience very strong perturbations compared to rays with the same launch angles or same launch and arrival points in a waveguide without internal-wave-induced sound-speed perturbations. To our knowledge, the first observation that strong perturbations of ray trajectories do not necessarily result in strong perturbations of signal arrival times was made by Palmer *et al.*¹⁰

There is an apparent contradiction between the stability of wave fronts and strong perturbations of the rays that comprise these wave fronts. A similar apparent contradiction has been noted in simulation studies of elastic wave propagation in the Earth's crust,¹¹ where it was observed that travel times obtained by tracing a finite number of rays exhibit rapid spatial variation, which is not present when a wave front is found by a finite-difference solution of the eikonal equation. Partial explanations of the apparent contradiction proposed in the literature include “wave front healing”¹¹ due to diffraction and clustering of chaotic rays (see reviews in Refs. 7 and 12). The clusters occur at propagation ranges, which are large compared to b^{-1} , where b is the Lyapunov exponent.

Typical values of b^{-1} are 100–300 km in deep water (Ref. 12, p. 68). In this paper, we offer an alternative explanation that does not invoke diffraction and is valid at ranges much shorter than those necessary for chaotic ray clusters to develop.

Below, we use a version of the ray perturbation theory to provide an analytic explanation of previously observed properties of ray and wave front perturbations and to relate the observed relative stability of wave fronts to the disparity between propagation range and spatial scales of the sound-speed perturbations. Within the ray approximation, we consider sound propagation in a medium where weak sound-speed perturbations are superimposed on an arbitrary given background which can be either homogeneous or inhomogeneous in one (layered), two (range-dependent) or three (horizontally inhomogeneous) dimensions. The environment is assumed to be time independent and quiescent. Sound-speed perturbations can be either random or deterministic. We will demonstrate that, at long propagation ranges, perturbations of a wave front are much weaker than perturbations of individual rays forming the wavefront (Fig. 1). Similar issues assuming a homogeneous background have been considered recently by Flatté and Colosi¹³ in a conference presentation. Our approach is based on an earlier work by Godin.¹⁴ In the underwater waveguide in a deep ocean, this theoretical approach applies at propagation ranges up to $O(b^{-1})$, or up to a few hundred kilometers.

The remainder of the paper is organized as follows. Section II contains basic equations for ray trajectories and their perturbations. Displacement of ray end points in a direction normal to an unperturbed wave front is discussed in Sec. III. Displacement of ray end points in the directions along the unperturbed wave front is discussed in Sec. IV for uniform, range-dependent, and three-dimensionally inhomogeneous

^{a)}Electronic mail: Oleg.Godin@noaa.gov

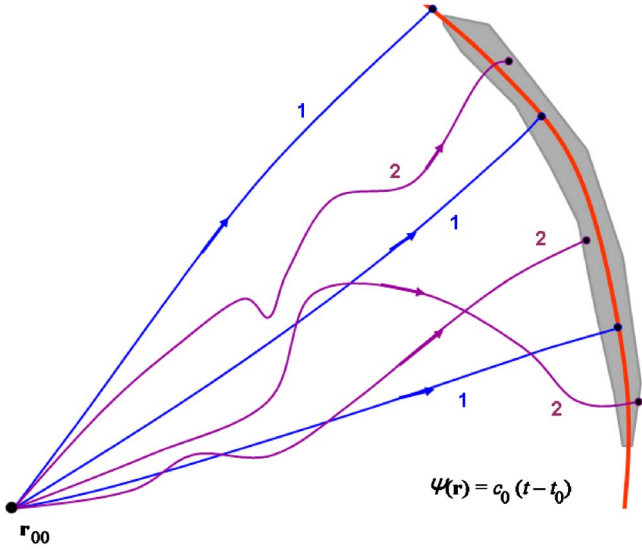


FIG. 1. (Color online) Anisotropy of ray scattering at long-range propagation in an inhomogeneous medium. Perturbed rays (2) can deviate far from unperturbed rays (1) launched from the same point source, but their end points at any given time t are located in the vicinity (shaded) of an unperturbed wave front (3).

unperturbed (background) sound-speed fields. An important special case of small-scale random environmental inhomogeneities is considered in Sec. V. A physical interpretation of the results is given in Sec. VI. Section VII offers conclusions of this work.

II. DIFFERENTIAL RAY EQUATIONS

In the geometric acoustics (ray) approximation, acoustic pressure is the sum of terms of the kind $P(\mathbf{r}, \omega) = A(\mathbf{r}, \omega) \exp(i\omega c_0^{-1} \Psi(\mathbf{r}))$, where \mathbf{r} is a position vector, ω is wave frequency, c_0 is the reference sound speed, A is the ray amplitude, and Ψ is the eikonal. The quantity $\mathbf{p} = \nabla \Psi$ has the meaning of dimensionless wave vector; $\mathbf{p}^2 = n^2$, $n = c_0/c(\mathbf{r})$, where c and n are sound speed and refraction index. An acoustic wave emitted at a point \mathbf{r}_0 at a moment t_0 arrives at point \mathbf{r} at the moment $t = t_0 + [\Psi(\mathbf{r}) - \Psi(\mathbf{r}_0)]/c_0$ (Fig. 1). Rays obey first-order differential equations^{15,16}

$$d\mathbf{r}/d\tau = \mathbf{p}, \quad d\mathbf{p}/d\tau = n \nabla n, \quad (1)$$

where parameter τ determines the position of the point along the ray. Study of wave fronts is significantly simplified by choosing time as an independent variable. We recast Eq. (1) as

$$d\mathbf{r}/dt = c_0 \mathbf{p}/n^2, \quad d\mathbf{p}/dt = c_0 (\nabla n)/n \quad (2)$$

or as a single second-order equation

$$\frac{d}{dt} \left(n^2 \frac{d\mathbf{r}}{dt} \right) = c_0^2 \frac{\nabla n}{n} \quad (3)$$

for the ray trajectory. Note that according to Eq. (2) the eikonal equation $(\nabla \Psi)^2 = n^2$ can be written as

$$n^2 (d\mathbf{r}/dt)^2 = c_0^2. \quad (4)$$

To specify an individual ray, Eq. (3) needs to be supplemented by appropriate initial (or boundary) conditions. In

view of Eq. (3), three-dimensional (3D) rays launched from a point source at a point \mathbf{r}_{00} can be uniquely specified by their launch time t_0 and two parameters $\gamma_{1,2}$, which determine launch direction. To simplify notation, we will write the ray trajectory as $\mathbf{r} = \mathbf{r}(t)$ omitting parameters \mathbf{r}_{00} , t_0 , and $\gamma_{1,2}$ for brevity, where it cannot lead to confusion.

For a derivative of the ray trajectory with respect to a parameter η in the initial or boundary conditions specifying the ray, we obtain

$$c_0^{-2} \frac{d}{dt} \left[\frac{d\mathbf{r}}{dt} \left(\frac{\partial \mathbf{r}}{\partial \eta} \cdot \nabla \right) n^2 + n^2 \frac{d}{dt} \frac{\partial \mathbf{r}}{\partial \eta} \right] - \left(\frac{\partial \mathbf{r}}{\partial \eta} \cdot \nabla \right) \frac{\nabla n}{n} = 0 \quad (5)$$

by differentiating Eq. (3) and taking into account that $n(\mathbf{r}(t))$ depends on η only through dependence of the ray trajectory $\mathbf{r}(t)$ on η . Here, η can be, for instance, t_0 , $\gamma_{1,2}$, or any component of \mathbf{r}_{00} . Moreover, $d\mathbf{r}/dt$ also satisfies Eq. (5) as long as the environment is time independent. Note that Eq. (5) for $\partial \mathbf{r}/\partial \eta$ is linear unlike Eq. (3) for $\mathbf{r}(t)$. We will use the following property of the ray trajectory derivative with respect to the parameter:

$$\frac{d}{dt} \left(n^2 \frac{d\mathbf{r}}{dt} \cdot \frac{\partial \mathbf{r}}{\partial \eta} \right) = 0, \quad (6)$$

which can be viewed as an extension of Eq. (4). To prove Eq. (6), it suffices to note that

$$c_0^{-2} \frac{d\mathbf{r}}{dt} \cdot \frac{d}{dt} \frac{\partial \mathbf{r}}{\partial \eta} = \frac{1}{2} \frac{\partial}{\partial \eta} n^{-2}(\mathbf{r}(t))$$

according to Eq. (4) and to use the ray differential equation (3).

Consider now perturbations in a ray trajectory resulting from a small perturbation in the sound speed. Let $n(\mathbf{r}) = n_0(\mathbf{r}) + \varepsilon n_1(\mathbf{r}) + \varepsilon^2 n_2(\mathbf{r}) + \dots$, $0 < \varepsilon \leq 1$. An unperturbed (background) medium is characterized by the refraction index $n_0(\mathbf{r})$, whereas $\varepsilon n_1(\mathbf{r})$ and higher-order terms describe sound-speed perturbation. We represent the ray trajectory in the perturbed environment as a power series in the small parameter ε :

$$\mathbf{r}(t) = \mathbf{r}_0(t) + \varepsilon \mathbf{r}_1(t) + \varepsilon^2 \mathbf{r}_2(t) + \dots, \quad (7)$$

where the ray trajectory $\mathbf{r}_0(t)$ in the unperturbed environment is assumed known; it satisfies Eq. (3) with n replaced by $n_0(\mathbf{r})$. Our task is to determine first-order perturbations $\mathbf{r}_1(t)$ in ray trajectories caused by small sound-speed perturbations. To justify development of $n_0(\mathbf{r}(t))$ into a series in powers of ε , we assume that $|\mathbf{r}(t) - \mathbf{r}_0(t)| \ll L_0, L_1$ where L_0 and L_1 are representative spatial scales of the background sound-speed inhomogeneities and sound-speed fluctuations, respectively. Then, substituting Eq. (7) into ray equation (3) and equating terms with the same power of ε , we obtain

$$c_0^{-2} \frac{d}{dt} \left[\frac{d\mathbf{r}_0}{dt} (\mathbf{r}_1 \cdot \nabla) n_0^2 + n_0^2 \frac{d\mathbf{r}_1}{dt} \right] - (\mathbf{r}_1 \cdot \nabla) \frac{\nabla n_0}{n_0} = \frac{\nabla n_1}{n_0} - 3n_1 \frac{\nabla n_0}{n_0^2} - 2 \frac{n_0^2}{c_0^2} \frac{d\mathbf{r}_0}{dt} \left(\frac{d\mathbf{r}_0}{dt} \cdot \nabla \right) \frac{n_1}{n_0}. \quad (8)$$

Note that Eq. (8) is linear and differs from Eq. (5) for $\partial \mathbf{r}_0/\partial \eta$ only by having a nonzero right-hand side. Therefore, as it is

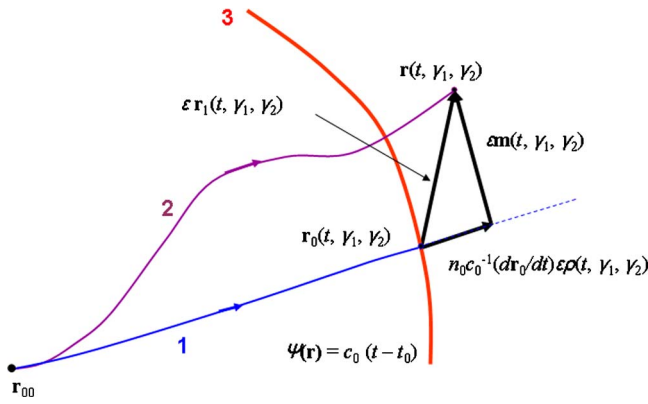


FIG. 2. (Color online) Ray displacement by environmental perturbations. Rays launched in the same direction from a source at the point \mathbf{r}_{00} in the background (1) and perturbed (2) environments arrive at points, $\mathbf{r}_0(t, \gamma_1, \gamma_2)$ and $\mathbf{r}(t, \gamma_1, \gamma_2)$, respectively, at time t . Ray displacement $\mathbf{r}(t, \gamma_1, \gamma_2) - \mathbf{r}_0(t, \gamma_1, \gamma_2) = \epsilon \mathbf{r}_1(t, \gamma_1, \gamma_2) + O(\epsilon^2)$ has components $\epsilon \mathbf{m}(t, \gamma_1, \gamma_2) + O(\epsilon^2)$ along and $\epsilon [n_0/c_0 (d\mathbf{r}_0/dt)] \rho(t, \gamma_1, \gamma_2) + O(\epsilon^2)$ across the unperturbed wave front (3).

well known in classical mechanics and geometrical optics,¹⁶ trajectory perturbations can be expressed in terms of derivatives of the unperturbed trajectory with respect to parameters. The second-order differential equation (8) needs to be supplemented by initial conditions. When comparing rays launched from the same point and in the same direction in unperturbed and perturbed media, we have

$$\mathbf{r}_1(t_0) = 0, \quad \frac{d\mathbf{r}_1}{dt}(t_0) = - \left(\frac{n_1}{n_0} \frac{d\mathbf{r}_0}{dt} \right)_{t=t_0}, \quad (9)$$

according to Eq. (4).

III. NORMAL DISPLACEMENT

A wave front in a perturbed medium is displaced relative to a wave front with the same value of t in the background medium. Consider first the component of the ray trajectory perturbation in the direction normal to the unperturbed wave front at a given point. The perturbed wave front can be obtained from the unperturbed wavefront with the same value of t by adding the normal component of the ray trajectory perturbation. The normal to the unperturbed wave front is parallel to an unperturbed ray, the direction of which is given by $d\mathbf{r}_0/dt$ (Fig. 2). Therefore, the normal component, ρ , of the ray trajectory perturbation [and, hence, the wave front displacement at point $\mathbf{r}_0(t)$] is

$$\rho(t) = \frac{n_0}{c_0} \frac{d\mathbf{r}_0}{dt} \cdot \mathbf{r}_1. \quad (10)$$

Although \mathbf{r}_1 satisfies the second-order equation (8), a simpler, first-order equation can be obtained for ρ using Eq. (4), which can be viewed as a ‘‘conservation law,’’ or a first integral of Eq. (3). Indeed, it follows from Eq. (4) that

$$c_0^{-2} \frac{d\mathbf{r}_0}{dt} \cdot \frac{d\mathbf{r}_1}{dt} = - \frac{n_1 + \mathbf{r}_1 \cdot \nabla n_0}{n_0^3}. \quad (11)$$

Equation (11) can be viewed as a first integral of Eq. (8). From Eqs. (10), (11), and (3) we find that

$$c_0^{-1} \frac{d}{dt}(n_0 \rho) = - \frac{n_1}{n_0}. \quad (12)$$

From Eqs. (9) and (12) it follows that

$$\rho(t) = \frac{-c_0}{n_0(\mathbf{r}_0(t))} \int_{t_0}^t dt_1 \frac{n_1(\mathbf{r}_0(t_1))}{n_0(\mathbf{r}_0(t_1))}. \quad (13)$$

Compare this result to the perturbation of the acoustic wave travel time T between points $\mathbf{r}_{00} \equiv \mathbf{r}_0(t_0)$ and $\mathbf{r}_0(t)$. The travel time $T = T_0 + \epsilon T_1 + \epsilon^2 T_2 + \dots$, where T_0 is the travel time in the background medium, and the first-order travel time perturbation is¹⁶

$$T_1 = \int_{t_0}^t dt_1 \frac{n_1(\mathbf{r}_0(t_1))}{n_0(\mathbf{r}_0(t_1))}. \quad (14)$$

Hence, we obtain the relation $\rho(t) = -c_0 T_1 / n_0(\mathbf{r}_0(t))$, which could have been anticipated from physical considerations.

Let us write the first-order trajectory perturbation $\mathbf{r}_1(t)$, which is also called ray displacement, as

$$\mathbf{r}_1 = \mathbf{m} + \rho n_0 c_0^{-1} d\mathbf{r}_0/dt. \quad (15)$$

It follows from Eqs. (10) and (15) that $\mathbf{m} \cdot d\mathbf{r}_0/dt = 0$. Hence, \mathbf{m} has the meaning of the ray displacement along the unperturbed wave front (Fig. 2). Substituting Eq. (15) into Eqs. (8) and (9) and using Eq. (12) for ρ and Eq. (3) for $\mathbf{r}_0(t)$, we obtain the following initial-value problem for \mathbf{m} :

$$c_0^{-2} \frac{d}{dt} \left[\frac{d\mathbf{r}_0}{dt} (\mathbf{m} \cdot \nabla) n_0^2 + n_0^2 \frac{d\mathbf{m}}{dt} \right] - (\mathbf{m} \cdot \nabla) \frac{\nabla n_0}{n_0} = \nabla \left(\frac{n_1}{n_0} \right) - \frac{n_0^2}{c_0^2} \frac{d\mathbf{r}_0}{dt} \left(\frac{d\mathbf{r}_0}{dt} \cdot \nabla \right) \frac{n_1}{n_0}, \quad (16)$$

$$\mathbf{m}(t_0) = 0, \quad \frac{d\mathbf{m}}{dt}(t_0) = 0. \quad (17)$$

IV. TANGENTIAL DISPLACEMENT

A. Uniform background

Before calculating along the front components of the ray displacement in a general case, consider a simple case when the background is uniform ($n_0 = \text{const.}$) and, therefore, there is no regular refraction. In this case, according to Eqs. (3) and (4), $n_0 d\mathbf{r}_0/dt$ is a constant unit vector, and $\mathbf{r}_0(t) = \mathbf{r}_{00} + (t - t_0) d\mathbf{r}_0/dt$. From Eqs. (8) and (9) we find the ray trajectory perturbation:

$$\mathbf{r}_1(t) = - \frac{n_1(\mathbf{r}_{00})}{n_0 c_0} \frac{d\mathbf{r}_0}{dt} (t - t_0) + \frac{1}{n_0^3} \int_{t_0}^t dt_1 (t - t_1) \left[c_0^2 \nabla n_1 - 2n_0^2 \frac{d\mathbf{r}_0}{dt} \left(\frac{d\mathbf{r}_0}{dt} \cdot \nabla n_1 \right) \right]_{\mathbf{r}=\mathbf{r}_0(t_1)}. \quad (18)$$

This result is by no means new (see, e.g., Ref. 17). It is reproduced here to illustrate anisotropy of the ray displacement in the simplest setting possible. When calculating dot product of Eq. (18) with $n_0 d\mathbf{r}_0/dt$ in order to find the normal component of the ray trajectory perturbation, a full derivative

$dn_1(\mathbf{r}_0(t_1))/dt_1$ appears in the integrand, and after integration by parts one finds that

$$\rho(t) = -c_0 n_0^{-2} \int_{t_0}^t dt_1 n_1(\mathbf{r}_0(t_1)) \quad (19)$$

in agreement with Eq. (13). For the component of ray trajectory perturbation in the direction of a constant unit vector \mathbf{s} orthogonal to the vector $d\mathbf{r}_0/dt$, one finds from Eq. (18) that

$$\mathbf{s} \cdot \mathbf{r}_1(t) = c_0^2 n_0^{-3} \int_{t_0}^t dt_1 (t-t_1) (\mathbf{s} \cdot \nabla n_1)_{\mathbf{r}=\mathbf{r}_0(t_1)}. \quad (20)$$

When n_1 and $\mathbf{s} \cdot \nabla n_1$ do not change their signs, from Eqs. (19) and (20) we find

$$|\rho| \approx n_0^{-2} \|n_1\| c_0 (t-t_0), \quad |\mathbf{s} \cdot \mathbf{r}_1| \approx n_0^{-3} \|n_1\| c_0^2 (t-t_0)^2 / 2L_1, \quad (21)$$

where $\|n_1\|$ and L_1 are a representative value of $|n_1|$ and a spatial scale of n_1 variation, respectively. In the opposite case, where n_1 oscillates around zero and changes its sign many times between points $\mathbf{r}_0(t_0)$ and $\mathbf{r}_0(t)$, we have

$$|\rho| \approx n_0^{-3/2} \|n_1\| \sqrt{L_1 c_0 (t-t_0)}, \quad |\mathbf{s} \cdot \mathbf{r}_1| \approx n_0^{-5/2} \|n_1\| \sqrt{c_0^3 (t-t_0)^3 / 3L_1}. \quad (22)$$

In both cases, the ray trajectory displacement along an unperturbed wave front increases much faster with time and, at large t [i.e., at ray lengths $R=c_0(t-t_0)/n_0$ large compared to L_1], becomes much larger than the ray trajectory displacement across the unperturbed wave front; $|\mathbf{s} \cdot \mathbf{r}_1|/|\rho| \approx R/2L_1$.¹⁸ Below, we show that these observations remain valid in a much more general case.

B. Range-dependent background and horizontal refraction

Introduce an orthogonal Cartesian coordinate system with horizontal coordinates x and y and a vertical coordinate z . Let the background environment be independent of y . To specify the initial direction of a ray launched from a point source, we will use grazing angle χ and azimuthal angle ψ , so that

$$\frac{1}{c_0 n} \frac{d\mathbf{r}}{dt} \Big|_{t=t_0} = (\cos \chi \cos \psi, \cos \chi \sin \psi, \sin \chi). \quad (23)$$

First, consider rays in the xz vertical plane, which correspond to $\psi=0$. Choosing the ray parameter η in Eq. (6) to be the grazing angle χ , we have $\partial\mathbf{r}_0/\partial\chi=0$ at $t=t_0$ and, hence, the unit vector $\mathbf{q}_1=(\partial\mathbf{r}_0/\partial\chi)/|\partial\mathbf{r}_0/\partial\chi|$ is orthogonal to $d\mathbf{r}_0/dt$ at $t>t_0$. Both vectors, as well as $\mathbf{r}_0(t)$, lie within the xz plane. ($\partial\mathbf{r}_0/\partial\chi=0$ and \mathbf{q}_1 becomes indeterminate on caustics. It is assumed that \mathbf{q}_1 is extended along the ray through each caustic's contact point as a continuous function of t .) When sound-speed perturbations are independent of y , ray trajectory perturbations have no y -component, and the ray displacement \mathbf{m} along the unperturbed wave front can be represented as

$$\mathbf{m} = \beta(t) \mathbf{q}_1(t). \quad (24)$$

Substituting \mathbf{m} as given by Eq. (24) into Eqs. (16) and (17) and using Eq. (5) for $\partial\mathbf{r}_0/\partial\chi$, we obtain a scalar initial-value problem for the unknown function β :

$$\begin{aligned} n_0^2 \frac{d^2 \beta}{dt^2} + \frac{dn_0^2}{dt} \frac{d\beta}{dt} - \frac{\beta}{|\partial\mathbf{r}_0/\partial\chi|} \frac{d}{dt} \left(n_0^2 \frac{d}{dt} \left| \frac{\partial\mathbf{r}_0}{\partial\chi} \right| \right) \\ = c_0^{-2} (\mathbf{q}_1 \cdot \nabla) \frac{n_1}{n_0}; \\ \beta(t_0) = \frac{d\beta}{dt}(t_0) = 0. \end{aligned} \quad (25)$$

The solution of the problem (25) can be expressed in terms of known solutions of the homogeneous second-order ordinary differential equation in Eq. (25). One solution of the homogeneous equation, $\beta_1(t) \equiv |\partial\mathbf{r}_0/\partial\chi|$, corresponds to the known solution, $\partial\mathbf{r}_0/\partial\chi$, of Eq. (5) (with $n=n_0$). A linearly independent solution to the homogeneous equation, $\beta_2(t) \equiv \mathbf{q}_1 \cdot \partial\mathbf{r}_0/\partial b$, is obtained by choosing parameter η of the ray trajectory as a magnitude b of the displacement $b\mathbf{q}_1(t_0)$ of the source position within the xz plane in a direction $\mathbf{q}_1(t_0) = (-\sin \chi, 0, \cos \chi)$ perpendicular to $(d\mathbf{r}_0/dt)|_{t=t_0}$. Linear independence of $\beta_1(t)$ and $\beta_2(t)$ follows from their values at the initial moment: $\beta_1(t_0)=0$, $\beta_2(t_0)=1$.

Using the variation of constants method, the solution to Eq. (25) can be expressed in terms of $\beta_1(t)$ as

$$\begin{aligned} \beta(t) = c_0^2 \beta_1(t) \int_{t_0}^t \frac{dt_2}{n_0^2(\mathbf{r}_0(t_2)) \beta_1^2(t_2)} \int_{t_0}^{t_2} dt_1 \beta_1(t_1) [\mathbf{q}_1(t_1) \cdot \nabla] \\ \times \frac{n_1}{n_0} \Big|_{\mathbf{r}=\mathbf{r}_0(t_1)} \end{aligned} \quad (26)$$

or in terms of $\beta_1(t)$ and $\beta_2(t)$ as

$$\begin{aligned} \beta(t) = \frac{c_0}{n_0(\mathbf{r}_{00})} \int_{t_0}^t dt_1 [\beta_1(t) \beta_2(t_1) - \beta_1(t_1) \beta_2(t)] \\ \times [\mathbf{q}_1(t_1) \cdot \nabla] \frac{n_1}{n_0} \Big|_{\mathbf{r}=\mathbf{r}_0(t_1)}. \end{aligned} \quad (27)$$

Equation (26) shows that an environmental perturbation results in a ray displacement along the unperturbed wave front which is equal to the displacement resulting from a small change, $\varepsilon \delta\chi$, where $\delta\chi=\beta(t)/\beta_1(t)$, in the ray launch angle in the unperturbed environment. The right-hand side of Eq. (26) diverges after the ray passes a caustic, and this equation is applicable only between the source and the first contact of a given unperturbed ray with a caustic in the background medium. Equation (27) does not have such limitations and describes ray displacement along the unperturbed wave front for rays that have an arbitrary number of contacts with caustics. An inspection shows that Eqs. (26) and (27) are equivalent in their common domain of validity. Equation (27) suggests a somewhat different interpretation of the ray displacement. According to Eq. (27), environmental-perturbation-induced ray displacement along the unperturbed

wave front is equivalent to the combination of variation $\varepsilon\chi_1$ of the launch angle and a shift $\varepsilon b_1\mathbf{q}_1$ of the source position, where

$$\chi_1(t) = \frac{c_0}{n_0(\mathbf{r}_{00})} \int_{t_0}^t dt_1 \beta_2(t_1) [\mathbf{q}_1(t_1) \cdot \nabla] \frac{n_1}{n_0} \Big|_{\mathbf{r}=\mathbf{r}_0(t_1)},$$

$$b_1(t) = - \frac{c_0}{n_0(\mathbf{r}_{00})} \int_{t_0}^t dt_1 \beta_1(t_1) [\mathbf{q}_1(t_1) \cdot \nabla] \frac{n_1}{n_0} \Big|_{\mathbf{r}=\mathbf{r}_0(t_1)}. \quad (28)$$

Generally, splitting of $\beta(t)$ into equivalent contributions due to change in launch angle and source position is nonunique, as illustrated by Eqs. (26) and (28). However, the ambiguity vanishes on caustics. On a caustic of the rays emitted by a point source $\beta_1(t)=0$. Then, the displacement of a ray touching a caustic in the unperturbed medium along the wave front due to environmental perturbation is equivalent to a displacement of the sound source in the background medium by $b_1\mathbf{q}_1(t_0)$ without any change in the launch angle.

As long as $\mathbf{r}(t) - \mathbf{r}(t_0) = c_0 O(t - t_0)$, we have $\beta_1(t) = c_0 O(t - t_0)$, and consequently, the quantity $\beta_1(t)\beta_2(t_1) - \beta_1(t_1)\beta_2(t) = c_0 O(t - t_1)$ in the integrand in Eq. (27). Comparing Eqs. (27) and (19) for the ray displacement along and across the unperturbed wave front, we see that, as in the special case of uniform background, the ratio $|\mathbf{m}|/|\rho| \sim R/\min(L_0, L_1)$ becomes large at ray lengths large compared to the spatial scale of the refraction index variation.

We now abandon the assumption that n_1 is independent of y . Then, perturbed rays with $\psi=0$, unlike rays in the background environment, are no longer confined to the xz vertical plane. In an underwater acoustics context, deviation of the rays from the vertical plane is referred to as horizontal refraction. Now, instead of Eq. (24) we have

$$\mathbf{m} = \alpha(t)\mathbf{q}_2(t) + \beta(t)\mathbf{q}_1(t), \quad \mathbf{q}_2 = (\partial\mathbf{r}_0/\partial\psi)/|\partial\mathbf{r}_0/\partial\psi|, \quad \mathbf{q}_1 = (\partial\mathbf{r}_0/\partial\chi)/|\partial\mathbf{r}_0/\partial\chi|. \quad (29)$$

Substituting Eq. (29) into Eqs. (16) and (17) and taking into account that $\mathbf{q}_1 \cdot \mathbf{q}_2 = 0$ (at $\psi=0$), we again obtain Eq. (25) for $\beta(t)$ and an initial value problem for $\alpha(t)$, which differs from Eq. (25) only by substitution of $\alpha(t)$ for $\beta(t)$ and \mathbf{q}_2 for \mathbf{q}_1 . The derivative $\partial\mathbf{r}_0/\partial\psi$ can be found explicitly. Here we are interested only in the case $\psi=0$, where

$$\left(\frac{\partial\mathbf{r}_0}{\partial\psi} \right)_{\psi=0} = n_0(\mathbf{r}_{00}) \cos \chi h(t) \nabla y, \quad h(t) = c_0 \int_{t_0}^t \frac{dt_1}{n_0^2(\mathbf{r}_0(t_1))|_{\psi=0}}. \quad (30)$$

Using Eq. (26) (with the appropriate substitutions indicated previously) and Eq. (30), after some algebra we find that

$$\alpha(t) = c_0 h(t) \int_{t_0}^t dt_1 \left(\frac{1}{n_0} \frac{\partial n_1}{\partial y} \right)_{\mathbf{r}=\mathbf{r}_0(t_1), \psi=0} \left[1 - \frac{h(t_1)}{h(t)} \right]. \quad (31)$$

Equation (31) means that displacement of the ray from the xz plane due to sound-speed perturbations is equivalent to change of the launch angle in the background medium from

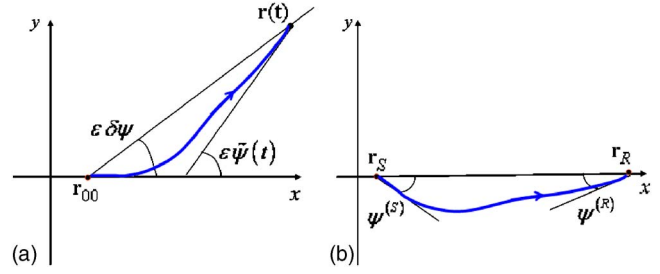


FIG. 3. (Color online) Horizontal refraction angles for (a) a ray launched within the xz plane and (b) an eigenray connecting a source and receiver lying within the xz plane.

the actual value $\psi=0$ to an effective value $\psi = \varepsilon\delta\psi(t)$ [see Fig. 3(a)], where

$$\delta\psi = \left| \frac{\partial\mathbf{r}_0}{\partial\psi} \right| \alpha(t) = \frac{c_0}{n_0(\mathbf{r}_{00}) \cos \chi} \int_{t_0}^t dt_1 \left(\frac{1}{n_0} \frac{\partial n_1}{\partial y} \right)_{\mathbf{r}=\mathbf{r}_0(t_1), \psi=0} \left[1 - \frac{h(t_1)}{h(t)} \right]. \quad (32)$$

This angle is generally different from the angle $\varepsilon\tilde{\psi}(t)$ [Fig. 3(a)] that a horizontal projection of the tangent to the ray makes with Ox coordinate axis. To first order in ε ,

$$\tilde{\psi}(t) = \frac{dy_1}{dt} \Big/ \frac{dx_0}{dt} = \frac{n_0}{c_0 \cos \tilde{\chi}} \frac{d}{dt} \left(\frac{\partial y_0}{\partial \psi} \delta\psi \right), \quad (33)$$

where $\tilde{\chi} = \arctan[(dz_0/dt)/(dx_0/dt)]$ is the current value of the ray grazing angle. From Eqs. (32) and (33) we find that

$$\tilde{\psi}(t) = \frac{c_0}{n_0(\mathbf{r}_0(t)) \cos \tilde{\chi}} \int_{t_0}^t dt_1 \left(\frac{1}{n_0} \frac{\partial n_1}{\partial y} \right)_{\mathbf{r}=\mathbf{r}_0(t_1), \psi=0}. \quad (34)$$

Horizontal refraction of rays has been studied by several authors, see Refs. 19 and 20 and references therein. In Refs. 19 and 20, ray perturbation theory was used to find azimuthal arriving angles $\psi^{(R)}$ [Fig. 3(b)] of eigenrays connecting a source and a receiver located within the xz vertical plane. In our notation, the result obtained in Refs. 19 and 20 becomes

$$\psi^{(R)} = \frac{\varepsilon c_0}{n_0(\mathbf{r}_0(t)) h(t) \cos \tilde{\chi}} \int_{t_0}^t dt_1 \left(\frac{1}{n_0} \frac{\partial n_1}{\partial y} \right)_{\mathbf{r}=\mathbf{r}_0(t_1), \psi=0} h(t_1). \quad (35)$$

Using reversibility of rays, which follows from the reciprocity principle, we obtain from Eq. (35) an expression for the azimuthal launch angle of the eigenray [Fig. 3(b)]:

$$\psi^{(S)} = \frac{-\varepsilon c_0}{n_0(\mathbf{r}_{00}) \cos \chi} \int_{t_0}^t dt_1 \left(\frac{1}{n_0} \frac{\partial n_1}{\partial y} \right)_{\mathbf{r}=\mathbf{r}_0(t_1), \psi=0} \left[1 - \frac{h(t_1)}{h(t)} \right]. \quad (36)$$

Geometrical considerations require that, within the first-order accuracy of our calculations, Fig. 3(b) should be reproduced by rotating Fig. 3(a) by the angle $\varepsilon\delta\psi$ in a clockwise direction, i.e., $\varepsilon\delta\psi = -\psi^{(S)}$. Inspection of Eqs. (32) and (36) shows that this requirement is met, and our results agree with earlier results of Refs. 19 and 20. Moreover, the change of the

y-component of the dimensionless wave vector $\mathbf{p}=\nabla\Psi$ along the perturbed ray, calculated using Eq. (34) as $p_2(t)-p_2(t_0)=\varepsilon\tilde{\psi}(t)n_0(\mathbf{r}_0(t))\cos\tilde{\chi}+O(\varepsilon^2)$ and using Eqs. (35) and (36) as $p_2(t)-p_2(t_0)=n_0(\mathbf{r}_0(t))\cos\tilde{\chi}\sin\psi^{(R)}-n_0(\mathbf{r}_0(t_0))\cos\chi\sin\psi^{(S)}+O(\varepsilon^2)$ are the same to the second order in ε .

Horizontal refraction induced by random²⁰ and deterministic¹⁹ sound-speed perturbations has been previously studied in detail. An inspection of the results^{19,20} shows that the ratio $|y_1(t)|/|\rho(t)|$ of components of ray displacement along and across the unperturbed wave front increases proportionally to $t-t_0$ at large t for both large-scale and small-scale sound-speed perturbations.

Another check of our results for ray displacement along the wave front in the case of a range-dependent background is provided by comparing Eqs. (26) and (31) to the explicit solution (20) in the case of a homogeneous background. When $n_0=\text{const.}$, we have $\beta_2=\text{const.}$, h and β_1 are linear functions of t , and integrals in Eqs. (26), (27), and (31) are easily calculated. An inspection shows that Eqs. (26), (27), and (31) reduce to Eq. (20) for arbitrary perturbations in the case of a homogeneous background.

C. Rays in a generic inhomogeneous medium

Let two parameters $\gamma_{1,2}$ uniquely specify the launch direction of a ray emitted from a point source and

$$\frac{\partial\mathbf{r}_0}{\partial\gamma_1}(t_0)\cdot\frac{\partial\mathbf{r}_0}{\partial\gamma_2}(t_0)=0. \quad (37)$$

(For instance, the angles χ and ψ introduced in Sec. IV B can be used as the parameters $\gamma_{1,2}$.) Let $b_{1,2}$ be magnitudes of source position displacement along unit vectors

$$\mathbf{q}_j=\frac{\partial\mathbf{r}_0}{\partial\gamma_j}\bigg/\left|\frac{\partial\mathbf{r}_0}{\partial\gamma_j}\right|, \quad j=1,2. \quad (38)$$

Designate

$$\mathbf{f}_j(t)=\frac{\partial\mathbf{r}_0}{\partial\gamma_j}, \quad \mathbf{f}_{j+2}(t)=\frac{\partial\mathbf{r}_0}{\partial b_j}, \quad j=1,2. \quad (39)$$

It follows from Eq. (6) that $\mathbf{f}_j(t)\cdot d\mathbf{r}_0/dt=0$ for all four vector functions $\mathbf{f}_j(t)$. These are known functions as long as the ray geometry in the background medium is known. We represent the displacement \mathbf{m} of the perturbed ray along the unperturbed wave front as

$$\mathbf{m}(t)=e_j(t)\mathbf{f}_j(t), \quad (40)$$

where the coefficients e_j are unknown. Here and in the following equations [Eqs. (41)–(43) and (47)], summation over repeated indexes $j=1,2,3,4$ is implied.

Representation (40) is nonunique. It is convenient to impose an additional requirement

$$\mathbf{f}_j de_j/dt=0. \quad (41)$$

The vector \mathbf{m} is orthogonal to $d\mathbf{r}_0/dt$, and therefore Eq. (41) is equivalent to two scalar equations

$$(\mathbf{q}_1\cdot\mathbf{f}_j)de_j/dt=0, \quad (\mathbf{q}_2\cdot\mathbf{f}_j)de_j/dt=0. \quad (42)$$

By substituting Eq. (40) into Eqs. (16) and (17), after using Eq. (41) and Eq. (5) for \mathbf{f}_j and projecting the result on vectors $\mathbf{q}_{1,2}$, we obtain two additional equations:

$$\begin{aligned} n_0^2\frac{de_j}{dt}\left(\mathbf{q}_1\cdot\frac{d\mathbf{f}_j}{dt}\right) &= c_0(\mathbf{q}_1\cdot\nabla)\frac{n_1}{n_0}, & n_0^2\frac{de_j}{dt}\left(\mathbf{q}_2\cdot\frac{d\mathbf{f}_j}{dt}\right) \\ &= c_0(\mathbf{q}_2\cdot\nabla)\frac{n_1}{n_0}, \end{aligned} \quad (43)$$

as well as the initial conditions:

$$e_j(t_0)=0, \quad j=1,2,3,4 \quad (44)$$

for the unknown coefficients e_j . Equations (42) and (43) provide a set of four linear algebraic equations from which time derivatives of the coefficients e_j are found: $de_j/dt=c_0E_j$, where

$$E_j=\frac{1}{n_0^2M}\left[M_{3j}(\mathbf{q}_1\cdot\nabla)\frac{n_1}{n_0}+M_{4j}(\mathbf{q}_2\cdot\nabla)\frac{n_1}{n_0}\right]\bigg|_{\mathbf{r}=\mathbf{r}_0(t)}, \quad (45)$$

$$M=\det\begin{pmatrix} \mathbf{q}_1\cdot\mathbf{f}_1 & \mathbf{q}_1\cdot\mathbf{f}_2 & \mathbf{q}_1\cdot\mathbf{f}_3 & \mathbf{q}_1\cdot\mathbf{f}_4 \\ \mathbf{q}_2\cdot\mathbf{f}_1 & \mathbf{q}_2\cdot\mathbf{f}_2 & \mathbf{q}_2\cdot\mathbf{f}_3 & \mathbf{q}_2\cdot\mathbf{f}_4 \\ \frac{\mathbf{q}_1\cdot d\mathbf{f}_1}{c_0 dt} & \frac{\mathbf{q}_1\cdot d\mathbf{f}_2}{c_0 dt} & \frac{\mathbf{q}_1\cdot d\mathbf{f}_3}{c_0 dt} & \frac{\mathbf{q}_1\cdot d\mathbf{f}_4}{c_0 dt} \\ \frac{\mathbf{q}_2\cdot d\mathbf{f}_1}{c_0 dt} & \frac{\mathbf{q}_2\cdot d\mathbf{f}_2}{c_0 dt} & \frac{\mathbf{q}_2\cdot d\mathbf{f}_3}{c_0 dt} & \frac{\mathbf{q}_2\cdot d\mathbf{f}_4}{c_0 dt} \end{pmatrix}, \quad (46)$$

and M_{kj} is the cofactor to the kj -element in the determinant in the right-hand side of Eq. (46). When parameters $\gamma_{1,2}$ are chosen to be the angles χ and ψ , $d\mathbf{f}_1/dt=c_0n_0^{-1}(\mathbf{r}_{00})\mathbf{q}_1$, $d\mathbf{f}_2/dt=c_0n_0^{-1}(\mathbf{r}_{00})\mathbf{q}_2$ at $t=t_0$, and $M(t_0)=n_0^{-2}(\mathbf{r}_{00})$.

With de_j/dt and the initial conditions (44) known, we finally find that

$$\mathbf{m}(t)=c_0\mathbf{f}_j(t)\int_{t_0}^t E_j(t_1)dt_1. \quad (47)$$

Note that, according to Eqs. (40) and (41) an expression for $d\mathbf{m}/dt$ is obtained from Eq. (47) by substitution of $d\mathbf{f}_j/dt$ for \mathbf{f}_j in front of the integral in the right-hand side of Eq. (47).

The solution obtained simplifies if $\mathbf{f}_1\cdot\mathbf{f}_2=0$ for all t (rather than just at $t=t_0$) and

$$\begin{aligned} \mathbf{f}_1\cdot\mathbf{f}_4=0, \quad \mathbf{f}_2\cdot\mathbf{f}_3=0, \quad \mathbf{f}_1\cdot\frac{d\mathbf{f}_2}{dt}=\mathbf{f}_1\cdot\frac{d\mathbf{f}_4}{dt}=0, \quad \mathbf{f}_2\cdot\frac{d\mathbf{f}_1}{dt} \\ =\mathbf{f}_2\cdot\frac{d\mathbf{f}_3}{dt}=0. \end{aligned}$$

(For instance, all these conditions are met in the special case of range-dependent background considered in Sec. IV B.) Then the matrix in Eq. (46) becomes sparse; half of its elements equal zero. From Eqs. (45) and (46), we find that

$$\begin{aligned}
E_1 &= -c_0 \left[(\mathbf{q}_1 \cdot \mathbf{f}_1) \left(\mathbf{q}_1 \cdot \frac{d\mathbf{f}_3}{dt} \right) - (\mathbf{q}_1 \cdot \mathbf{f}_3) \right. \\
&\quad \left. \times \left(\mathbf{q}_1 \cdot \frac{d\mathbf{f}_1}{dt} \right) \right]^{-1} n_0^{-2} (\mathbf{q}_1 \cdot \mathbf{f}_3) (\mathbf{q}_1 \cdot \nabla) \frac{n_1}{n_0}, \\
E_2 &= -c_0 \left[(\mathbf{q}_2 \cdot \mathbf{f}_2) \left(\mathbf{q}_2 \cdot \frac{d\mathbf{f}_4}{dt} \right) - (\mathbf{q}_2 \cdot \mathbf{f}_4) \right. \\
&\quad \left. \times \left(\mathbf{q}_2 \cdot \frac{d\mathbf{f}_2}{dt} \right) \right]^{-1} n_0^{-2} (\mathbf{q}_2 \cdot \mathbf{f}_4) (\mathbf{q}_2 \cdot \nabla) \frac{n_1}{n_0}, \\
E_3 &= c_0 \left[(\mathbf{q}_1 \cdot \mathbf{f}_1) \left(\mathbf{q}_1 \cdot \frac{d\mathbf{f}_3}{dt} \right) - (\mathbf{q}_1 \cdot \mathbf{f}_3) \right. \\
&\quad \left. \times \left(\mathbf{q}_1 \cdot \frac{d\mathbf{f}_1}{dt} \right) \right]^{-1} n_0^{-2} (\mathbf{q}_1 \cdot \mathbf{f}_1) (\mathbf{q}_1 \cdot \nabla) \frac{n_1}{n_0}, \\
E_4 &= c_0 \left[(\mathbf{q}_2 \cdot \mathbf{f}_2) \left(\mathbf{q}_2 \cdot \frac{d\mathbf{f}_4}{dt} \right) - (\mathbf{q}_2 \cdot \mathbf{f}_4) \right. \\
&\quad \left. \times \left(\mathbf{q}_2 \cdot \frac{d\mathbf{f}_2}{dt} \right) \right]^{-1} n_0^{-2} (\mathbf{q}_2 \cdot \mathbf{f}_2) (\mathbf{q}_2 \cdot \nabla) \frac{n_1}{n_0}. \quad (48)
\end{aligned}$$

To compare Eqs. (47) and (48) for the ray displacement along the unperturbed wave front with the results obtained in Sec. IV B, we choose $\gamma_1 = \chi$, $\gamma_2 = \psi$. Then the functions $\beta_{1,2}(t)$, introduced in Sec. IV B, and functions \mathbf{f}_j are related as follows: $\beta_1 = \mathbf{q}_1 \cdot \mathbf{f}_1$, $\beta_2 = \mathbf{q}_1 \cdot \mathbf{f}_3$, and

$$\begin{aligned}
(\mathbf{q}_1 \cdot \mathbf{f}_1) \left(\mathbf{q}_1 \cdot \frac{d\mathbf{f}_3}{dt} \right) - (\mathbf{q}_1 \cdot \mathbf{f}_3) \left(\mathbf{q}_1 \cdot \frac{d\mathbf{f}_1}{dt} \right) &= \beta_1 \frac{d\beta_2}{dt} \\
- \beta_2 \frac{d\beta_1}{dt} &= -c_0 \frac{n_0(\mathbf{r}_0(t_0))}{n_0^2(\mathbf{r}_0(t))} \quad (49)
\end{aligned}$$

as a Wronskian of solutions to homogeneous Eq. (25). Now it is easy to see that the \mathbf{q}_1 component of the ray displacement \mathbf{m} Eq. (47) equals $\beta(t)$ Eq. (27). Similarly, one establishes agreement between the \mathbf{q}_2 component of the ray displacement \mathbf{m} Eq. (47) and $\alpha(t)$ Eq. (31).

For a generic inhomogeneous background, at large t we have $\mathbf{r}_0(t) - \mathbf{r}_0(t_0) = O(t - t_0)$, $d\mathbf{r}_0/dt = O(1)$ and $\mathbf{f}_{1,2}(t) = O(t - t_0)$, $\mathbf{f}_{3,4}(t) = O((t - t_0)^{\kappa_{3,4}})$, $d\mathbf{f}_j/dt = O(1)$, $\kappa_k = 0$ or 1 . Then $M(t) = O((t - t_0)^2)$, $M_{kj}(t) = O((t - t_0)^2)$, where $k = 3, 4$ and $j = 1, 2, 3, 4$. According to Eq. (45), $E_j(t) = O((t - t_0)^0 / \min(L_0, L_1))$. Comparison of Eqs. (47) and (13) then shows that $|\mathbf{m}|/|\rho| \sim R / \min(L_0, L_1)$, where, as before, R , L_0 , and L_1 designate ray length and representative spatial scales of $n_0(\mathbf{r})$ and $n_1(\mathbf{r})$ variation.

V. SMALL-SCALE RANDOM INHOMOGENEITIES

The ray perturbation theory as used earlier involves an approximation of the ray trajectory in the argument of refraction index $n(\mathbf{r})$ by the first terms of the power series (7) with the subsequent development of $n(\mathbf{r}_0(t) + \varepsilon \mathbf{r}_1(t) + \varepsilon^2 \mathbf{r}_2(t) + \dots)$ in powers of ε . To justify this procedure, one has to assume that $\varepsilon \mathbf{r}_1(t)$ is small compared to spatial scales L_0 and L_1 of the background sound-speed inhomogeneities and sound-speed perturbations. In many applications, including

underwater sound propagation, perturbations have much smaller correlation length than the spatial scales of the background environment variation. In the case of small-scale random perturbations superimposed on a large-scale background, where $L_0 \gg L_1$, the validity domain of the perturbation theory can be greatly extended by the following modification of the above analysis.

We represent the value of the refraction index on the perturbed trajectory as

$$\begin{aligned}
n(\mathbf{r}(t)) &= n_0(\mathbf{r}_0(t)) + \varepsilon [\mathbf{r}_1(t) \cdot \nabla] n_0(\mathbf{r}_0(t)) + \varepsilon n_1(\mathbf{r}(t)) \\
&\quad + O(\varepsilon^2) \quad (50)
\end{aligned}$$

without expanding $n_1(\mathbf{r}(t))$ into a power series. Then Eqs. (8), (9), (12), and (16) and their corollaries remain unchanged, except n_1 should be understood as $n_1(\mathbf{r}(t))$. The advantage of such a representation is that it is valid when $\varepsilon |\mathbf{r}_1| \ll L_0$; the requirement $\varepsilon |\mathbf{r}_1| \ll L_1$ no longer applies.

For deterministic inhomogeneities, Eqs. (13) and (18) and others do not give an explicit solution for ray trajectory perturbations when $n_1 = n_1(\mathbf{r}(t))$. Explicit solutions, however, can be derived for statistical moments of ray trajectory perturbations induced by random environmental perturbations. Let us assume that the refraction index perturbations are locally statistically homogeneous, have zero mean: $\langle n_1(\mathbf{r}) \rangle = 0$, and are characterized by a differentiable correlation function

$$\langle n_1(\mathbf{r}_A) n_1(\mathbf{r}_B) \rangle = C \left(\mathbf{r}_A - \mathbf{r}_B, \frac{\mathbf{r}_A + \mathbf{r}_B}{2} \right). \quad (51)$$

Here and in the following angular brackets denote an ensemble average. For statistically homogeneous perturbations, the correlation function and other two-point statistical moments are functions of the position difference $\mathbf{r}_A - \mathbf{r}_B$ only. For locally statistically homogeneous perturbations, spatial scale L_2 of the correlation function variation with respect to the second argument, $(\mathbf{r}_A + \mathbf{r}_B)/2$, is much larger than with respect to the first argument, $\mathbf{r}_A - \mathbf{r}_B$: $L_2 \gg L_1$.

For the across wave front component of the random ray displacement, we have from Eqs. (13) and (51) $\langle \rho(t) \rangle = 0$ and

$$\begin{aligned}
\langle \rho^2(t) \rangle &= \frac{c_0^2}{n_0^2(\mathbf{r}_0(t))} \int \int_{t_0}^t \frac{dt_1 dt_2}{n_0(\mathbf{r}_0(t_1)) n_0(\mathbf{r}_0(t_2))} C \left(\mathbf{r}_0(t_1) \right. \\
&\quad \left. - \mathbf{r}_0(t_2), \frac{\mathbf{r}_0(t_1) + \mathbf{r}_0(t_2)}{2} \right). \quad (52)
\end{aligned}$$

In derivation of Eq. (52) we assumed that the ray trajectory perturbation is small compared to L_2 : $|\varepsilon \mathbf{r}_1(t)| \ll L_2$. We also assumed that the ray trajectory perturbation is small compared to L_1 over propagation distances of the order of L_1 , i.e., $\varepsilon |\mathbf{r}_1(t_1) - \mathbf{r}_1(t_2)| \ll L_1$ when $|t_1 - t_2| \leq c_0^{-1} n_0 L_1$. As in Ref. 14, for small-scale inhomogeneities the double integral in Eq. (52) can be reduced to a single integral along an unperturbed ray. Taking into account that correlation length is small compared to spatial scales of n_0 and $d\mathbf{r}_0/dt$ variation, we change variables of integration in Eq. (52) to $(t_1 - t_2)$ and $(t_1 + t_2)/2$, neglect terms of the order L_1/L_2 and L_1/L_0 , and after some algebra obtain

$$\langle \rho^2(t) \rangle = \frac{2c_0}{n_0^2(\mathbf{r}_0(t))} \int_{t_0}^t \frac{dt_1}{n_0(\mathbf{r}_0(t_1))} D\left(\frac{n_0(\mathbf{r}_0(t_1))}{c_0} \frac{d\mathbf{r}_0}{dt_1}; \mathbf{r}_0(t_1)\right), \quad (53)$$

where

$$D(\mathbf{s}; \mathbf{r}) = \int_0^{+\infty} C(\mathbf{s}l, \mathbf{r}) dl, \quad \mathbf{s}^2 = 1. \quad (54)$$

The quantity D is an integral characteristic of environmental perturbations; $D(\mathbf{s}; \mathbf{r})/\langle n_1^2(\mathbf{r}) \rangle$ has the meaning of the integral scale of refraction index perturbations around point \mathbf{r} in the direction defined by the unit vector \mathbf{s} . In the case of isotropic sound-speed fluctuations, D is independent of \mathbf{s} . According to Eq. (53), $\langle \rho^2(t) \rangle = 2n_0^{-2}(\mathbf{r}_0(t))R(t)\bar{D}$, where \bar{D} is the value of D spatially averaged along the unperturbed ray and $R(t)$ is the length of the unperturbed ray.

Quite similarly, from general equations (45) and (47) (with $n_1 = n_1(\mathbf{r}(t))$) for along the wave front ray displacement \mathbf{m} , we find that $\langle \mathbf{m} \rangle = 0$ and

$$\begin{aligned} \langle \mathbf{m}^2(t) \rangle &= 2c_0 \mathbf{f}_j(t) \cdot \mathbf{f}_i(t) \int_{t_0}^t \frac{dt_1}{n_0^5(\mathbf{r}_0(t_1))M^2(t_1)} \\ &\times D_{km} \left(\frac{n_0(\mathbf{r}_0(t_1))}{c_0} \frac{d\mathbf{r}_0}{dt_1}; \mathbf{r}_0(t_1) \right) \times [M_{3j}(t_1)q_{1k}(t_1) \\ &+ M_{4j}(t_1)q_{2k}(t_1)][M_{3i}(t_1)q_{1m}(t_1) \\ &+ M_{4i}(t_1)q_{2m}(t_1)], \end{aligned} \quad (55)$$

where the summation over indices $j, i = 1, 2, 3, 4$ and $k, m = 1, 2, 3$ is implied and an integral characteristic D_{km} of environmental perturbations is defined as follows:

$$D_{km}(\mathbf{s}; \mathbf{r}) = - \int_0^{+\infty} \left. \frac{\partial^2 C(\mathbf{x}, \mathbf{r})}{\partial x_k \partial x_m} \right|_{\mathbf{x}=\mathbf{s}l} dl, \quad \mathbf{s}^2 = 1. \quad (56)$$

Note that

$$D_{km}(\mathbf{s}; \mathbf{r}) \sim L_1^{-2} D(\mathbf{s}; \mathbf{r}). \quad (57)$$

Like $D(\mathbf{s}; \mathbf{r})$, $D_{km}(\mathbf{s}; \mathbf{r})$ is independent of the unit vector \mathbf{s} , which specifies the direction of integration in Eq. (56), in the case of isotropic perturbations. The quantities $D(\mathbf{s}; \mathbf{r})$ and $D_{km}(\mathbf{s}; \mathbf{r})$ are independent of \mathbf{r} in the case of statistically homogeneous perturbations. Applying to Eq. (55) the same reasoning as used in the end of Sec. IV B to evaluate \mathbf{m} in the case of deterministic perturbations, we obtain $\langle \mathbf{m}^2(t) \rangle \approx n_0^{-2}L_1^{-2}R^3(t)\bar{D}$. Hence, the ratio of variances of ray displacement along and across the unperturbed wave front

$$\langle \mathbf{m}^2(t) \rangle / \langle \rho^2(t) \rangle \approx L_1^{-2}R^2(t) \quad (58)$$

is independent of environmental perturbation magnitude (within the validity domain of the perturbation theory), is sensitive to the spectral content of the environmental perturbations, rapidly increases with propagation range, and becomes large compared to unity when $R \gg L_1$.

It should be emphasized that, contrary to assertions by other authors (see, e.g., Ref. 22, where the case of homogeneous background is considered), the requirement $\varepsilon|\mathbf{r}_1| \ll L_1$, or $\varepsilon^2\langle \mathbf{r}_1^2 \rangle \ll L_1^2$, is *not* necessary for the perturbation theory

results (53) and (55) to be valid. The condition $\varepsilon^2\langle \mathbf{r}_1^2 \rangle \ll L_1^2$ is replaced by a less restrictive condition $\varepsilon^2c_0^{-2}n_0^2\langle (d\mathbf{r}_1/dt)^2 \rangle \ll 1$. Using the previous estimates of $\langle \rho^2(t) \rangle$ and $\langle \mathbf{m}^2(t) \rangle$, the condition $\varepsilon^2\langle \mathbf{r}_1^2 \rangle \ll L_1^2$ can be written as

$$R^3 \ll n_0^2L_1^4/\varepsilon^2\bar{D}, \quad (59)$$

whereas the actual validity condition $\varepsilon^2c_0^{-2}n_0^2\langle (d\mathbf{r}_1/dt)^2 \rangle \ll 1$ imposes much weaker restriction

$$R \ll n_0^2L_1^2/\varepsilon^2\bar{D} \quad (60)$$

on the propagation range. In deriving an estimate $\langle (d\mathbf{m}/dt)^2 \rangle \approx c_0^{-2}n_0^{-4}L_1^{-2}R(t)\bar{D}$, which underlies the inequality (60), we utilized the relation between $d\mathbf{m}/dt$ and \mathbf{m} , discussed after Eq. (47).

It follows from Eqs. (59) and (60) that the statistical predictions, Eqs. (53) and (55), of the ray displacement remain valid under much less restrictive conditions and at longer propagation ranges than deterministic predictions. This conclusion is consistent with the observation made by Beron-Vera and Brown⁹ from their numerical simulations of sound propagation in an underwater waveguide perturbed by internal gravity waves that, for steep rays, the first-order ray travel time correction T_1 [see Eq. (14)] remains a good statistical predictor of travel time fluctuations well beyond the ranges where T_1 serves as a good deterministic predictor.

Equation (55), which is valid for an arbitrary background environment, can be simplified considerably in special cases. When the background environment is homogeneous and perturbations are statistically homogeneous, from Eq. (20) [with $n_1 = n_1(\mathbf{r}(t))$] we find that

$$\langle \mathbf{m}^2(t) \rangle = 2c_0^3n_0^{-5}[D_{22}(\mathbf{s}) + D_{33}(\mathbf{s})](t - t_0)^3/3. \quad (61)$$

For comparison, in this special case Eq. (53) gives

$$\langle \rho^2(t) \rangle = 2c_0n_0^{-3}D(\mathbf{s})(t - t_0). \quad (62)$$

For definiteness, it is assumed here that the unperturbed ray propagates along the Ox coordinate axis, so that $\mathbf{s} = c_0^{-1}n_0d\mathbf{r}_0/dt = (1, 0, 0)$.

In another special case of range-dependent background, which we considered in Sec. IV B, we average Eqs. (31), (32), and (34) for the y -component of the ray displacement, α , and for horizontal refraction angles $\delta\psi$ (an effective correction to azimuthal launch angle) and $\tilde{\psi}$ (an instantaneous value of the azimuthal angle) to find that

$$\begin{aligned} \langle \alpha^2(t) \rangle &= 2c_0h^2(t) \int_{t_0}^t \frac{dt_1}{n_0(\mathbf{r}_0(t_1))} \left[1 \right. \\ &\left. - \frac{h(t_1)}{h(t)} \right]^2 D_{22} \left(\frac{n_0(\mathbf{r}_0(t_1))}{c_0} \frac{d\mathbf{r}_0}{dt_1}; \mathbf{r}_0(t_1) \right) \Big|_{\psi=0}, \end{aligned} \quad (63)$$

$$\langle (\delta\psi)^2 \rangle = \frac{2c_0}{n_0^2(\mathbf{r}_{00})\cos^2\chi} \int_{t_0}^t \frac{dt_1}{n_0(\mathbf{r}_0(t_1))} \left[1 - \frac{h(t_1)}{h(t)} \right]^2 \left. D_{22} \left(\frac{n_0(\mathbf{r}_0(t_1))}{c_0} \frac{d\mathbf{r}_0}{dt_1}; \mathbf{r}_0(t_1) \right) \right|_{\psi=0}, \quad (64)$$

$$\langle \tilde{\psi}^2(t) \rangle = \frac{2c_0}{n_0^2(\mathbf{r}_0(t))\cos^2\chi} \int_{t_0}^t \frac{dt_1}{n_0(\mathbf{r}_0(t_1))} \times \left. D_{22} \left(\frac{n_0(\mathbf{r}_0(t_1))}{c_0} \frac{d\mathbf{r}_0}{dt_1}; \mathbf{r}_0(t_1) \right) \right|_{\psi=0}. \quad (65)$$

Equation (64) is consistent with the results of Ref. 20, where statistics of horizontal refraction angles of eigenrays have been investigated. If the background is range-independent (stratified), the integrand in Eq. (65) is a periodic function of t_1 . When an acoustic wave propagates in a waveguide with a stratified background, random perturbations are statistically homogeneous in the horizontal plane, and the propagation range is large compared to the skip distance (i.e., spatial period of a background ray), we find from Eqs. (64), (65), and (30) that

$$\langle (\delta\psi)^2 \rangle = \langle \tilde{\psi}^2(t) \rangle / 3. \quad (66)$$

The same result is also obtained in the case of a homogeneous background and statistically homogeneous perturbations. In the general case of a range-dependent background, it follows from Eqs. (63) and (64) that

$$\langle (\delta\psi)^2 \rangle = \frac{2\bar{D}_{22}R(t)}{3n_0^2\cos^2\chi}, \quad \langle \alpha^2(t) \rangle = \frac{2\bar{D}_{22}R^3(t)}{3n_0^2}, \quad (67)$$

where \bar{D}_{22} is the value of D_{22} spatially averaged along the unperturbed ray.

For the component of the along the wave front ray displacement within the xz plane, averaging of Eq. (27) gives

$$\begin{aligned} \langle \beta^2(t) \rangle &= \frac{2c_0}{n_0^2(\mathbf{r}_{00})} \int_{t_0}^t \frac{dt_1}{n_0(\mathbf{r}_0(t_1))} [\beta_1(t)\beta_2(t_1) \\ &\quad - \beta_1(t_1)\beta_2(t)]^2 \times [q_{11}^2(t_1)D_{11}(\mathbf{s};\mathbf{r}) \\ &\quad + 2q_{11}(t_1)q_{13}(t_1)D_{13}(\mathbf{s};\mathbf{r}) \\ &\quad + q_{13}^2(t_1)D_{33}(\mathbf{s};\mathbf{r})]_{\mathbf{s}=[n_0(\mathbf{r}_0(t_1))/c_0][d\mathbf{r}_0/dt_1], \mathbf{r}=\mathbf{r}_0(t_1)}, \end{aligned} \quad (68)$$

where $\mathbf{q}_1 \equiv (q_{11}, 0, q_{13})$. As noted after Eq. (28) in Sec. IV B, the quantity in the first square brackets in the integrand in Eq. (68) is $c_0 O(t-t_1)$. Then it again follows from Eq. (68) that variance of the along the wavefront component of the ray displacement β accumulates with ray length as $R^3(t)$, as α does, but, in the case of anisotropic environmental perturbations, is controlled by different statistical characteristics of the latter. An inspection shows that Eqs. (61), (63), and (68) agree with respective corollaries of the general Eq. (55).

VI. DISCUSSION

Time derivatives of the components of a ray trajectory perturbation across and along an unperturbed wave front are of the same, first order in the magnitude of refraction index perturbations. Therefore, at first glance, it appears puzzling that ray displacements along an unperturbed wave front can be much larger than the displacements across the wavefront. It is shown that not the weakness of the environmental perturbations but a different small parameter is responsible for the anisotropy of ray scattering relative to an unperturbed wave front. This parameter is proven to be the ratio L_1/R of a representative spatial scale of perturbations and the ray length (or propagation range). The reciprocal ratio R/L_1 can be interpreted as the number of uncorrelated inhomogeneities encountered by the acoustic wave or as the number of independent scattering events, which needs to be large for the anisotropy of the ray scattering to emerge.

Ultimately, the anisotropy stems from the fact that the eikonal equation imposes a local relation (11) between the refraction index perturbations at a point in space and time derivative of the normal ray displacement at this point. Because of the local relation, the time derivative of the normal displacement, unlike the time derivatives of the tangential components of the ray displacement, does not accumulate with propagation range, and the tangential components grow much faster with range than the normal component. From a different perspective, the local relation (11) can be viewed as a first integral of the Cauchy problem (8) and (9) for time-dependent ray displacement relative to an unperturbed ray.

The anisotropy of ray scattering can be concisely formulated as the following statement: perturbations (displacements) of wave fronts are much smaller than perturbations (displacements) of rays forming these wave fronts. The geometry of a wave front is sensitive only to ray trajectory perturbations across the wave front; displacements of rays along the wavefront are manifested in variations of wave amplitude and possibly in splitting of an eigenray in the background environment into a number of micromultipaths. When micromultipathing occurs, a single-sheet branch of the unperturbed wave front is transformed into a wave front consisting of many closely spaced sheets.^{5,23} In this case, normal displacement of the perturbed rays gives an upper bound on the wavefront “widening,” i.e., the distance between various sheets of the perturbed wave front.

It should be emphasized that, somewhat surprisingly, it is the properties of the sound-speed fluctuations rather than the properties of the background sound-speed field that prove to be responsible for the wave front stability. In particular, guided propagation is not a prerequisite of wave fronts being stable compared to rays. The proportionality constant in the asymptotic relation $[\langle \rho^2(t) \rangle / \langle \mathbf{m}^2(t) \rangle]^{1/2} \propto L_1/R(t)$ is, of course, dependent on the background sound-speed field. Quantitative dependence of wave front stability on background sound-speed profile at long-range sound propagation in weakly range-dependent waveguides has been previously studied by Beron-Vera and Brown.⁹

In the underwater acoustic waveguide, the anisotropy ratio $[\langle \mathbf{m}^2(t) \rangle / \langle \rho^2(t) \rangle]^{1/2}$ is much larger for steep refracted

rays than for flat rays, when the source and receiver are located close to the waveguide's axis. According to Eq. (58), the difference is associated with the difference in the effective value of correlation length L_1 of environmental perturbations. Two mechanisms contribute to the difference in the anisotropy ratio. First, turning points of steep rays are located at shallower depths, where shorter internal waves exist.²⁴ Flat rays do not deviate far from the waveguide axis and sense only long internal waves. Second, internal gravity wave-induced sound-speed perturbations are strongly anisotropic, with their horizontal correlation length being 1–2 orders of magnitude larger than their vertical correlation length.²⁴ Because of the anisotropy of the environmental perturbations, L_1 in Eq. (57) is close to the horizontal correlation length for flat rays and is much smaller for steep rays. Thus, in agreement with observations,^{1–4} end points of perturbed steep rays are much closer to corresponding unperturbed wavefronts than end points of flat rays.

For orientation, taking $L_1=1$ km for steep refracted rays and $R=300$ km, from Eq. (58) we have $[\langle \mathbf{m}^2(t) \rangle / \langle \rho^2(t) \rangle]^{1/2} \approx 300$. It means that across-the-wave-front ray displacement is just a few meters (and, hence, root-mean-square travel-time fluctuations are just a few milliseconds), when along the front ray perturbations approach the vertical extent of the waveguide.

Previously, we have related the anisotropy of multiple ray scattering to the property of the eikonal equation to constrain the time derivative of the normal ray displacement by a local value of the refraction index perturbation. An alternative and arguably more intuitive explanation of the anisotropy can be obtained using Fermat's principle.

From geometrical considerations, first-order ray displacement \mathbf{m} across an unperturbed ray is proportional to a change in the ray direction and, hence, to perturbation in the refraction index gradient [cf. differential equation (16) for \mathbf{m}]. Change in the ray direction by an angle $\delta=O(\varepsilon)$ leads to an insignificant ray displacement of the order of $1-\cos \delta=O(\varepsilon^2)$ along the unperturbed ray. A different mechanism, namely, the sound-speed variation, is responsible for the displacement ρ along the unperturbed ray. This displacement can be compensated for by an appropriate change in the launch time of the ray and, therefore, equals $-c_0 T_1$, where T_1 is the first-order perturbation in the travel time between the source and the end point of the unperturbed ray. According to Fermat's principle, T_1 is given by an integral of the refraction index perturbations n_1 along the unperturbed ray, while the ray displacement across the unperturbed ray is given by an integral of a component of the n_1 gradient, which is proportional to n_1/L_1 . For \mathbf{m} and ρ to have the same dimensionality, the integral for \mathbf{m} has to accumulate faster with ray length R . From $\mathbf{m} \propto L_1^{-1}$ it follows that $|\mathbf{m}|/\rho \propto L_1^{-1}R$. Thus, the anisotropy of the ray scattering, including its scaling with the main parameters of the problem, follows directly from rather general physical considerations.²⁵

Previously, we implied that the medium is nondispersive and utilized proportionality between the ray travel time and the eikonal increment. In dispersive media, where refraction index is a function of wave frequency, transient signals do not necessarily possess wavefronts in the time domain. How-

ever, the differential ray equations are unaffected by the dispersion, and all the above results apply also to continuous waves (cw) in dispersive media as long as wave fronts are understood as surfaces of constant wave phase (or eikonal). In particular, at long-range propagation in a horizontally inhomogeneous ocean, wave fronts of cw adiabatic normal modes, which are curves in the horizontal plane,¹⁵ are more stable than horizontal (modal) rays comprising these wave fronts.

Anisotropy of multiple ray scattering by weak refraction index perturbations is a basic and essentially geometrical mechanism of wave front stability. Additional physical mechanisms can be involved (and, conceivably, be more efficient) in specific problems, including finite-frequency effects such as wave amplitude diffusion along wave fronts. Moreover, the previous analysis does not apply at very large propagation ranges, where second- or higher-order (in the magnitude of the sound-speed perturbations) contributions overtake the first-order contribution to the ray displacement.

VII. CONCLUSIONS

We have demonstrated that, at long-range propagation in an arbitrary inhomogeneous medium, wavefronts are much more stable with respect to environmental perturbations than individual rays, which form the wave fronts.

The relative stability of the wave fronts takes place because, at any given moment, scattering of the end points of rays resulting from weak perturbations in the refraction index occurs primarily along wave fronts of the unperturbed wave with the same travel time. Scattering of the ray end points can be characterized by the anisotropy ratio, defined as a ratio of norms of components of the end points' displacement along and across the unperturbed wave front. At large propagation ranges, the anisotropy ratio is shown to increase as the ratio of the ray length to a representative spatial scale of the refraction index perturbations.

The origin of the relative stability of the wave fronts is traced back to a specific property of the eikonal equation, which relates the time derivative of the normal ray displacement to a local value of the refraction index perturbation [see Eq. (12)]; this derivative does not accumulate with propagation range. There is no such restriction on the time derivative of the along the wavefront ray displacement; and this derivative, in fact, accumulates with range.

Fermat's principle provides a more intuitive explanation of the anisotropy of the multiple ray scattering. Theoretically derived dependencies of the anisotropy ratio on ray length and spatial scale of environmental inhomogeneities are reproduced in a simple dimensional analysis based on Fermat's principle.

Our results provide a theoretical explanation of and indicate environmental parameters responsible for the relative stability of acoustic wave fronts in the underwater waveguide in the deep ocean, which has been repeatedly observed previously in experimental data^{1–4} as well as in results of numerical simulations^{5–9} Accuracy of the ray perturbation theory, which underlies our analysis, decreases with propagation range and amplitude of environmental perturbations.

The limiting range is estimated to be a few hundred kilometers for steep refracted rays in the deep ocean. Further research, including numerical simulations, is necessary to quantify and test the limits of validity of the theory.

Further research is also necessary to determine whether the properties of ray scattering anisotropy, which have been established in this paper and are responsible for wave front stability, remain valid for waves in nonstationary, moving, and anisotropic media, as well as in problems that involve wave scattering at a rough surface.

ACKNOWLEDGMENTS

This work was supported, in part, by the Office of Naval Research. The author is grateful to John A. Colosi (Naval Postgraduate School, Monterey, CA) for drawing author's attention to this problem, to Valery U. Zavorotny (NOAA, Earth System Research Laboratory, Boulder, CO) and Michael Charnotski (Zel Technologies, Boulder, CO) for valuable discussions, and to Michael G. Brown (University of Miami, Miami, FL) for his critical review of an early version of this manuscript.

- ¹J. L. Spiesberger, R. C. Spindel, and K. Metzger, "Stability and identification of ocean acoustic multipaths," *J. Acoust. Soc. Am.* **67**, 2011–2017 (1980).
- ²T. F. Duda, S. M. Flatté, J. A. Colosi, B. D. Cornuelle, J. A. Hildebrand, W. S. Hodgkiss, P. F. Worcester, B. M. Howe, J. A. Mercer, and R. C. Spindel, "Measured wavefront fluctuations in 1000-km pulse propagation in the Pacific Ocean," *J. Acoust. Soc. Am.* **92**, 939–955 (1992).
- ³P. F. Worcester, B. D. Cornuelle, J. A. Hildebrand, W. S. Hodgkiss, T. F. Duda, J. Boyd, B. M. Howe, J. A. Mercer, and R. C. Spindel, "A comparison of measured and predicted broadband acoustic arrival patterns in travel time-depth coordinates at 1000 km range," *J. Acoust. Soc. Am.* **95**, 3118–3128 (1994).
- ⁴P. 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. Birdsall, and A. B. Baggeroer, "A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean," *J. Acoust. Soc. Am.* **105**, 3185–3201 (1999).
- ⁵J. Simmen, S. M. Flatté, and G.-Yu. Wang, "Wavefront folding, chaos and diffraction for sound propagation through ocean internal waves," *J. Acoust. Soc. Am.* **102**, 239–255 (1997).
- ⁶J. A. Colosi, E. K. Scheer, S. M. Flatté, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. Birdsall, and A. B. Baggeroer, "Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern north Pacific Ocean," *J. Acoust. Soc. Am.* **105**, 3202–3218

(1999).

- ⁷M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, "Ray dynamics in long-range deep ocean sound propagation," *J. Acoust. Soc. Am.* **113**, 2533–2547 (2003).
- ⁸F. J. Beron-Vera, M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, "Ray dynamics in a long-range acoustic propagation experiment," *J. Acoust. Soc. Am.* **114**, 1226–1242 (2003).
- ⁹F. J. Beron-Vera and M. G. Brown, "Travel time stability in weakly range-dependent sound channels," *J. Acoust. Soc. Am.* **115**, 1068–1077 (2004).
- ¹⁰D. R. Palmer, T. M. Georges, and R. M. Jones, "Classical chaos and the sensitivity of the acoustic field to small-scale ocean structure," *Comput. Phys. Commun.* **65**, 219–223 (1991).
- ¹¹O. Witte, M. Roth, and G. Müller, "Ray tracing in random media," *Geophys. J. Int.* **124**, 159–169 (1996).
- ¹²A. L. Virovlyansky, *Luchevaya teoriya dal'nego rasprostraneniya zvuka v okeane (Ray Theory of Long-Range Sound Propagation in the Ocean)* (Institute of Applied Physics, Nizhny Novgorod, 2006) (in Russian).
- ¹³S. M. Flatté and J. A. Colosi, "Anisotropy in the spatial distribution of the arrival points of a ray after a fixed travel time," *J. Acoust. Soc. Am.* **120**, 3061 (2006).
- ¹⁴O. A. Godin, "Systematic distortions of signal propagation times in random inhomogeneous media," *Dokl. Phys.* **48**, 389–394 (2003).
- ¹⁵L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media. 2: Point Sources and Bounded Beams*. 2nd, extended ed. (Springer, Berlin, 1999), Chaps. 5 and 7.
- ¹⁶Yu. A. Kravtsov, *Geometrical Optics in Engineering Physics* (Alpha Science, Harrow, UK, 2005), Chap. 2.
- ¹⁷S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radiophysics. 4: Wave Propagation through Random Media* (Springer, Berlin, 1989), Chap. 1.
- ¹⁸The order-of-magnitude estimates (21) and (22) refer to the case when the source is located at a generic point in an inhomogeneous medium. According to Eqs. (19) and (20), $|\rho|$ and $|\mathbf{m}|$ increase at the same rate with time in the degenerate special case when the refraction index perturbations depend on a single coordinate and there are no perturbations at the source location [i.e., $n_1(\mathbf{r}_{00}) \equiv 0$].
- ¹⁹O. A. Godin, "A 2-D description of sound propagation in a horizontally-inhomogeneous ocean," *J. Comput. Acoust.* **10**, 123–151 (2002).
- ²⁰O. A. Godin, V. U. Zavorotny, A. G. Voronovich, and V. V. Goncharov, "Refraction of sound in a horizontally-inhomogeneous, time-dependent ocean," *IEEE J. Ocean. Eng.* **31**, 384–401 (2006).
- ²¹This estimate does not apply to closed systems (resonators) after ray length exceeds linear dimensions of the resonator.
- ²²M. Roth, G. Müller, and R. Snieder, "Velocity shift in random media," *Geophys. J. Int.* **115**, 552–563 (1993).
- ²³F. D. Tappert and X. Tang, "Ray chaos and eigenrays," *J. Acoust. Soc. Am.* **99**, 185–195 (1996).
- ²⁴S. M. Flatté, R. Dashen, W. H. Munk, K. M. Watson, and F. Zachariassen, *Sound Transmission through a Fluctuating Ocean* (Cambridge University Press, Cambridge, UK, 1979), Chap. 3.
- ²⁵The same dimensional arguments can be used to explain why ray travel times are more stable and predictable than ray amplitudes.

Assessing a dual-frequency identification sonars' fish-counting accuracy, precision, and turbid river range capability

Suzanne L. Maxwell^{a)} and Nancy E. Gove

Alaska Department of Fish and Game, 43961 Kalifornsky Beach Rd, Suite B, Soldotna, Alaska 99669

(Received 30 April 2007; revised 18 September 2007; accepted 20 September 2007)

Accurately assessing migrating salmon populations in turbid rivers with hydroacoustics is challenging. Using single, dual, or split-beam sonars, difficulties occur fitting acoustic beams between the river's narrow boundaries, distinguishing fish from nonfish echoes, and resolving individual fish at high densities. To address these issues, the fish-counting capability of a dual-frequency identification sonar (DIDSON), which produces high resolution, video-like images, was assessed. In a clear river, fish counts generated from a DIDSON, an echo counter, split-beam sonar, and video were compared to visual counts from a tower, a method frequently used to ground-truth sonars. The DIDSON and tower counts were very similar and showed the strongest agreement and least variability compared to the other methods. In a highly turbid river, the DIDSON's maximum detection range for a 10.16 cm spherical target was 17 m, less than absorption and wave spreading losses predict, and 26 m in clear water. Unlike tower and video methods, the DIDSON was not limited by surface disturbances or turbidity. DIDSON advantages over other sonars include: better target resolution; wider viewing angle; better coverage of the water column; accurate direction of travel; and simpler to aim and operate.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799500]

PACS number(s): 43.30.Sf, 43.30.Vh [RCG]

Pages: 3364–3377

I. INTRODUCTION

Counting migrating fish in turbid rivers has always been challenging. Sonar systems traditionally used for counting fish include Bendix counters (515 kHz single-beam echo counters) (Gaudet, 1990), dual-beam sonars (Maxwell, 2000), and split-beam sonars (Daum and Osborne, 1998; Enzenhofer *et al.*, 1998; Miller and Burwen, 2002). In 2002, a dual-frequency identification sonar (DIDSON) first became available for fish counting. The standard DIDSON is a multi-beam, high-frequency (1.8 and 1.0 MHz) sonar coupled with an acoustic lens that focuses sound energy to create video-like images where fish can be observed moving across the static background of the river bottom when deployed at a fixed location (Belcher *et al.*, 2002). The static background image can be removed using a background subtraction algorithm, leaving the moving fish targets (Fig. 1). This study, previously reported in-house (Maxwell and Gove, 2004), was designed to assess the use of the DIDSON for fish enumeration by comparing migrating fish counts from the DIDSON in a clear river with counts from a tower observer and determining the DIDSON's range capacity in a highly turbid river. Since this study, use of the DIDSON in fisheries has spread primarily because of the high resolution fish images, and unlike the more traditional sonars, the DIDSON is relatively easy to use and images are easy to interpret. A later clear-river comparison was performed (Holmes *et al.*, 2006), but only at low fish migration rates. Researchers have used the DIDSON for a number of fishery-related problems including monitoring smolt behavior (Moursund *et al.*, 2003),

studying salmon *Oncorhynchus spp.* redds (Tiffan *et al.*, 2004), verifying split-beam sonar estimates of upstream fish passage (passage and passage rates refer to the number of fish migrating past the sonar site) (Xie *et al.*, 2005), and studying fish behavior around traps and entrances to fishways (Baumgartner *et al.*, 2006).

In many turbid rivers, sonar fish counters are important tools in the management of salmon fisheries. Annual salmon counts from Bendix counters have been used to set escape-ment goals and provide in season estimates of migrating salmon for the Alaska Department of Fish and Game (ADF&G) for over 30 years (Gaudet, 1990). The Bendix counter was not without problems. The primary feedback is a voltage display on an oscilloscope. The automated output consists of counts per range sector. Low voltages can be produced from small fish crossing the beam center or large fish grazing the beam edge. Voltage spikes that cross a counting threshold a certain number of times (dependent on preset, range-dependent hit criteria) are counted. The narrow circular beams of the Bendix counter, that are necessary to eliminate mixing fish echoes with echoes from boundary layers, may allow fish to pass undetected over (or under) the beams at close range. We evaluated the DIDSON for the purpose of counting sockeye salmon as a potential replacement for the Bendix counter after assessing and rejecting the split-beam sonar because of the difficulties outlined below (Maxwell *et al.*, 2007).

Although split-beam sonars have been successfully used to count Chinook salmon *O. tshawytscha*, chum salmon *O. keta* (Miller and Burwen, 2002; Daum and Osborne, 1998), and other migrating fish species (Lilja *et al.*, 2003), these sonars have been less successful for counting the dense schools of migrating sockeye salmon *O. nerka*. Salmon

^{a)}Author to whom correspondence should be addressed. Electronic mail: suzanne.maxwell@alaska.gov

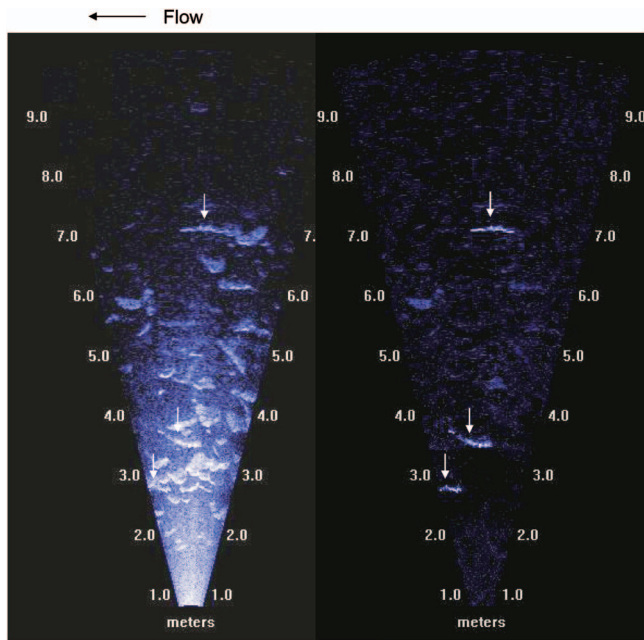


FIG. 1. A DIDSON image of three fish (noted with arrows) swimming over a rocky background (left) and the same three fish with the static background removed (right).

counts from split-beam sonars have compared well with visual counts up to 2000 fish/h, but were shown to level off as visual counts continued to increase (Enzenhofer *et al.*, 1998). We encountered several problems using split-beam sonars to count sockeye salmon. In rivers, a small change in the vertical position of the split-beam transducer (sometimes half of a degree) can mean the difference between high fish detection and no fish detection. The narrow beam required to ensound the desired sampling range provides limited coverage of the water column at close range, where the majority of sockeye salmon pass. At high densities, overlapping echoes reduce the number of echoes obtained per fish (Cronkite and Enzenhofer, 2002). As fish travel through the beam, numerous echoes generated from multipathing mingle with echoes from the river bottom, debris, bubbles, etc. to compromise fish detection. Determining the direction of travel from a split-beam sonar is difficult because unwanted echoes mix with fish echoes to produce confusing tracks, and at close range, the fish occupies the entire beam making it impossible to determine which side of the beam the fish entered first. Our choice of an elliptical beam allowed better coverage along the river bottom, but also created opportunity for multiple fish at one range to simultaneously enter the beam. The higher energy of the split-beam sonar compared to the Bendix counter created more problems from boundary and volume reverberation and multipathing. In addition, the filtering process for the echo returns from Hydroacoustics Technologies, Inc. (HTI) split-beam sonars considerably reduced the number of received echoes. Because the HTI sonars do not include the raw, unfiltered sample data electronically, the low numbers of echoes received required us to use antiquated chart recorders coupled with dot-matrix, continuous-feed printers. At sites where split-beam sonars are used for management, these same recorders and printers are still in use.

For these reasons, we began testing the DIDSON as a potential Bendix replacement.

Holmes *et al.* (2006) showed that DIDSON counts of migrating adult salmon were both accurate and precise, but fish passage rates in the study were low (for the unimpeded fish, less than 300 fish/h). In many rivers, hourly passage rates exceed 2000 fish/h (Brazil, 2007; Westerman and Willette, 2006). According to Holmes *et al.*, the density of fish schools does not change as hourly passage rates increase. At low passage rates, an hour of fish passage is often broken up into short periods of dense fish schools interspersed with long periods of no fish passage (gaps). If higher hourly passage rates are created by filling in the gaps, we would not expect the density of fish within an individual school to change. However, we have observed that as passage rates increase, schools grow larger, denser, and more continuous making visual counting difficult.

Ground-truthing sonars in turbid rivers is difficult, so evaluations are often performed in clear rivers. However, significant signal loss has been observed under conditions of high turbidity even with lower frequency (120 kHz) sonars (Maxwell, 2000). Signal loss results from a variety of factors in addition to wave spreading including absorption, scattering from macroscopic particles, and boundary interactions. In rivers with narrow water columns, interactions of the sound waves with boundary layers may restrict the wave spreading and result in a focusing of the beam, rather than a loss. In freshwater, absorption is often viewed as the primary cause of signal attenuation outside of wave spreading and although absorption coefficients are small for low frequency sonars, they become large at high frequencies (Francois and Garrison, 1982). Scattering due to macroscopic particles is less understood but is known to substantially attenuate the returning signal, with greater loss observed at high frequencies (Richards *et al.*, 1996). At the standard DIDSON's high frequencies, we expected a large degree of signal loss in turbid rivers.

Our primary objectives were to compare the similarity between salmon counts from the DIDSON and tower in a clear river and evaluate the DIDSON's range capacity in a turbid river. For the range comparison, an artificial target was used to evaluate the DIDSON's detection limits in a turbid river and clear lake, and range-dependent signal loss was examined. In a clear river, we obtained synchronized counts of migrating salmon from three sonars, the DIDSON, Bendix counter, and split-beam sonar, along with counts from a video and an observer in a tower, with the tower counts used as the ground truth. Counting towers have been operated by ADF&G for over 40 years, and the accuracy and precision of the counts has been validated (Becker, 1962). Secondary objectives included examining the accuracy of the other sonar and video methods under the same conditions used to assess the DIDSON. To determine whether results from this study could be applied outside the range of hourly passage rates tested, we calculated the relative precision of DIDSON estimates between multiple observers grouped by passage rate and looked at fish density as a function of hourly passage rate. If fish density does not change as hourly passage rates

increase, the results should be applicable outside the range of this study; if not, the results should only be applied within the range of passage rates observed.

II. METHODS AND MATERIALS

A. Study areas

The clear river evaluation was conducted at a fish-counting tower site along the Wood River, approximately one mile below the outlet of Lake Aleknagik (outside of Dillingham, Alaska). Here, the sockeye salmon run is large and condensed into approximately two weeks. During the year of this study, a total of 1.284 million sockeye salmon were counted at the tower site with a maximum daily rate of 239,000 salmon (West, 2003). At the deployment site, the cross-river profile of the bottom is mostly linear with an 8.5° slope and substrate material composed predominantly of pebble-size gravel mixed with sand. At the time of the study, the water temperature measured 9 °C (sound speed of 1443 m/s) and the turbidity 8 NTUs (nephelometric turbidity units) with a factory calibrated Global Water Model WQ770 turbidity meter. Although the area is tidally influenced with 60 cm water level fluctuations, no salt water reaches the site, and the 1–2 m/s current flow discourages salmon from milling.

Range tests were conducted in the lower portion of the Copper River near the mile 27 bridge (road miles) and at a barge on Lake Washington, Seattle, owned by the University of Washington's Applied Physics Laboratory. The river bottom at the Copper River site was composed of gravel mixed with silt. The river is highly turbid from glacial silt, and its appearance is completely opaque. During 1991 and 1993, (U.S. Geological Survey) measured suspended sediment levels at the mile 27 bridge of 0.5–2.3 kg/m³ (Brabets, 1997). Turbidity measured using the Global Water meter in 2002 oscillated considerably, but for the most part hovered above 800 NTUs (the meter's limit was 1000 NTUs). The water temperature was 7.4 °C, a sound speed of 1437 m/s. The site was selected for its high turbidity, and because ADF&G operates a Bendix counter at the river near the outlet of Miles Lake. Lake Washington has a mostly mud bottom with some sections of gravel. The depth of the lake where the range tests were performed ranged from 4–10 m. The water of Lake Washington was very clear at the time of the study. No turbidity measures were taken, but the lake has reported secchi depth readings of 3.7–5.5 m (McIntyre *et al.*, 2006). The water temperature was 10 °C, a sound speed of 1447 m/s.

B. Clear river evaluations

1. Equipment and setup

For the clear river comparison, we installed a standard DIDSON, a Bendix counter, a split-beam sonar, and a video camera at the Wood River tower site. The counting tower was located along the shore with the front legs submerged. The video camera was installed on the counting tower at a height of 4 m above the river's surface. The DIDSON was deployed in front of the tower so the DIDSON, video, and tower all sampled the same region of water. The Bendix counter and split-beam sonar were positioned just upriver.

Each sonar was deployed nearshore with the beam directed offshore, perpendicular to current flow. Initially, we deployed green flash panels on the river bottom to improve viewing from the tower. The fish avoided the panels, so we moved one panel offshore and used it to divert fish toward shore. The nearshore edge of the panel, 8.5 m from the DIDSON transducer, served as the end range for data collection. We did not install a weir to keep fish from traveling inshore of our transducers; however, fish appeared to stay offshore possibly due to the tower's influence.

Visual, real-time counts from an observer sitting on top of a counting tower (tower counts) served as the control for the study. Becker (1962) determined that although the accuracy of tower counts was affected by different counting personnel, migration intensities, weather conditions, and disrupted migration patterns, none of the factors were found to inject significant directional errors in the counts, but showed definite tendencies to cancel out. He established confidence limits of ±3.99% at a 95% level of significance. For our study, a single observer counted and classified each sample as very good, good, moderate, or uncertain depending on viewing conditions and passage rates. For periods rated moderate or uncertain, video images would be counted and substituted for the tower counts. The video camera (Supercircuits Model PC-33C, high resolution color camera) was equipped with a Computar 3.5–8.0 mm varifocal, auto-iris lens fitted with a polarizing filter. The varifocal lens was set at a focal length of 6.0 mm for a 7 m horizontal field of view at a range of 8.5 m. Video images were stored on an analog SVHS recorder recording 9.1 m/s and a desktop PC recording 4.6 m/s. Salmon were counted from analog and digitized images, but observers reported poor quality of the digitized images, so only the analog counts were used.

The standard DIDSON's 1.8 MHz beam is divided into 96 0.3° × 12° beams for an overall field of view of 28.8° × 12°, with range settings extending to 15 m. The 1.0 MHz beam is divided into 48 0.6° × 12° beams, with range settings extending to 50 m. (Note: in all later DIDSON models, the beam elevation is 14°, rather than 12°.) The two-way sidelobes are down 25–30 decibels (dB). The multiple beams are not equal in intensity. The beams along the edge are 15–20 dB down from the center beam. At the time of this study, the DIDSON had no time-varied gain (TVG) or absorption functions, so the intensity down range was not constant. We anticipate little difference in the DIDSON's image quality with the addition of a TVG function because the noise in a riverine environment is primarily from reverberation off boundaries and particles in the moving water. The near field boundaries for the DIDSON's high and low frequencies are 24.2 and 13.4 m, respectively. The near field boundary range is reduced by focusing the lens, which changes the relative delays of the wavelengths producing a coherent wave closer to the transducer. With the focus mechanism set at its minimum range, the DIDSON produces a sharp image at 1½ m, and images that are acceptable, but not as sharp at 1 m. Like a camera lens, when you focus on a close range object, the depth of field becomes blurry. A

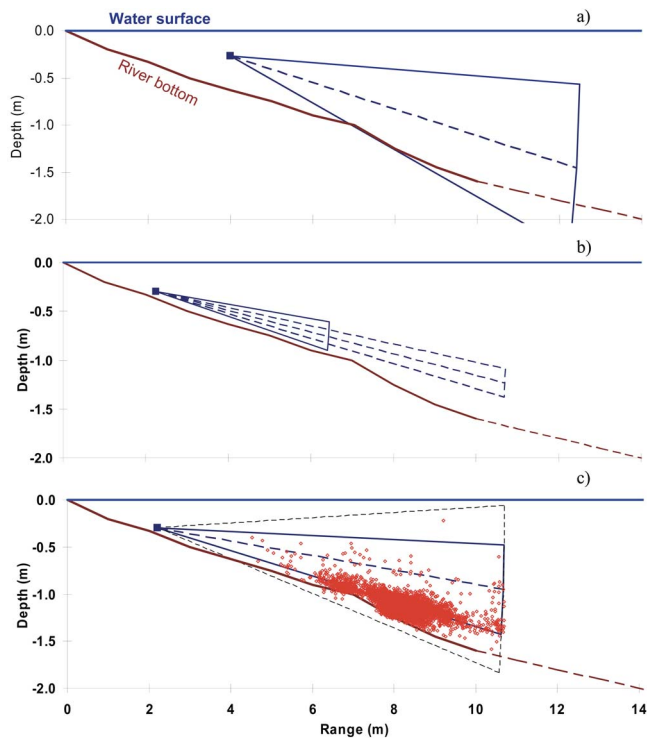


FIG. 2. (Color online) The Wood River bottom profile and the nominal vertical beam of (a) the DIDSON's 12° beam, (b) the Bendix counter's 4° nearshore and 2° offshore beams, and (c) the split-beam sonar's 6.4° beam (solid lines) and the width of the effective beam (dotted lines), along with fish targets (dots) obtained from split-beam echoes.

focus range of 7 m is no different than a focus range of infinity; objects viewed at 7 m to the end of the range will be in focus.

The DIDSON transducer was deployed with a BioSonics' heading, pitch, and roll sensor and an HTI automated rotator, and positioned 36 cm (to the lower edge) above the river bottom in water 63 cm deep. The transducer was angled 8.0° down from level. Typically, with split-beam and single-beam sonars, the transducer is positioned just above the river bottom. We intentionally raised the DIDSON higher above the river bottom with a steeper downward angle. This beam geometry was intended to facilitate coverage of the nearshore range and reduce shadowing from fish traveling close to the transducer. The approximate beam geometries of the three sonars are illustrated in Fig. 2 along with the position of fish targets determined from split-beam echoes. Sampling settings for the DIDSON included high frequency mode (1.8 MHz), a 0.75 m start range, a 9.0 m window length, a 5.08 m focus, a 40 dB receiver gain, and a frame rate of 8 frames/s. The pulse length, which is dependent on the window length and frequency, was 32 μ s. Sound speed was calculated based on the water temperature (Simmonds and MacLennan, 2005) and input into the DIDSON's initialization file.

The frequency and window settings were selected to maximize the resolution of the fish targets. The clarity of the image is a combination of the window length and frequency. At the same window length, the higher frequency beam provides a sharper image. The DIDSON collects 512 data values over the length of the window. The down range resolution is

a function of the length of the window divided by 512. For the selected 9 m window, the down range resolution is 1.8 cm. As the window lengths increase, the fish images blur and eventually become blob-like with little detail. The 9 m window was the shortest window length that could be used to image the range where fish were actively migrating. The cross-range resolution is a function of the size of the individual beams. At the furthest range sampled, 9 m, the individual high frequency beams measure 4.7 cm across.

Three observers independently counted the DIDSON samples. Observers chose their preferred frame rate during playback. Settings kept constant between observers were the default background subtraction parameters and a 3 dB threshold setting. The intensity setting was adjusted to maximize the brightness of the fish without over-brightening the background. Using the background subtraction feature created bright fish images against a mostly black background. Fish moving upriver were manually counted from the DIDSON images using a tally counter. Holding and downriver-moving fish were not counted.

The Bendix counter was a 515 kHz, single-beam, echo-counting sonar, alternately transmitting between 4° nearshore and 2° offshore beams. We mounted the Bendix transducer 10 cm above the river bottom in water 39 cm deep, and adjusted the tilt angle until echoes reflected from bottom structure were just below the counting threshold. The pulse width was 0.1 ms, the range 0.75–8.5 m. The Bendix counter was calibrated hourly by counting voltage spikes produced by passing fish on an oscilloscope and adjusting the ping rate until the oscilloscope count matched the automated machine count. The counter was set to output counts at the end of each sample.

A BioSonics' 201 kHz, 6.4° circular, split-beam transducer and attitude sensor were deployed 12 cm above the river bottom in water 36 cm deep, and angled 4.4° down from level. Other settings included: 17.2 pings/s, 0.2 ms transmit pulse width, -50 dB data collection and editing threshold, 1.0–8.5 m range, and single target criteria including a -50 dB target threshold, 0.02–0.6 pulse width acceptance measured 6 dB below the pulse peak, 10 dB maximum beam compensation, and 3 dB maximum standard deviation of the alongship and athwartship angles. Split-beam sonar data were displayed using SonarData's Echoview software, and fish traces were manually counted using a tally counter.

Sonar calibrations were field tested using standard tungsten carbide spheres positioned beyond the near field of the transducers. For the Bendix counter and split-beam sonars, the near fields are slightly less than 1 m (based on Simmonds and MacLennan, 2005). The transmit level for the Bendix counter is set by adjusting the sensitivity screw and measuring the resulting peak-peak voltage on an oscilloscope. Peak voltages from a 3.81 cm calibration sphere crossed the fish-counting threshold of the Bendix counter, so no adjustments were made to the transmit level. The split-beam sonar's calibration was tested using a 3.6 cm tungsten carbide sphere. Target echo amplitudes from this sphere averaged -40.4 ± 2.8 dB (s.d.). For the DIDSON, no calibration was attempted, but the 3.81 cm calibration sphere appeared bright in DIDSON images at 3 m. Using calibration spheres

to field test the DIDSON is difficult because the calibration sphere is not a point source in relation to the DIDSON's pulse length. The pulse length (converted to cm using a 1443 m/s sound speed) is 28.86 cm for the split-beam sonar and 14.43 cm for the Bendix counter, both larger than the calibration spheres, while the DIDSON's pulse length of 0.046 cm is considerably smaller. In addition, the individual horizontal beams of the DIDSON are smaller than the calibration sphere out to 8 m. Theoretical target strengths of the calibration spheres at 9 °C for the Bendix counter, split-beam sonar, and two DIDSON frequencies are -43.2, -39.6, -38.7, and -38.3 dB, respectively (Faran, 1951).

2. Data analyses

Migrating salmon were counted in 15 min synchronized samples using each assessment method. Paired datasets were obtained by aligning the tower count with each of the other methods. Our initial analyses included plotting raw data from all of the assessment methods, residual plots of each data pair, and scatter plots using the original then logged (ln) coordinates from each paired dataset.

We selected two data analysis techniques used to measure agreement or reproducibility between variables, Bland and Altman's method of calculating the 95% limits of agreement (Bland and Altman, 1986) and the concordance correlation coefficient (CCC) (Lin *et al.*, 2002). Unlike regression analyses, these methods do not rely on the assumption of no error in the independent variable. For the Bland-Altman method, if the errors are constant, the differences between paired observations are plotted against their means along with lines representing the mean difference and limits of agreement, defined as the mean difference ± 2 standard deviations. If the errors are proportional, the data are logged, the limits of agreement calculated, and the ratios of each paired observation plotted against their geometric means in the original units. In this case, the limits of agreement are defined as the mean ratio ± 2 standard deviations.

Three factors contribute to the CCC, which ranges from 0 (no agreement) to 1 (perfect agreement): the analytical range of the dataset (i.e., homogenous data can create small r_c values), the precision (r), and the accuracy (c_a). The equation for the accuracy coefficient is

$$c_a = \frac{2}{\frac{(\bar{y} - \bar{x})^2}{s_y s_x} + \frac{s_y}{s_x} + \frac{s_x}{s_y}}, \quad (1)$$

where s_y is the sample variance in y , s_x is the sample variance in x , \bar{y} is the mean of y , and \bar{x} is the mean of x . The accuracy coefficient is equal to one when both the mean and variance of x and y are equal. In this statistic, r is the Pearson's correlation coefficient. The CCC (r_c) is the product of the accuracy (c_a) and precision (r) coefficients:

$$r_c = r \cdot c_a. \quad (2)$$

Where the Pearson correlation coefficient measures how well the relationship between two variables can be explained by a straight line, the CCC measures how well the relation-

ship between two variables can be explained by a line with a slope of 0 and an intercept of 1. For each method, we graphically examined the data to determine if differences or ratios were more appropriate. Differences would be used when the variation was constant compared to the counts, and logged ratios would be used when the variation was proportional, increasing as the counts increased.

The relative precision of DIDSON counts from multiple observers was examined using the average percent error (APE) (Chilton and Beamish, 1982), the same statistic utilized by Holmes *et al.* (2006). The DIDSON data were sorted by hourly passage rate and divided into the following groups (fish/h): 1) <101; 2) 101-300; 3) 301-1000; 4) 1001-3000; and 5) >3000. Groups 1 and 2 match the hourly passage rates of Holmes' unconstrained fish samples allowing a direct comparison of Holmes' conclusion that higher passage rates would result in lower APE values among observers. The remaining groups were divided based on breaks in the dataset.

To test the hypothesis that fish density within schools remains relatively constant as hourly passage rates increase, we used flux as a surrogate for density. School flux was defined as the number of fish/s within an individual school. Both density and flux terms require knowing the number of fish per unit volume. Because the same beam angle was used during collection of all DIDSON samples and fish schools appeared to pass randomly within a narrow range bin (3-7 m), we assumed a similar volume. Dividing the DIDSON samples into the same passage rate groups defined above, we obtained the average school flux for each group by counting the number of fish in individual schools and dividing by the time it took the nose of the first fish and the tail of the last fish in the school to cross the center beam. For passage rate groups 1-4, fish were defined to be within a single school if the gap between individual fish exceeded the window width in the DIDSON view (approximately 2.5 m at the center of the range where most fish passed). Group 5 had few or no gaps that met these criteria. For these files, we picked a spot where a single fish crossed the beam center and counted fish for 30 s. While group 5 does not actually consist of schools, this grouping gives a measure of flux over time, allowing for a similar comparison with the other groups. We analyzed 30 schools from each group, dividing the number of samples desired between the files collected. The data were analyzed by plotting the school flux (fish/s) with a standard deviation from each passage rate group against the group's average hourly passage rate.

Although the tower counts were not made available until after all counting was completed, sonar and video operators were located in the same structure and may have been influenced by each other; hence, this was not a true blind test. The Bendix counts, which are operator dependent, were obtained on site. The DIDSON, split-beam, and video files were counted at a later time by different people.

C. Turbid river evaluation

A spherical artificial target was used to evaluate the range capability of the DIDSON at the turbid Copper River and clear Lake Washington sites. At the Copper River, the

DIDSON was installed close to the shoreline with the beam directed perpendicular to current and angled above the river bottom. At Lake Washington, the DIDSON was lowered from a mount on a barge and angled downward until the beam was below the water's surface. At both sites, the DIDSON beam was positioned below the surface, but raised high enough that no reflections from the river (or lake) bottom were visible. No surface disturbances were apparent in the DIDSON images. The low frequency setting (1.0 MHz) was selected at both sites to allow us to sample beyond 15 m. At the Copper River, DIDSON settings included: a 9.0 m start range, a 9.0 m window length, a 40 dB receiver gain, and a frame rate of 6 frames/s. For these settings, the pulse length is 0.32 μ s, and the down range resolution is 1.8 cm. At Lake Washington, settings included a 5.1 m start range, a 36.0 m window length, and 4 frames/s, resulting in a 0.64 μ s pulse length and 7.0 cm down range resolution. The start range was extended to reduce the window length, improving the target resolution. We obtained images of the target at close range (2 m) at the Copper River site using low frequency, a 0.83 m start range, and a 20.01 m window length (0.64 μ s pulse length).

A 10.16 cm plastic sphere filled with 4.4 mm copper-plated steel shot (bb's) was selected for the target. The spherical target was selected over tethered fish because it is difficult to know the fish's aspect in turbid water. The acoustic size of the plastic sphere, measured by 200 kHz split-beam systems, is similar to a sockeye salmon (Maxwell *et al.*, 2007). In DIDSON images, the entire surface of the fish reflects the high-frequency sound. The reflective surface of an adult sockeye salmon is considerably larger than the plastic sphere, so the maximum detection range should be similar to, or greater than, the detection range of the plastic sphere. At each site, the plastic sphere was suspended from an extension pole off the side of a boat and lowered in the front of the transducer until the brightest image was obtained. The boat was slowly moved offshore keeping the target image in the center of the beam until it was no longer visible. If we were uncertain the image was the target, the target was raised out of the water until its image disappeared then returned to its original position.

To compare the observed maximum ranges to theoretical values, spreading loss, spreading loss plus absorption (transmission loss), and the signal-to-noise ratio (SNR) were calculated for the plastic sphere. For our purposes, the signal refers to the intensity of the plastic sphere, the noise to the background reverberation, which is a combination of volume reverberation and reflections from the river bottom. Loss due to scattering from macroscopic particles was not included in the theoretical calculations. For each site, we calculated the theoretical intensity of the plastic sphere after two-way spreading loss (I_S) and transmission loss (I_T) in 1 m range increments (R_i) by:

$$I_S = I_o - 40 \cdot \text{Log} \left(\frac{R_i}{R_o} \right), \quad (3)$$

$$I_T = I_S - 2 \cdot \alpha \cdot (R_i - R_o), \quad (4)$$

where

- I_o = is the initial intensity of the plastic sphere,
- R_o = is the range where the initial intensity of the plastic sphere was determined, and
- α = (alpha) is the absorption coefficient (Francois and Garrison, 1982).

The DIDSON software's "measure box statistics" function in combination with the zoom was used to measure all intensity values from raw, or unenhanced, DIDSON images. The I_o value was obtained by measuring and averaging intensity values of the plastic sphere positioned close to the transducer. The α values were based on water temperatures of 7.4 °C (Copper River) and 10 °C (Lake Washington) and a transducer depth of 0.5 m. To obtain the SNR, we measured the average target intensity and regions of background near the target, excluding target echoes, approximately every five frames in DIDSON images then paired the target and background intensities and averaged them in 1 m range bins. The paired target and background intensities were plotted in range increments against a transmission loss curve. A 10 dB SNR is deemed acceptable (Dahl *et al.*, 2001), but is often not achieved in riverine environments. The theoretical ranges used for comparison were points where the transmission loss curve reached 0, 5, and 10 dB above the average background reverberation. Without bottom profiles, we did not know the absolute position of the target within the beam. Moving the target up and down vertically will affect the target's intensity and detection range. Although we attempted to place the target where the image was brightest, some signal loss may also have resulted from the position of the target within the beam.

III. RESULTS

A. Clear river evaluation

A total of 40 paired samples were collected from the DIDSON and tower. During six of the sample periods, the DIDSON and tower observer were moved to the opposite side of the river in an attempt to obtain higher passage rates. The move resulted in fewer samples from the other assessment methods, 33 from the Bendix and split-beam sonars and 34 from the video. We experienced problems obtaining an optimal tilt angle for the split-beam sonar. During samples 1–3 (sample 2 was missed) the operator felt the aim was not optimum. The data were included, but the qualification must be noted.

During the study, the water remained clear, and viewing conditions from the tower were good. Winds were absent or minimal. There was a light rippling of the water surface during a few of the samples. The tower observer rated each count "very good" or "good," so no video substitutions were made. Instead, we used the video counts as a separate comparison to the tower counts. The majority of fish observed were sockeye salmon.

Fish images from the DIDSON were clearly visible, and individual fish were distinguishable even during the highest observed passage rates (Fig. 3). The still images fail to illus-

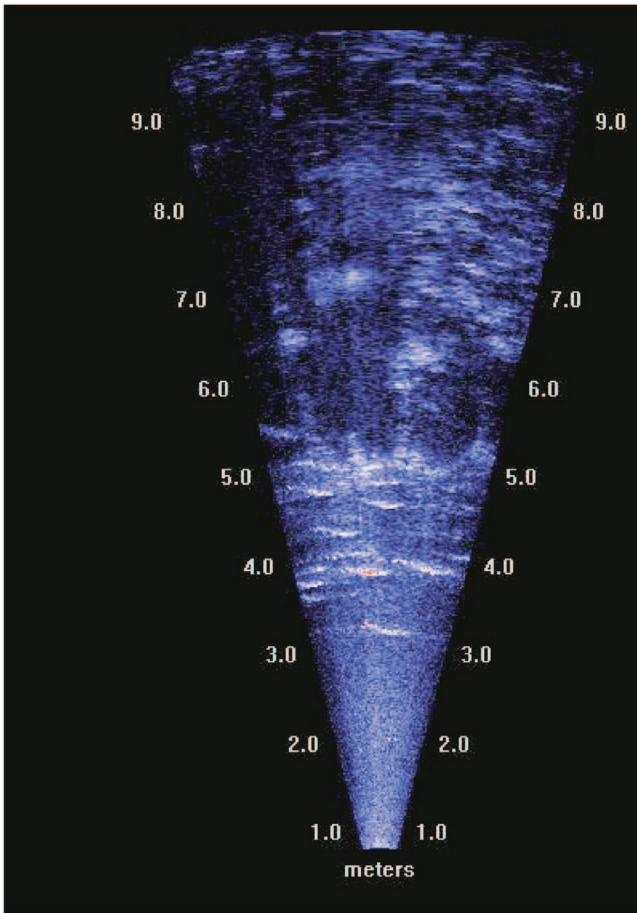


FIG. 3. A DIDSON image of fish at high passage rates.

trate the ease with which moving fish are detected. Acoustic shadows, created as fish swim through the beam and block a portion of the beam's energy, are visible in Fig. 3 between 5 and 9 m as dark oblong shadows that obscure the river bottom image. Figure 4 shows images of the river bottom with no fish passing and with several fish in the beam. Although acoustic shadows blocked out the signal from the river bottom, a fish is clearly visible at 5 m traveling through the shadow.

The voltage peaks observed on the oscilloscope coupled to the Bendix counter were strong and relatively easy to count when salmon passed through the transducer beam. From the split-beam sonar, we obtained positional data that allowed us to plot the vertical position of individual fish targets. Although the angle information from split-beam sonars in rivers is often poor (Smith and Maxwell, 2007), averaged vertical fish positions provide a general idea of where the fish are in relation to the river bottom. Fish targets, tracked with the split-beam sonar, were predominantly bottom oriented, and the majority (over 90%) passed within 4–7 m of the transducer [Fig. 2(c)]. The target strength of the tracked fish averaged -32.2 ± 2.9 dB.

Fish counts obtained from each sonar and video method were similar to counts made by the tower observer, which ranged from 8 to 1330 fish per 15 min sample (Fig. 5). The relationship between each paired dataset was similar to a slope of one for the original and logged data (Fig. 6). Total

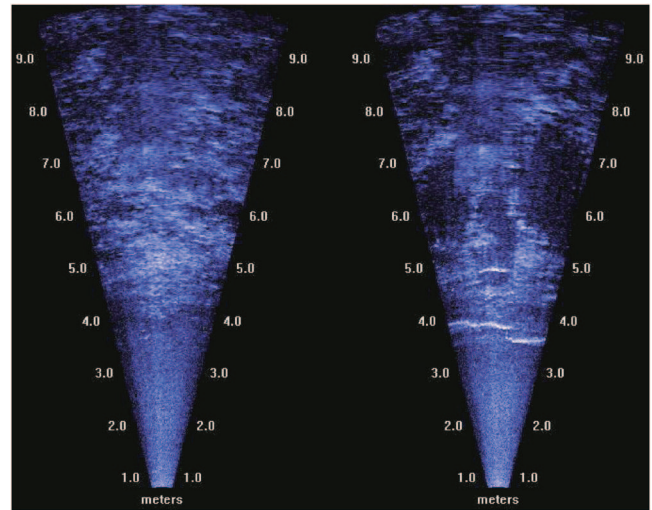


FIG. 4. A DIDSON image of the background with no fish present (left), and with fish passing through (right). A fish is visible at 5 m in one of the dark, oblong acoustic shadows created by closer-range fish partially blocking the beam's energy.

DIDSON counts were 6.9% higher, and Bendix counts 5.6% higher than tower counts. Total video counts were 3.2% lower than the tower counts, while split-beam counts were 10.8% lower. Heteroscedasticity was observed in residual plots of the paired datasets containing the DIDSON, split beam, and video and was less obvious in the Bendix dataset. We used ratios and ln transformations for the Bland–Altman and CCC analyses; however, it should be noted that the ln transformation overcorrects for the increasing variance.

The Bland–Altman ratios (Fig. 7) and CCC statistics (Table 1) indicated strong agreement between each paired dataset. For each pair, the Bland–Altman mean ratios were slightly larger than one, and the CCC's lower 95% confidence limits were all higher than 0.94 with the exception of the split-beam tower pair. The DIDSON-tower pair showed the highest levels of agreement with the least amount of variability. The video-tower pair showed the same level of agree-

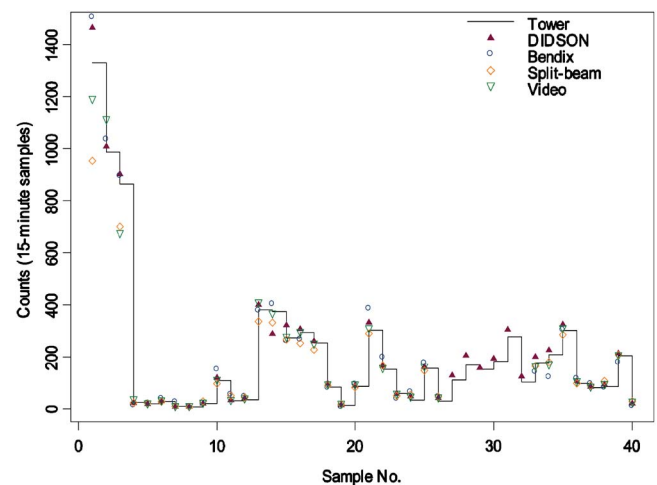


FIG. 5. (Color online) Tower, DIDSON, Bendix, split-beam, and video counts of migrating sockeye salmon in 15 min samples from the Wood River. Salmon passage occurred primarily between 4 and 7 m from the transducer during these counts.

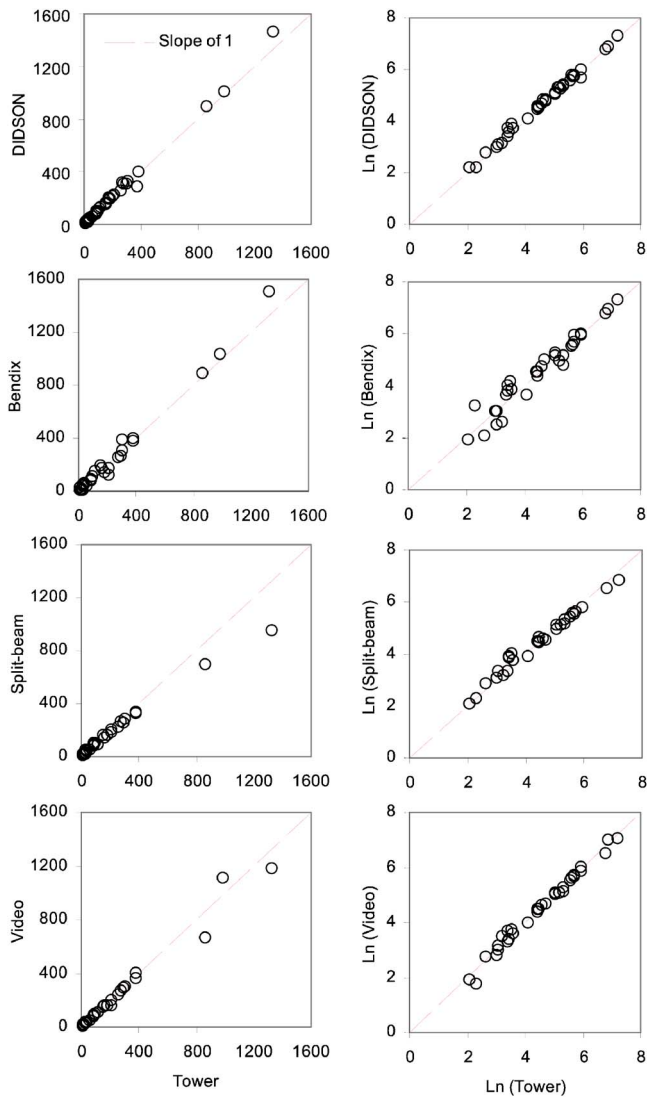


FIG. 6. (Color online) Comparisons of tower counts of migrating sockeye salmon paired with DIDSON, Bendix, split-beam, and video counts using the original (left) and logged (right) variables.

ment, with slightly more variability. The Bland–Altman limits of agreement were widest for the Bendix-tower and split-beam tower pairs. The CCC's lower 95% confidence limits for the split-beam tower pair diverged more from the tower counts compared to the other methods. Examination of the precision and accuracy coefficients showed that the wider confidence intervals were caused by lower accuracy and lower precision.

The APE from DIDSON counts made by multiple observers was 5.4%. When grouped according to passage rate

(Table II), the APE values declined as passage rates increased through group 3, then began to increase.

The school flux remained constant from group 1 to group 2, while the hourly passage rate increased. Within the remaining groups, the obvious trend was an increase in school flux as hourly passage rate increased (Fig. 8). The largest increase was observed between groups 4 and 5 where the ratio of change was 1.8. After observing this increase, we were interested in learning whether the trend would continue. Using the same sampling method, we added a group 6 containing salmon passage rates beyond 5000 fish/h obtained from DIDSON data from other rivers. School flux from group 5 to the new group 6 increased by a ratio of 1.4.

B. Turbid river evaluations

In DIDSON images, the plastic sphere was detectable to ranges of 26 m in Lake Washington and 17 m in the Copper River. The initial intensity for the plastic sphere was obtained at the Copper River with the sphere positioned 2 m from the DIDSON in the center of the lens. Raw intensity measures (i.e., not corrected for spreading loss or absorption) from 30 DIDSON images of the sphere averaged 75.2 ± 3.7 dB (s.d.). This starting point was used to generate transmission loss curves and theoretical range limitations for both sites. Because the alpha for the two sites differs by less than 0.07 dB/m between 1 and 2 m, this initial point should suffice for both.

The most striking features in the plotted intensity measures of the plastic target versus range (Fig. 9) are the high degree of scatter in both datasets, the flat slope of the Lake Washington target intensities, and the declining slope in the Copper River data. The background intensities were similar across the observed ranges at both sites. Target intensities were mostly below the transmission loss (TL) curve out to 20 m and mostly above beyond 20 m at Lake Washington, remaining wholly below the TL curve at the Copper River. Background intensity measures ranged from 2.2 to 8.6 dB, averaging 5.2 dB at Lake Washington; and 1.5–8.2 dB, averaging 4.8 dB at the Copper River. The theoretical TL curve for the plastic sphere dropped to 10, 5, and 0 dB above the average background intensity at ranges of 26.5, 30.5, and 34.9 m for Lake Washington, and 25.5, 29.2, and 33.2 m for the Copper River. At Lake Washington, the last echoes were detected at the range where the theoretical TL curve dropped to 10 dB above the average intensity of bottom echoes. At the Copper River, the detection limit (17 m) was considerably less than the range of the theoretical 10 dB point above the average background intensity.

TABLE I. The concordance correlation coefficient (CCC) and related statistics for sockeye salmon counts in the Wood River made by an observer in a tower paired with counts from the DIDSON, Bendix counter, split-beam sonar, and video.

Comparison	CCC (r_c)	95% Lower Limit (r_c)	Precision (r)	Accuracy (c_a)
Tower—DIDSON	0.994	0.990	0.997	0.998
Tower—Bendix counter	0.967	0.942	0.968	0.999
Tower—Split-beam sonar	0.951	0.891	0.966	0.985
Tower—Video	0.994	0.989	0.994	1.000

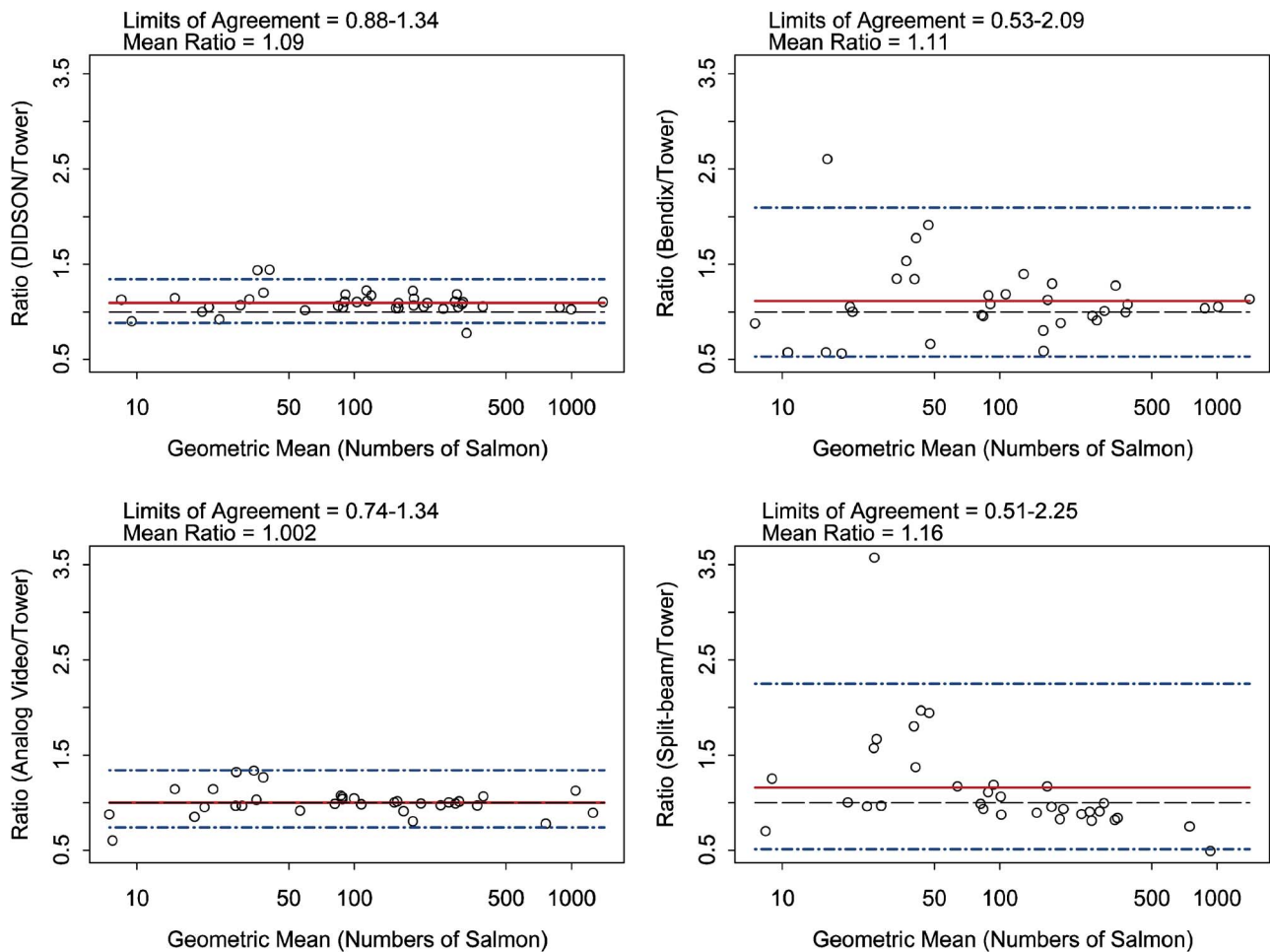


FIG. 7. (Color online) Assessment of agreement between tower counts paired with DIDSON, Bendix, split-beam, and video counts showing the mean ratio (solid line), a ratio of one (dotted line), and the limits of agreement (long and short dashes).

A variety of coefficients for the transmission loss equation were tested to find a curve that better fit each dataset (Fig. 10). For the Lake Washington dataset, the theoretical transmission loss curve (spreading loss+absorption) was $-[40 \text{ Log}(R)+2\cdot 0.31(R)]$. The equation that intersected more of the data points was $-[40 \text{ Log}(R)+2\cdot 0.02(R)]$; however, the flat slope of the points made it difficult to fit to a curve. For the Copper River plot, we took the liberty of adding two additional datasets, the plastic target data at 2 m and intensity measures from free-swimming salmon targets in the Copper River (3.0 m start range, 18.0 m window length, and $0.64 \mu\text{s}$ pulse length). The theoretical TL equation for the Copper River data was $-[40 \text{ Log}(R)$

$+2\cdot 0.35(R)]$, the equation that better fit the data points was $-[48 \text{ Log}(R)+2\cdot 0.75(R)]$. Changing the log and linear coefficients of the equation created a model that more closely matched the curve of the data.

IV. DISCUSSION

Counts of migrating fish from the DIDSON showed strong agreement with the tower counts. Our results indicate the DIDSON can be used to accurately count fish, within its range capacity, up to passage rates of 6000 fish/h. The split-

TABLE II. The average percent error (APE) for fish counts grouped by hourly passage.

Group	Hourly passage (Fish/h)	APE (%)	n^a
1	1–100	9.2	7
2	101–300	7.5	6
3	301–1000	3.0	11
4	1001–3000	3.8	7
5	>3000	4.7	3

^a n is the number of samples from each group.

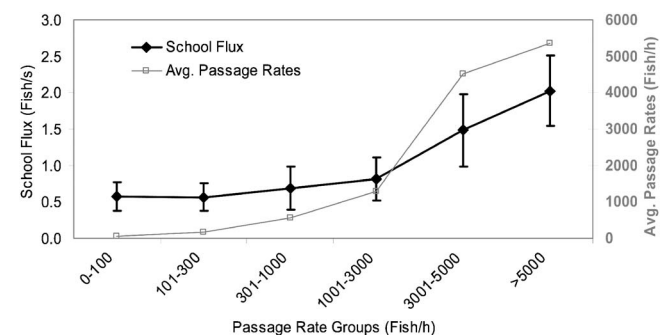


FIG. 8. School flux with standard deviation error bars for each hourly passage rate group. The average hourly passage for each group is plotted on the secondary axis (right).

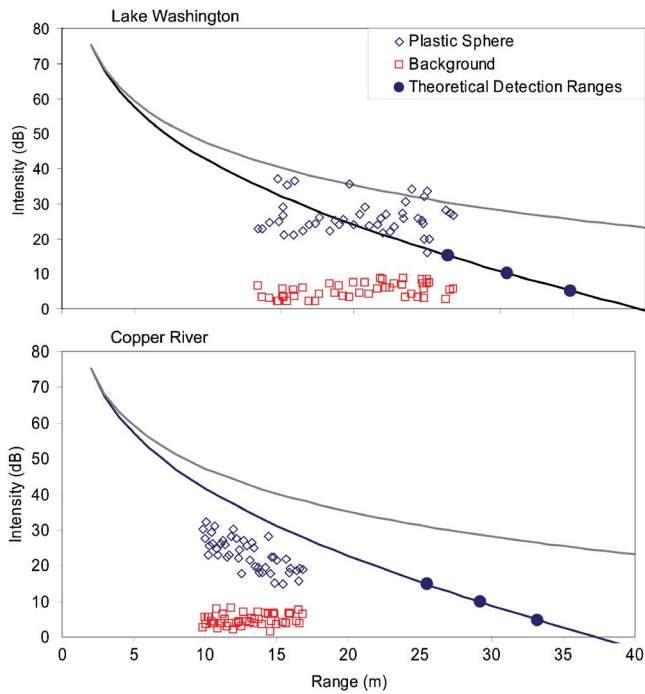


FIG. 9. (Color online) Intensity of the plastic sphere and background reverberation by range compared against theoretical spreading loss (gray line) and transmission loss (black line) for the plastic sphere. The circles represent the theoretical detection range of the plastic sphere where the transmission curve reaches 0, 5 and 10 dB above the average background intensity.

beam counts were the least similar to tower counts, saturating as passage rates exceeded 2000 fish/h, the same level as previously reported (Enzenhofer *et al.*, 1998). Individual fish became difficult to resolve in split-beam sonar echograms as hourly passage rates increased, while DIDSON fish images remained distinct. Overall, the total counts from the DIDSON were 6.9% higher than tower counts. However, we cannot simply conclude the DIDSON overestimated counts. The DIDSON was not subject to the same potential problems as the tower observer. Although our tower observer rated each count as good or very good, there were a few sample periods when clouds or light winds may have compromised the viewing quality. These difficulties and potential error sources are well outlined in Becker (1962). The same surface conditions affected the video images, but had no effect on DIDSON images. In addition, the tower counts were observed in real time, there was no opportunity to slow the counting or replay the image if the observer got behind. After this study, we became convinced that the DIDSON could be used to verify tower counts, especially during periods of reduced visibility, i.e., turbidity events and surface disturbances. Other advantages of using a DIDSON rather than real-time tower observers to count fish is that DIDSON images can be slowed down, replayed, or sped up depending on the level of fish passage. In addition, counting fish with the DIDSON has the potential to be automated.

The Bendix counter also performed well compared with tower counts, but like the DIDSON, the counts were slightly higher. The Bendix counter was developed and extensively tested at the Wood River site (Gaudet, 1990), so we expected good agreement between the Bendix and tower counts. The

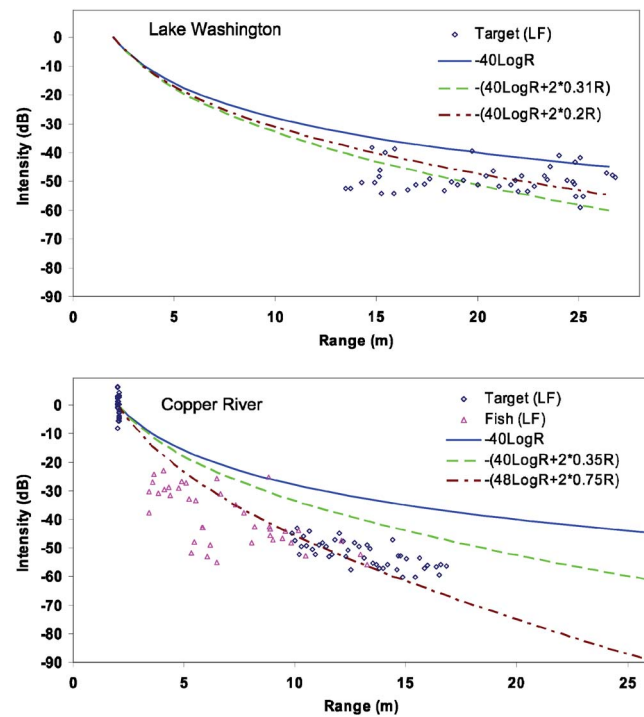


FIG. 10. (Color online) Theoretical and fitted transmission loss curves for Lake Washington (top) and Copper River (bottom). The Copper River data include measured intensities for the target at close range, and for free-swimming fish targets.

similarity between the counts from the Bendix counter and DIDSON will make it easier to substitute one type of counter for another. However, we cannot assume this similarity will cross over to other sites. If similar conditions prevail, i.e., fish travel close to the shoreline and near the river bottom, we would expect the two technologies to produce similar counts. Changing water levels and current flow may alter fish behavior. If fish rise up above the river bottom or if bottom obstructions prevent positioning the beam near the river bottom, we would expect to count fewer fish with the Bendix counter's narrow beam. To determine whether counts from the two technologies are similar enough to continue a historic dataset requires side-by-side testing at each site that encompasses the range of current velocities and water levels that may differentially affect counts.

Estimates of the relative counting precision between multiple DIDSON counters varied depending on passage rate. Although observers were more confident of their DIDSON counts when fish passage was low, these groups had the highest APE values. For similar low passage groups, Holmes *et al.* (2006) measured a higher precision, with APE values of 3.29% (group 1) and 2.04% (group 2) for the unconstrained fish. In the constrained portion of the same study (fish were backed up behind a weir and the gate opened periodically), passage rates were not stated, and the high 13.96% APE value for the low passage group was not explained. For comparison purposes, we examined the relative counting precision between tower observers using the data from Becker's Table 3 (1962). The relative counting precision was higher for the tower counts from this study with APE values of 2.38%, 1.64%, and 2.78% for passage rates

similar to our groups 3, 4, and 5, respectively (no passage rates lower than group 3 were reported). Neither the DIDSON nor tower estimates followed the simple trend of increased precision with increased passage observed by Holmes *et al.*

Holmes *et al.* (2006) states that fish density is not always reflected in fish-passage rates, and uses this statement to justify their classification of fish/event without including a time component for an event in the unconstrained portion of his study. Although not apparent at the low passage rates observed by Holmes *et al.* (less than 300 fish/h in the free-swimming fish study), we observed a strong positive correlation between the flux rate of individual schools and hourly passage rate. Sockeye salmon traveled in schools that grew increasingly larger as passage rates increased. At passage rates beyond 3000 fish/h, the gaps between schools disappeared and passage became continuous. We also observed that sockeye salmon tended to compact closer to shore as passage rates increased. The relationship between the flux rates of individual schools and hourly passage rate suggests that caution needs to be applied when extending the levels of agreement observed in this study outside the range of passage rates studied (i.e., 6000 fish/h).

The river bottom topography and beam geometry play a role in the production of acoustic shadows. As fish pass through the beam, shadows of the fish are often observed at further ranges. The effects of these shadows on fish counts are unknown; however, we have often detected fish traveling through the shadowed regions (Fig. 4). Obviously, close-range fish swimming directly in front of the transducer may block the beam entirely. A weir is recommended to prevent fish from traveling too close to (or behind) the transducer. Few fish traveled closer to the transducer than 3 m during this study. At 3 m, the horizontal beam width of the DIDSON's composite beam is 205 cm. The 50.8 cm average length of a Wood River sockeye salmon (West, 2003) occupies only a fraction of the composite beam, allowing a significant portion of the beam energy to travel beyond the fish and reflect off from fish at further ranges. Although shadows often obscured the bottom image, the comparison results indicated fish detection was unaffected by the shadowing, even at the high migration rates observed.

Multipathing, or sound that forward scatters from fish and reflects off bottom or surface before returning to the transducer, was not prevalent at the Wood River site. Multipath echoes take the form of multiple traces of the same fish displayed at a further range, or fountains of echoes that cloud the echogram from the range of the fish to the end range. The extra tracings are usually lower in amplitude and less coherent than the direct reflection, although in some cases, boundary focusing effects can create higher amplitude returns from an indirect pathway. We have frequently observed multipathing on split-beam echograms at other sites, and although more difficult to discern, occasionally in oscilloscope displays generated by the Bendix counter. At the Wood River site, it is likely the relationship of the sonar beams and the river environment was not conducive to returning forward-scattered echoes. However, we have used DIDSON at other sites where multipathing occurred with every fish that en-

tered the beam of a split-beam system, and did not observe the phenomenon in DIDSON images. The geometry of the DIDSON may be a factor. Raising it higher on the mount with a greater downward tilt angle may result in multipath images that reflect off bottom or surface, but not at an angle that returns the sound energy to the transducer. We have observed minimal multipathing at a site where the DIDSON was deployed with a geometry more similar to traditional sonars. The multipath images were low in amplitude and only visible at the lowest threshold setting. Another potential reason for the limited multipathing is the higher frequency beam, which scatters off smaller objects including small particles in the water, and is more likely to lose intensity as it travels the indirect path back to the transducer.

The maximum detection ranges for the plastic sphere at Lake Washington and the Copper River were lower than expected. Holmes *et al.* (2006) concluded, without testing, that the effects of water temperature and turbidity on signal attenuation and scattering at the ranges covered by the DIDSON (a stated 15 m in high-frequency mode and 50 m in low-frequency mode) are negligible. Many factors can limit a target's detection range. At the high DIDSON frequencies, the small wavelengths scatter off smaller objects, and can rapidly lose intensity with range. In addition, the target must be visible above the background reverberation levels. At Lake Washington, the intensity of the last target image was above the transmission loss curve, and the SNR was over 20 dB. The transmission loss curve (Fig. 9) suggests that we would not be able to detect the target beyond 30 m, where the background reverberation level is expected to approach the target intensity. The limiting factor appeared to be the size and shape of the target, which became similar in size and difficult to distinguish from background reverberation. A more oblong target, like a salmon, is likely to be detected farther down range. Although the theoretical TL curve tends to flatten with range, the expected loss of intensity from the start to the end range of the target is 16 dB. At Lake Washington, the observed change in intensity within the window length sampled was almost zero (a regression slope through the data points was -0.06). The reason for the lack of a decline was unclear. We considered multipathing issues, which could potentially increase a target's intensity if the multipath image appeared within the 7 cm down range resolution of the direct reflection. The water column was deeper compared to riverine sites, so multipathing should have been less of an issue, and we did not observe a lengthening of the echoes. The high degree of scatter in the intensity measures may have obscured a decline.

At the Copper River, all target intensity measures for the plastic sphere remained below the theoretical TL curve and declined with range. The additional loss could not be modeled by simply increasing the linear coefficient, indicating the loss is a more complex function. The high levels of turbidity and associated scattering loss are likely the reasons for the substantial signal loss. Other reasons may stem from the nature of the experiment. Field tests are always less reliable than laboratory ones. At the Copper River, it was necessary to hold the boat in place against a strong current while attempting to suspend the target in the beam. The current is

always tugging against the target, making depth measures questionable. A more controlled experiment is needed if we are to model the effects of turbidity on intensity loss. What we hoped to gain was knowledge of the DIDSON's range limitation at the most turbid site where we deploy sonars, and a method to obtain reasonable transmission loss coefficients at a field site. Targets have been viewed at considerably longer ranges with lower frequency sonars. In the turbid Yukon River (Maxwell, 2000), chum salmon have been detected beyond 200 m from the transducer using a 120 kHz sonar. If ensonification at longer ranges is required, a lower frequency sonar is a better option. The Copper River is highly turbid. In less turbid rivers, we would expect to detect the same target farther offshore.

Since this study, Sound Metrics added a transmission loss algorithm with the option of adding gain from the start range or subtracting from the end range using the function:

$$TL = 2 * \alpha(R) + N \text{Log}(R),$$

where α is the absorption coefficient (linear function), R is the end range, and N is the spreading loss coefficient (non-linear function). Ideally, we should be able to empirically measure total signal loss in the sampling environment. Signal loss depends on many factors including an unknown amount of spreading loss as the beam is squeezed between the river's boundaries, absorption, scattering from macroscopic particles, shadowing by rocks or other river bottom obstructions, the combination of river bottom topography and beam geometry which determines what portion of the beam the fish will swim through, and other unknown factors. Rather than understanding the contribution from each factor, we should be able to smooth the echo intensity down range by: 1) setting up the sonar with the same parameters and transducer geometry used for sampling fish; 2) dragging a spherical target down range through the sonar beam near the river bottom, where we expect fish to migrate; 3) measuring the intensity of the target image incrementally down range; 4) plotting the target intensities against the TL curve; and 5) altering the linear and nonlinear coefficients to best match the measured intensity curve. The resulting coefficients would represent the actual sampling environment, rather than a theoretical one.

V. CONCLUSIONS AND FUTURE WORK

To achieve high detection levels, the DIDSON's geometry and settings are important. Although the DIDSON can be tilted across a wider tilt angle before compromising fish detection, it is important to understand the relationship between the river bottom topography and the position of the DIDSON. Maxwell and Smith (2007) describe how the DIDSON can be used to obtain profiles of the river bottom. By plotting the river bottom profile and a model of the transducer beam within the same coordinate system, the effects of potential tilt angles can be viewed. Using the DIDSON image of the river bottom only to determine the best tilt angle can be misleading. In situations where this was done, we have observed a tendency to tilt the DIDSON to a greater tilt angle, pushing the maximum response axis deep into the

river bottom (Maxwell and Smith, 2007). We recommend positioning the DIDSON closer to the river bottom and angling the beam similar to the angle of the river bottom. Although there is greater potential for shadowing, this beam position will extend the effective range and provide greater coverage of the near-bottom regions at close range. It is also important to consider the threshold and intensity settings. Because of the large amounts of signal subtracted when the background subtraction feature is used, it is important to keep the threshold as low as possible and adjust the intensity to brighten the fish images when using this feature.

The DIDSON has many advantages over the Bendix counter and split-beam sonar. As shown in Fig. 3, and modeled in Fig. 2(a), there was some interaction of the DIDSON beam with the gravel of the river bottom. For the Bendix counter and split-beam sonar, it is critical that the beam fit entirely between the river's boundary layers. If strong bottom echoes are present at the range of fish passage, the bottom echoes will obscure fish echoes. It is less important for the DIDSON beam to fit within the water column. We intentionally tilted the DIDSON down far enough to keep the beam's upper edge below the surface, while pushing the remaining beam into the river bottom. The narrow, horizontal DIDSON beams finely resolve the river bottom substrate. Because the DIDSON displays a moving, video-like, image rather than a static one, fish can be detected moving across the static bottom. The fine resolution of the images makes it easier for the bottom-subtraction feature to remove the static echoes, while retaining the moving ones. Surface disturbances and moving sand may interfere with fish detection. In very narrow water columns, it may be necessary for the beam to interact with the surface in order to position the maximum response axis of the beam above the river bottom. As long as surface disturbances are minimal, the image quality may remain usable. Currently, condenser lenses, which reduce the size of the vertical beam, are available as add ons to existing DIDSONs. In very narrow water columns, we recommend adding a condenser lens and reducing the receiver gain when sampling nearshore. More energy is transmitted with the narrow lens, so reducing the receiver gain will create a better image. Large rocks may be a problem if they completely shadow fish passage regions. In these situations, it is advisable to remove them, or find a new site.

Other advantages of the DIDSON include better resolution of multiple targets, better coverage of the water column, and accurate direction of travel information. Multiple fish traveling at a similar range are distinguishable because they weave in and out as they swim, allowing the viewer to separate the individual fish images. In strong current situations where fish are migrating close to shore, the DIDSON's wider vertical beam provides greater coverage of the water column at close range. The narrow, horizontal beams allow direction of fish travel to be accurately determined, while this same information is often poor from the split-beam sonar, and absent from the Bendix counter.

There are numerous continued avenues of study for the DIDSON. Two autotracking algorithms included with the DIDSON initially showed promise, but Baumgartner *et al.* (2006) showed that the automated counts were lower than

manual counts; while in our experience, the automated counts were usually substantially higher. Shadowing from passing fish is a large part of the problem. As the shadows move across the image, the river bottom echoes flicker off and on resembling movement. To date, the autotrackers tested are unable to distinguish this flickering from actual fish at the Wood River site. Echogram counting, which provides range and time information for targets, is available with the DIDSON software and may be faster in some situations. In addition, a manually counting algorithm has been added since this study that allows the user to mark fish on the electronic video image, and then export information about the location of the marks. This may prove to be a more accurate way to keep track of the fish that have been counted. Potentially, the DIDSON's fish-sizing algorithms may be used to separate fish species with distinct length curves. Because of the range limitations of the standard DIDSON, a lower frequency (1.1 and 0.7 MHz), long-range DIDSON was developed. Initial validation tests have shown good agreement with visual counts (Galbreath and Barber, 2005), but additional testing needs to be done at higher fish passage rates.

ACKNOWLEDGMENTS

The authors would like to thank Ed Belcher, Bill Hanot, and Joe Burch for providing a DIDSON for this study, training us, and making numerous software modifications to suit our needs. John Holmes, Tim Mulligan, Scott Raborn, Lowell Fair, and April Smith reviewed various drafts of the manuscript. Dan Huttunen, Don Degan, Lee McKinley, Anna-Maria Mueller, Ted Otis, Debby Burwen, and Jenny Cope assisted with the fieldwork and/or data processing. Peter Dahl wrote the program to calculate theoretical target strength values for the tungsten carbide sphere. Jay Johnson, Brian Marston, Ken Vartan, Corey Schwanke, and Joe Winter provided logistical support. The ADF&G and the U.S. Forest Service provided funding.

- Baumgartner, L. J., Reynoldson, N., Cameron, L., and Stanger, J. (2006). "Assessment of a Dual-frequency Identification Sonar (DIDSON) for application in fish migration studies," NWS Department of Primary Industries Fisheries, Final Report Series No. 84, ISSN 1449-9967, Narrandera.
- Becker, C. D. (1962). "Estimating red salmon escapements by sample counts from observation towers," *Fishery Bulletin* 192: Vol. 61. Fish and Wildlife Service, Washington.
- Belcher, E. O., Hanot, W., and Burch, J. (2002). "Dual-frequency identification sonar," in *Proceedings of the 2002 International Symposium on Underwater Technology*, Tokyo, April 16-19, pp. 187-192.
- Bland, J. M., and Altman, D. G. (1986). "Statistical methods for assessing agreement between two methods of clinical assessment," *Lancet* 327, 307-310.
- Brabets, T. P. (1997). "Geomorphology of the Lower Copper River, Alaska," U. S. Geological Survey Professional Paper 1581.
- Brazil, C. (2007). "Sonar enumeration of Pacific salmon escapement into the Nushagak River, 2003," ADF&G, Commercial Fisheries Division, Fishery Data Series, No.07-37, Anchorage.
- Chilton, D. E., and Beamish, R. J. (1982). "Age-determination methods for fishes studied by the groundfish program at the Pacific Biological Station," *Can. Spec. Publ. Fish. Aquat. Sci.* 60, 10-12.
- Cronkite, G. M. W., and Enzenhofer, H. J. (2002). "Observations of controlled moving targets with split-beam sonar and implications for detection of migrating adult salmon in rivers," *Aquat. Living Res.* 15(1), 1-11.
- Dahl, P. H., Geiger, H. J., Hart, D. A., Dawson, J. J., Johnson, S. V., and Degan, D. J. (2001). "The environmental acoustics of two rivers and its relation to salmon counting sonars," Technical Report. Applied Physics Laboratory, University of Washington, APL-UW TR 2001, Seattle.
- Daum, D. W., and Osborne, B. M. (1998). "Use of fixed-location, split-beam sonar to describe temporal and spatial patterns of adult fall chum salmon migration in the Chandalar River, Alaska," *N. Am. J. Fish. Manage.* 18, 477-486.
- Enzenhofer, H. J., Olsen, N., and Mulligan, T. J. (1998). "Fixed-location hydroacoustics as a method of enumerating migrating adult Pacific salmon: Comparison of split-beam acoustics vs. visual counting," *Aquat. Living Res.* 11(2), 61-74.
- Faran, J. J., Jr. (1951). "Sound scattering by solid cylinders and spheres," *J. Acoust. Soc. Am.* 23, 405-418.
- Francois, R. E., and Garrison, G. R. (1982). "Sound absorption based on ocean measurements. Part II: Boric acid contribution and equation for total absorption," *J. Acoust. Soc. Am.* 72, 1879-1890.
- Galbreath, P. F., and Barber, P. E. (2005). "Validation of a long-range dual frequency identification sonar (DIDSON-LR) for fish passage enumeration in the Methow River," Final Report—PSC Southern Fund 2004/2005 Project.
- Gaudet, D. M. (1990). "Enumeration of migrating salmon populations using fixed-location sonar counters," *Rapports et Proces-Verbaux des Reunions, Conseil International pour l'Exploration de la Mer* 189, 197-209.
- Holmes, J. A., Cronkite, M. W., Enzenhofer, H. J., and Mulligan, T. J. (2006). "Accuracy and precision of fish-count data from a "Dual-frequency Identification Sonar" (DIDSON) imaging system," *ICES J. Mar. Sci.* 63, 543-555.
- Lilja, J., Keskinen, T., Marjomaki, T. J., Valkeajarvi, P., and Karjalainen, J. (2003). "Upstream migration activity of cyprinids and percids in a channel, monitored by a horizontal split-beam echosounder," *Aquat. Living Res.* 16(3), 185-190.
- Lin, L., Hedayat, A. S., Sinha, B., and Yang, M. (2002). "Statistical methods in assessing agreement: Models, issues, and tools," *J. Am. Stat. Assoc.* 97, 257-270.
- Maxwell, S. L. (2000). "Yukon River sonar project report, 1998," ADF&G, Commercial Fisheries Division, Regional Information Report No. 3A00-4, Anchorage.
- Maxwell, S. L., and Gove, N. E. (2004). "The feasibility of estimating migrating salmon passage rates in turbid rivers using a dual-frequency identification sonar (DIDSON) 2002," ADF&G, Commercial Fisheries Division, Regional Information Report No. 2A04-05, Anchorage.
- Maxwell, S. L., Degan, D., Smith, A. V., McKinley, L., and Gove, N. E. (2008). "The feasibility of using a split-beam sonar to estimate salmon passage on the Nushagak River as a potential replacement for an echo-counting Bendix Sonar," Alaska Department of Fish and Game, Fishery Manuscript No. 08-XX, Anchorage (in press).
- Maxwell, S. L., and Smith, A. V. (2007) "Generating river bottom profiles with a dual-frequency identification sonar (DIDSON)," *N. Am. J. Fish. Manage.* 27(4).
- McIntyre, J. K., Beauchamp, D. A., Mazur, M. M., and Overman, N. C. (2006). "Ontogenetic trophic interactions and benthopelagic coupling in Lake Washington: Evidence from stable isotopes and diet analysis," *Trans. Am. Fish. Soc.* 135, 1312-1338.
- Miller, J. D., and Burwen, D. (2002). "Estimates of Chinook salmon abundance in the Kenai River using split-beam sonar, 2000," ADF&G, Sport Fish Division, Fishery Data Series No. 02-09, Anchorage.
- Moursund, R. A., Carlson, T. J., and Peters, R. D. (2003). "A fisheries application of a dual-frequency identification sonar, acoustic camera," *ICES J. Mar. Sci.* 60, 678-683.
- Richards, S. D., Heathershaw, A. D., and Thorne, P. D. (1996). "The Effect of suspended particulate matter on sound attenuation in seawater," *J. Acoust. Soc. Am.* 100, 1447-1450.
- Simmonds, E. J., and MacLennan, D. N. (2005). *Fisheries Acoustics: Theory and Practice* (Blackwell Science Ltd., Oxford, UK), pp. 39, 69.
- Smith, A. V., and Maxwell, S. L. (2008). "The feasibility of using a split-beam sonar to estimate salmon passage in the Kenai River as potential replacement for an echo-counting Bendix sonar," ADF&G, Commercial Fisheries Division, Fisheries Manuscript No. 08-XX, Anchorage (in Press).
- Tiffan, K. F., Rondorf, D. W., and Skalicky, J. J. (2004). "Imaging fall Chinook salmon redds in the Columbia River with a dual-frequency identification sonar," *N. Am. J. Fish. Manage.* 24, 1421-1426.

West, F. W. (2003). "Abundance, age, sex, and size statistics for Pacific salmon in Bristol Bay, 2002," ADF&G, Commercial Fisheries Division, Regional Information Report No. 2A03-23.

Westerman, D. L., and Willette, T. M. (2006). "Upper Cook Inlet salmon escapement studies, 2004," ADF&G, Commercial Fisheries Division,

Fishery Data Series No. 06-49, pp. 80–82.

Xie, Y., Gray, A. P., Martens, F. J., Boffey, J. L., and Cave, J. D. (2005). "Use of dual-frequency identification sonar to verify split-beam estimates of salmon flux and to examine fish behavior in the Fraser River," Pacific Salmon Commission Technical Report No. 16, 58 p.

Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999–2002

Kathleen M. Stafford^{a)}

Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

David K. Mellinger

Oregon State University and NOAA Pacific Marine Environmental Laboratory, 2030 SE Marine Science Drive, Newport, Oregon 97365

Sue E. Moore

NOAA Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, Washington 98115

Christopher G. Fox

National Geophysical Data Center, E/GC 325 Broadway, Boulder, Colorado 80305-3328

(Received 7 March 2007; revised 20 August 2007; accepted 25 September 2007)

Five species of large whales, including the blue (*Balaenoptera musculus*), fin (*B. physalus*), sei (*B. borealis*), humpback (*Megaptera novaeangliae*), and North Pacific right (*Eubalaena japonica*), were the target of commercial harvests in the Gulf of Alaska (GoA) during the 19th through mid-20th Centuries. Since this time, there have been a few summer time visual surveys for these species, but no overview of year-round use of these waters by endangered whales primarily because standard visual survey data are difficult and costly. From October 1999–May 2002, moored hydrophones were deployed in six locations in the GoA to record whale calls. Reception of calls from fin, humpback, and blue whales and an unknown source, called Watkins' whale, showed seasonal and geographic variation. Calls were detected more often during the winter than during the summer, suggesting that animals inhabit the GoA year-round. To estimate the distance at which species-diagnostic calls could be heard, parabolic equation propagation loss models for frequencies characteristic of each of each call type were run. Maximum detection ranges in the subarctic North Pacific ranged from 45 to 250 km among three species (fin, humpback, blue), although modeled detection ranges varied greatly with input parameters and choice of ambient noise level.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799905]

PACS number(s): 43.30.Sf, 43.80.Ka [WAA]

Pages: 3378–3390

I. INTRODUCTION

The Gulf of Alaska (GoA) is a large marine ecosystem with complex interannual and seasonal variability in ocean circulation and productivity (Whitney and Freeland, 1999; Brickley and Thomas, 2004). This productivity supports large numbers of fish, sea birds, and marine mammals. Most of what is known of marine mammals in the GoA has focused on pinnipeds with relatively little attention paid to large baleen whales. In part, this may be because five species of large whale, including blue (*Balaenoptera musculus*), fin (*B. physalus*), sei (*B. borealis*), humpback (*Megaptera novaeangliae*), and North Pacific right (*Eubalaena japonica*), were greatly depleted as the targets of commercial harvests in the GoA from the 19th to the mid-20th Centuries (Springer *et al.*, 1999; Shelden *et al.*, 2005) and are presently listed as endangered under United States law (NMFS 2005). North Pacific right whales were essentially extirpated by the 1930s and have rarely been seen in this region since then (Brownell *et al.*, 2001; Waite *et al.*, 2003; Mellinger *et al.*, 2004b). Very few large whales (<50 fin whales and <200 humpback whales) were seen during two postwhaling ship-

board and aerial surveys for marine mammals conducted along the eastern Aleutians in the near-shore Gulf of Alaska in the early and mid-1980s, despite searching in regions and during seasons in which whalers were most successful (Rice and Wolman, 1982; Stewart *et al.*, 1987). However, recent surveys in the western GoA and the Aleutians have provided data to support abundance and rate-of-increase estimates for fin, humpback and minke whales suggesting that these species may be increasing there (Zerbini *et al.*, 2006). Additionally, localized studies of humpback whales have been undertaken in the near-shore (on shelf) GoA (Baker *et al.*, 1985; Witteveen *et al.*, 2004). Offshore areas in the deep GoA have been less well monitored because of the logistical difficulties of working in such areas.

Acoustic surveys of cetaceans are a powerful means of identifying species' seasonal occurrence (Watkins *et al.*, 2000a, b; Moore *et al.*, 2006) and regional dialects (Stafford *et al.*, 2001) in remote habitats. During the past decade, large-scale acoustic surveys have been conducted in the North Atlantic (Clark and Charif, 1998; Charif *et al.*, 2001; Nieu Kirk *et al.*, 2004), eastern tropical Pacific (Stafford *et al.*, 1999a), Indian Ocean (Stafford *et al.*, 2004), and off the Antarctic Peninsula (Širović *et al.*, 2004). In the North Pacific, acoustic surveys have been used to define boundaries

^{a)}Electronic mail: stafford@apl.washington.edu

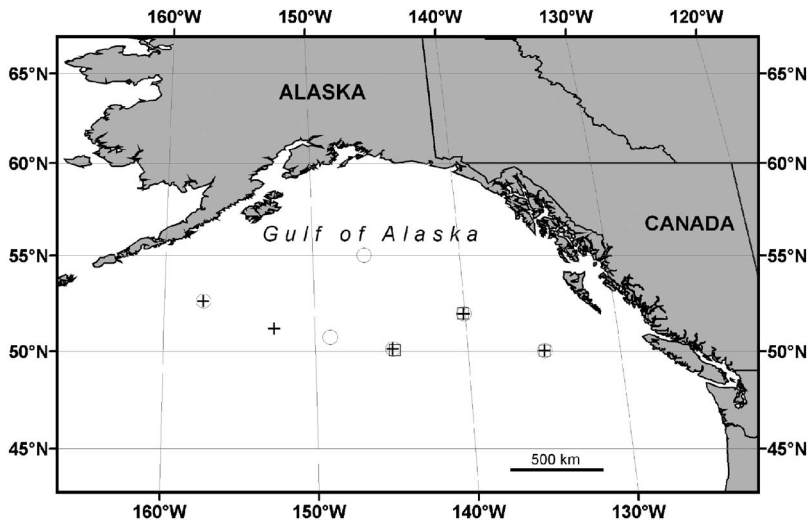


FIG. 1. Locations of moored hydrophones in the Gulf of Alaska (1999 ○; 2000 +; 2001 □).

for a northeast Pacific population of blue whales (Stafford *et al.*, 1999b), to determine seasonality and broad habitat characteristics of blue whales over a large region of the subarctic northwest Pacific (Moore *et al.*, 2002), and to document seasonal occurrence of large whales over basin scales (Watkins *et al.*, 2000a, b; Stafford *et al.*, 2001).

Assessments of endangered whales in the GoA by standard visual surveys are difficult and costly to conduct. There are summaries of seasonal occurrence of calls for blue, fin, North Pacific right and sperm whales (*Physeter macrocephalus*) in and near the GoA (Watkins *et al.*, 2000a, b; Moore *et al.*, 1998, 2002; Mellinger *et al.*, 2004b), but no multispecies analysis of year-round use of these waters. None of the acoustic summaries accounted for transmission loss, which varies among the diagnostic calls used for assessment of species' occurrence.

Automatic detection methods were used to examine the data set presented here for sperm and right whale calls (Mellinger *et al.*, 2004a, b) and an account has been published of visual inspection of the first 18 months of data for blue whale calls (Stafford, 2003). Here we expand on these analyses to present a multispecies (blue, fin, and humpback) acoustic survey of the GoA that includes estimates of the detection distances of diagnostic calls produced by each species. We also include results for a unique 52-Hz signal (Watkins *et al.*, 2004). As the late W. A. Watkins first documented this call type (Watkins *et al.*, 2000b), it is hereafter referred to as "Watkins' whale."

II. METHODS

A. Autonomous hydrophones

In October 1999, the National Oceanic and Atmospheric Administration's Pacific Marine Environmental Laboratory deployed moored autonomous hydrophones in the GoA to record whale calls (Fig. 1). Details of the instruments are provided in Fox *et al.* 2001. Briefly, each mooring package consisted of an anchor, an acoustic release, an International Transducer Corporation 1032 hydrophone (flat response from 1 Hz to 32 kHz and sensitivity of -192 dB re 1 V/ μ Pa; with preamplifier, system sensitivity was

-160 dB re 1 V/ μ Pa), a recorder in a pressure-resistant titanium case, and flotation. The instruments were moored with the hydrophones suspended in the deep sound channel at depths of 340–450 m. Data were archived onboard the instruments as digital files. The hydrophones sampled data at 1000 Hz with low-pass antialias filtering above 440 Hz, which is sufficient bandwidth to record the low-frequency calls of blue and fin whales (cf. Rivers, 1997; Watkins, 1981) and to record the low frequency components of both North Pacific right and humpback whales (Thompson *et al.*, 1986; McDonald and Moore, 2002). Because of filter rolloff, sounds up to approximately 470 Hz were analyzable.

Instruments were deployed for roughly one year each between 1999 and 2002 (Table I). Six hydrophones were deployed initially in October 1999 and were recovered in May 2000 (Fig. 1). Data were extracted and five of the instruments were redeployed. In June 2001, these five recorders were recovered and their data extracted, and three were redeployed. The last three instruments were recovered from the GoA in July 2002.

B. Call occurrence

A subsample of 20% of hours of available data was randomly chosen for analysis. The same hours of the day were examined for each hydrophone, resulting in a total sample of 17 185 hours of the ~ 85 000 h available from the fourteen recovered instruments. The sampled hours were manually examined as spectrograms (512 point Fast Fourier transform, 75% overlap, Hanning window, for a filter bandwidth of 8 Hz), and the presence (+) or absence (–) of a number of call types were tallied. The percentage of hours per month with calls was plotted for each species to explore seasonal trends. Averages by month over all years were used to produce seasonal histograms for each species.

Diagnostic call types whose bandwidths were within the 1–470 Hz range of the instruments included the well-documented moans of blue whales and pulses of fin whales (Watkins, 1981; Thompson and Friedl, 1982). Also falling into this frequency range were the 50 Hz calls of Watkins' whale (Watkins *et al.*, 2004). Humpback whale vocalizations are extremely varied and are typically grouped into three

TABLE I. Deployment details for hydrophones placed in the Gulf of Alaska.

Year deployed	Location	Depth (m)	Dates	Remarks
1999	50°02' N 134°59' W	349	21 Oct 1999–26 Jun 2000	
1999	50°07' N 144°55' W	340	11 Oct 1999–14 Jun 2001	
1999	50°44' N 149°02' W	367	10 Oct 1999–1 Jun 2000	
1999	51°57' N 140°05' W	349	19 Oct 1999–6 Jun 2001	
1999	52°38' N 157°24' W	340	6 Oct 1999–11 Jun 2001	
1999	55°02' N 146°31' W	367	17 Oct 1999–25 May 2000	Not redeployed
2000	50°02' N 135°00' W	450	23 May 2000–26 Jun 2001	Failed 30 d into deployment
2000	50°07' N 144°56' W	405	2 June 2000–14 Jun 2001	
2000	51°11' N 152°42' W	404	31 May 2000–15 Dec 2000	Failed midway through deployment
2000	51°58' N 140°06' W	404	24 May 2000–6 Jun 2001	
2000	52°38' N 157°23' W	410	29 May 2000–11 Jun 2001	
2001	50°02' N 134°59' W	441	27 July 2001–31 May 2002	
2001	50°03' N 144°47' W	340	2 Sep 2001–31 May 2002	
2001	51°58' N 140°07' W	439	28 July 2001–31 May 2002	

general categories: song (20 Hz–8 kHz), social sounds (50 Hz–10 kHz), and feeding sounds (~20 Hz–2 kHz) (Payne and McVay, 1971; Thompson *et al.*, 1986; Cerchio and Dahlheim, 2001). The low-frequency parts of humpback whale vocalizations, including song, were detected on these instruments. These were identified by long repeated sequences of short duration frequency- and amplitude-modulated phrases with energy from approximately 100 to 470 Hz (Payne and McVay, 1971).

C. Detection distance range modeling

To estimate the maximum detection range for the diagnostic call of each species, transmission loss (TL) was modeled for characteristic frequencies of each call (Table II). Low-frequency attenuation has been shown to be region-specific based on environmental properties, especially temperature layering (Kibblewhite *et al.*, 1977; Chow and Browning, 1983), and to increase with latitude, probably due to the shoaling of the deep sound channel axis (Chow and Turner, 1982). Almost all propagation loss studies have de-

ployed explosive charges to actively measure TL between a source and a receiver. As such an approach is impossible for whale calls, we modeled TL based on estimates of call source level and ambient noise from the published literature. Estimates of TL also requires information on sound speed profiles and bathymetry along the transmission path. Because none of these parameters were measured for the instrument sites in the GoA during the period of this study, they were estimated, as follows:

- *Frequency, source level, and depth of the vocalizing whale.* These were obtained from the literature (Table II). We assumed that sounds are emitted omni-directionally from the animals, which is probably reasonable for low-frequency baleen whales.
- *Depth of the receiver (hydrophone).* This is known from an acoustic ping-and-response survey taken after the instrument was deployed. See Table I.
- *Sound speed profile along the transmission path from source to receiver.* This was obtained from the climatic (averaged) data from the U.S. Naval Oceanographic Of-

TABLE II. Parameters used in detection distance modeling. Vocalization depth was held at 20 m for all species except fin whale (50 m, Watkins *et al.* 1987). Numbers in parentheses are the values used for propagation loss modeling. For ambient noise levels, we used values given, in the order listed in the table, from Wenz (1962), Urick (1983, moderate and high shipping) and Andrew *et al.* (2002, median values).

Species	Frequency (Hz)	SL (dB	Ambient noise levels at target frequency (dB re 1 $\mu\text{Pa}^2/\text{Hz}$)	References
		re 1 μPa at 1 m)		
Blue whale—eastern	16	180	74, 79, 83, 90	Thode <i>et al.</i> (2000)
Blue whale—western	20	180	75, 82, 85, 91	Assumed to be same as above
Watkins' whale	50	175	75, 82, 83, 90	Assumed to be same as blue whale third harmonic as per Thode <i>et al.</i> (2000)
Fin whale	20–35 (25)	171	76, 82, 85, 91	Watkins <i>et al.</i> 1987; Charif <i>et al.</i> (2002)
Humpback whale	100–470 ^a (350)	160	55, 60, 66, 75	Thompson <i>et al.</i> (1986); Au <i>et al.</i> (2006)

^aThe upper limit indicated here for humpback whale is the upper limit of our instrument rather than the upper limit of this species' repertoire.

face's General Digital Environmental Model (GDEM; Teague *et al.*, 1990) Version 3.0 (<https://128.160.23.42/gdemv/gdemv.html>). GDEM provide monthly averages at 0.25° intervals; we used a subset of these profiles at 0.5° intervals for latitudes from 50° N to 58° N and from 130° W to 158° W.

- *Ocean bottom composition along the transmission path.* This was determined from bore-hole data from hole 1023A from leg 168 of the Ocean Drilling Project (<http://www-odp.tamu.edu>).
- *Bathymetry along the transmission path.* This came from the GEODAS database (http://www.ngdc.noaa.gov/mgg/gdas/gd_designagrid.html) of the National Geophysical Data Center.
- *Ambient noise level.* Estimates of ambient noise levels were taken from three sources, including the most recent data available for the North Pacific, as provided in Andrew *et al.* (2002) for a region off central California; the Wenz curves (Wenz, 1962); and levels for moderate and high shipping presented by Urlick (1983).
- *Detectability of a call in noise.* At what signal-to-noise ratio is a call just detectable *visually* on a spectrogram? We selected several fin, blue, and humpback calls of that we considered just detectable and measured their signal-to-noise ratio (SNR). The ratio was calculated using only the noise in the frequency band of the call, not all of the noise in the recording at that time. The ratio of the very loudest part of the call to this noise level was the SNR. The SNRs of just-detectable fin, blue, and humpback calls measured this way were 5–8 dB. As a conservative value, 8 dB was used for the detection threshold SNR of a just-detectable call.
- The parabolic equation acoustic propagation model random-access memory (Collins, 1993, 1995) was used. Given the environmental and other parameters described above, this model provides propagation loss estimates along the range-depth plane from source to receiver. To explore horizontal range independence of measurements, we modeled propagation loss for species-specific frequencies (Table II) along four transects at 90° intervals to a distance of 400 km at 0.5° steps for the portions of the GoA where water depth along the transects was greater than 2000 m. For locations within 400 km of the continental shelf, some transects were less than 400 km. Because propagation loss can vary substantially with small changes in depth and range, loss was averaged over a small region of the range-depth plane, a region centered at the desired range/depth point and extending from 5% below to 5% above it in both depth and range. Loss measurements were iterated for four months (January, April, July, and October), resulting in more than 5200 transects per frequency. As no significant seasonal differences were found, these were combined to produce overall detection probability plots for each whale species in the GoA.

III. RESULTS

Many recorded sounds could be definitively attributed to fin, humpback, and blue whales. The source of one of the

common biological sounds, Watkins' whale, is not known with certainty, although the call has characteristics similar to eastern North Pacific blue whale sounds. Reception of calls from each species and Watkins' whale showed seasonal and geographic variation. Although the overall occurrence of each of these sound types varied somewhat among years, the seasonal patterns among years were similar and we therefore present mean annual occurrence over all years for each site.

A. Fin whales

Fin whale regular pulses [Watkins, 1981, Fig. 2(a)] were the most frequently recorded cetacean call in the GoA. In a few instances, irregular "social" sounds were recorded; as these were recorded infrequently they were combined with the regular pulses. Fin whale calls were recorded from mid-summer through early winter, with highest call occurrence rates from August through December and lowest occurrence from February through July [Figs. 3(a)–3(g)]. Call occurrence increased from July to a peak in October through December and then tapered off through March. Although the seasonal pattern of fin whale calls was roughly the same on all hydrophones, the overall occurrence of fin whale pulses was lowest in 1999 and highest in 2000, with 2001 occurrence falling in between the other two years. There was no obvious geographic differences; hydrophones closer to coast had similar seasonal patterns and percent occurrence as the mid-Gulf hydrophones [Figs. 3(a)–3(g)].

B. Humpback whales

Sounds of humpback whales [Fig. 2(b)] were recorded seasonally from fall through winter, but most often from September through February [Figs. 3(a)–3(g)]. As with fin whales, sounds were recorded more often in 2000 and 2001 than in 1999. We qualified most of these sounds as "song" based on characteristics that included the repetition of sounds such as cries, whoops and moans in long patterns (Norris *et al.*, 1999). Calls were recorded on all hydrophones, but most frequently on the instrument at 53° N 157° W which was the only one on which humpback calls were recorded in every month of the year. They were recorded from the beginning of the dataset at the end of October 1999 through February 2000 on all hydrophones. Calls were not recorded again until early September 2000, with the exception of calls at 53° N 157° W in early June 2000. In 2000–2001, similar call rates occurred on the instruments at 53° N 157° W and 50° N 150° W and the instruments at 51° N 140° W and 50° N 145° W. In the first two locations, sounds were detected at high levels from October to mid-February, whereas in the latter two locations calls were detected later (mid-November) but also decreased in mid-February and were seldom seen from March through the end of the deployment in June 2001. Humpback whale calls were recorded from November through February during 2001–2002 on the three recovered instruments. The instrument at 53° N 157° W, which was the site where humpback whales were recorded most often, was not recovered in 2002.

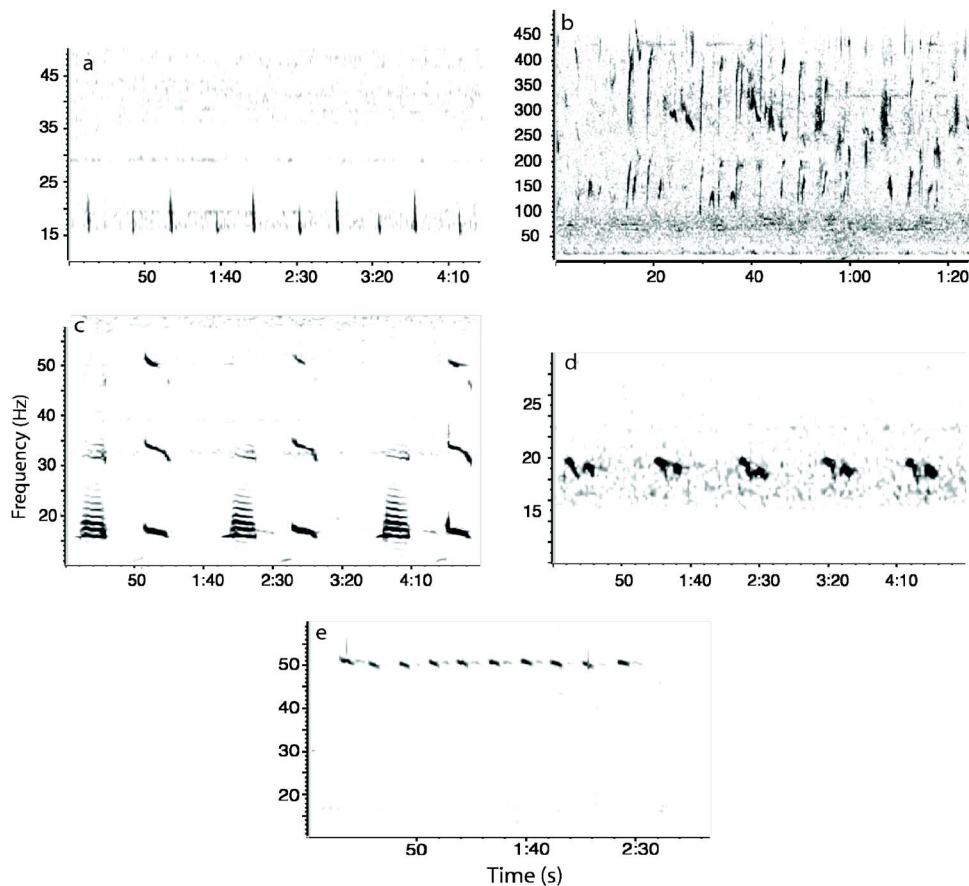


FIG. 2. (Color online) Spectrograms of (a) fin whale calls recorded at 50° N 135° W on 25 November 2001 (spectrogram parameters: FFT 2048 points, 75% overlap, Hanning window, for a filter bandwidth of 1.9 Hz); (b) humpback whale calls recorded at 53° N 157° W on 8 November 1999 (spectrogram parameters: FFT 512 points, 75% overlap, Hanning window, for a filter bandwidth of 7.8 Hz); (c) eastern North Pacific and (d) western North Pacific blue whale calls recorded at 50° N 135° W. The eastern calls were recorded on 18 September 2000 and the western on 17 October 2001. (e) A series of calls from “Watkin’s whale” recorded at 54° N 145° W on 20 October 1999 (spectrogram parameters: FFT 4096 points, 75% overlap, Hanning window, for a filter bandwidth of 0.97 Hz).

C. Blue whales

Calls produced by blue whales in the Northeast Pacific and Northwest Pacific are distinctive [Figs. 2(c) and 2(d)]. The overall seasonal pattern for the two call types was the same: sounds were recorded only from late summer (July) until mid- to late-December [Figs. 3(a)–3(g)]. Both eastern and western blue whale calls show the same pattern in the 2001–2002 dataset as they did in the 1999–2001 data (Stafford, 2003). The NEP call type was most common in October and November at 50° N 135° W and from August to September at 51° N 140° W. For western call types, the highest call detections for all hydrophones in all years were in August, with the exception of the instrument at 53° N 157° W, where call occurrence was slightly greater in July than August. No NEP blue whales were recorded on this hydrophone at 51° N 149° W and so few were recorded on the 53° N 157° W hydrophone that the proportion of hours with detections was less than 0.01.

D. Watkins’ whale

Because the sounds attributed to Watkins’ whale [Fig. 2(e)] have not previously been described quantitatively, measurements were made for 30 call series and given as means +1 standard deviation. Measurements were made for an av-

erage of 8 ± 4 sequences per series, where a sequence contains calls at regular intervals and is bounded on each end by a long silent period, thought to be a blow cycle (Cummins and Thompson, 1971). Overall, each sequence consisted of 11.8 ± 2.7 frequency-modulated calls descending from 50.6 ± 0.34 Hz down to 49.40 ± 0.40 Hz over 6.14 ± 0.91 s ($n=30$). Time from the end of one call to the start of the next was 8.93 ± 1.47 s.

Watkins’ whale calls were detected in the same seasonal pattern as both blue whale call types [Figs. 3(a)–3(g)]. In 1999, they were recorded on all hydrophones except 53° N 157° W from the beginning of recording (end of October or beginning of November) to the end of December. In 2000, they were recorded on all four functioning hydrophones from the end of August until the end of December. They were recorded most often on the hydrophone at 50° N 150° W from August to October and then on the hydrophone at 50° N 145° W from October to the end of December. In 2001, these calls were only recorded in November and December on the three recovered hydrophones. Watkins’ whale calls were often recorded during the same hours on multiple hydrophones and were never recorded on more than three instruments at one time.

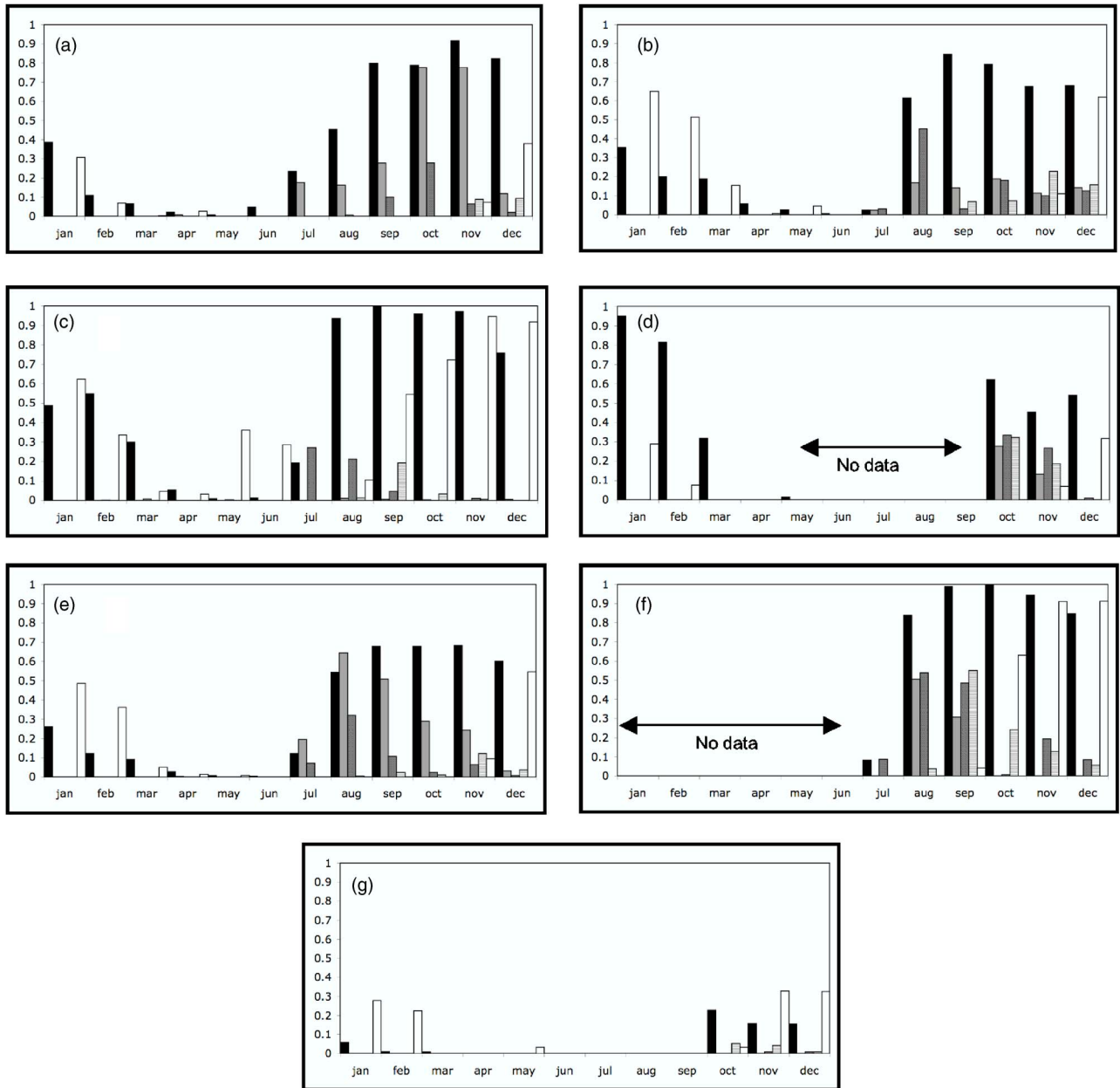


FIG. 3. (Color online) Annual detections of whale calls in the Gulf of Alaska averaged by month from October 1999 to May 2002. Histograms show the proportion of hours per month in which calls were detected and arrows indicate times when data were not available. Fin whales are shown as black bars, humpbacks as white bars, eastern North Pacific blue whales as dark gray, western North Pacific blue whales as medium gray and Watkins' whale as light gray. (a) 50°N 135°W; (b) 50°N 145°W; (c) 53°N 157°W; (d) 55°N 145°W; (e) 52°N 140°W; (f) 51°N 152°W; and (g) 50°N 149°W.

E. Detection distance estimation

For all modeled frequencies, the choice of ambient noise levels had the biggest influence on the detection probabilities when all else (source level, detection threshold) was held constant. For all but the lowest ambient noise levels (from Wenz, 1962) the 50% detection probability distance was less than 50 km for all frequencies modeled with the exception of Watkins' whale (50 Hz).

The modeled detection distance for fin whale calls (25 Hz) was less than 10 km for the two higher ambient noise levels although at the lowest ambient noise level calls over 100 km away should be detectable at least 50% of the time, and in a convergence zone seen at the second-lowest level, calls to 50 km should be detectable [Fig. 4(a)]. Detec-

tion radii of 100 km for all locations in the GoA, then, show that the same fin whale calls should likely not be recorded on more than one instrument at a time [Fig. 5(a)].

Less surprisingly, humpback whale calls (350 Hz) would not be detected if the animals were within 2 km of the instrument due to shadowing, but would be detected with 100% probability between 5 and 12 km. There is evidence of a convergence zone at the lowest ambient noise levels indicating that some humpback whale calls might be recorded up to 45 km away, but not between 15 and 40 km [Fig. 4(b)]. Calls from an animal recorded on one hydrophone should not be heard on any other hydrophones in the array [Fig. 5(b)].

Eastern blue whale detection probabilities were very similar to those of fin whales at all ambient noise levels

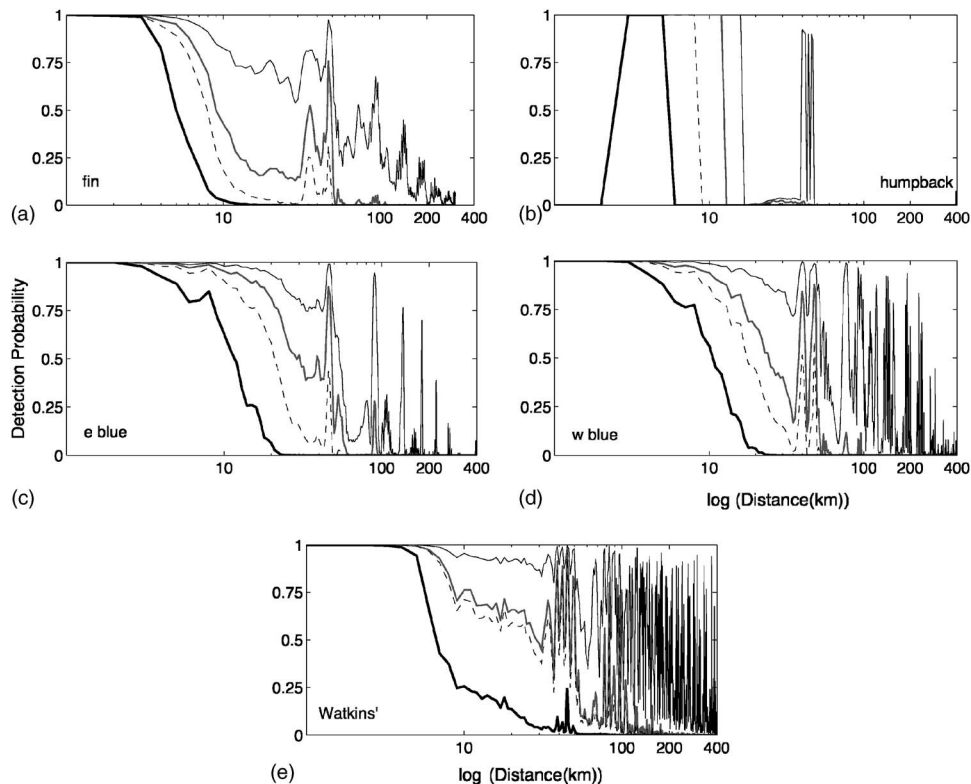


FIG. 4. Detection probabilities for (a) 25 Hz fin whale; (b) 350 Hz humpback whale; (c) 16 Hz eastern North Pacific blue whale; (d) 20 Hz western North Pacific blue whale; (e) 50 Hz Watkins' whale. Heavy black line is detection probability for highest ambient noise (AN) levels from Andrew *et al.* (2002), thin black line is minimum AN levels (Wenz, 1962), thick gray line is detection probability from Urick (1983) moderate AN levels, and the dashed line is from Urick (1983) high AN levels and thick black line is from Andrew *et al.* (2002).

besides the lowest; for the higher ambient noise levels, the 50% detection probability ranged from 10 to 20 km, whereas for the two lower ambient noise levels, the detection probability was over 75% at 45 km away. With the model parameters used here, only at very low ambient noise levels (75 dB) and in caustics with roughly 40 km spacing as far away as 190 km would eastern blue whale calls be detectable [Fig. 4(c)]. Despite this increased detection range, recording the same calls on more than one instrument at a time is highly unlikely [Fig. 5(c)].

Surprisingly, western Pacific blue whales, modeled at 20 Hz, could be detected with 50% probability roughly 60 km further than the 16 Hz eastern Pacific blue whales at the lowest ambient noise level. At the other three ambient noise levels, detection distances were similar to eastern Pacific blue whales [Fig. 4(d)]. This additional 60 km should sometimes permit the same call to be recorded on two of the more closely spaced hydrophones [Fig. 5(d)].

Watkins' whale (50 Hz) should be reliably detected at almost twice the distance of eastern and western Pacific blue whales, with the 50% detection probability distance as great as 50 km for the two middle noise levels and perhaps further than the 400 km modeled at the very lowest level. This should allow the same sound to be detected on up to three instruments at the same time [Figs. 4(e) and 5(e)]. This phenomenon of very low 50 Hz attenuation has been noted in both the subarctic North Pacific and sub-Antarctic South Pacific (Kibblewhite *et al.*, 1977; Chow and Turner, 1982).

When we plot the 50% detection distances as radii

around our locations for both conservative and absolute maximum distances (Fig. 5), it is clear that at medium and high ambient noise levels, no individual could be recorded on more than one instrument at once. But when the minimum noise levels and convergence zone distances are used, Watkins' whale could be detected on up to three instruments at a time, whereas blue whales might be detected on at most two instruments simultaneously and fin and humpback whales only on a single instrument.

IV. DISCUSSION

A. Fin whales

Fin whale calls showed the same seasonality throughout the GoA in all three years, with uniformly high occurrence from late summer/fall through early winter. Watkins *et al.* (2000a, b) reported a similar pattern where peak calling was in midwinter and fin whale calls were seldom detected in summer. Most of their detections were from the northeast Pacific. Watkins *et al.* interpreted the patterned series as a winter reproductive display. Consistent with this, Croll *et al.* (2002) reported that all fin whales in the Sea of Cortez making long, patterned series of pulses were male. Fin whales are regularly seen in coastal waters of the GoA during summer months and were taken in almost all months of the year by whalers in deep water (Zerbini *et al.*, 2006; S. A. Mizroch *et al.*, private communication, 2007). We must assume that they are present, but perhaps not vocalizing in summer. It is also possible that fin whales move closer to shore during summer,

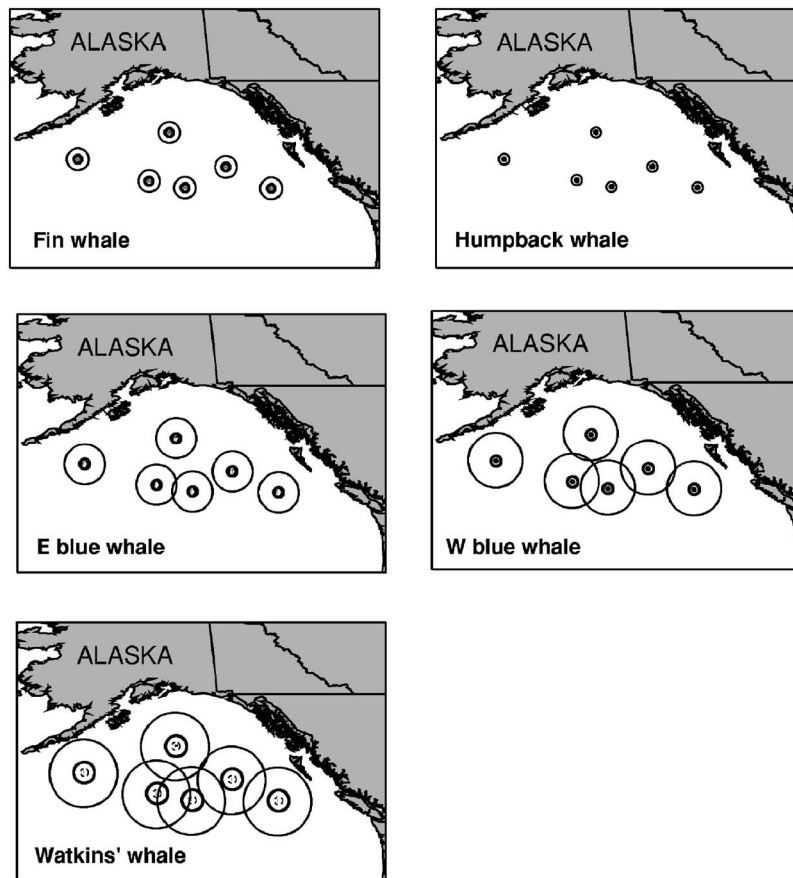


FIG. 5. Modeled detection ranges based on maximum distance for 50% detection probability for the different ambient noise levels and modeling parameters used in the Gulf of Alaska by species at the six deployment sites from 1999. Lines shown for each detection distance are the same as those in Fig. 4 although the distances for higher ambient noise levels were sometimes so short that they do not show up as distinct from the hydrophone location symbol. (a) Fin whales; (b) humpback whales; (c) eastern North Pacific blue whales; (d) western North Pacific blue whales; and (e) Watkins' whale.

out of acoustic range of our deep-water hydrophones, particularly given the apparent constraints on detection distances. This limited detection range has recently been shown for fin whales off the Antarctic Peninsula, where the maximum detection range determined from multipath arrivals was found to be only 56 km (Širović *et al.*, 2007).

What seems most likely, based on a combination of both acoustic (this study) and sighting (Zerbini *et al.*, 2006; S. A. Mizroch *et al.*, private communication.) data, is that fin whales are present in the GoA year-round but they do not produce long series of pulses during the summer. Watkins (1981) found that fin whales in the Atlantic Ocean produced long series of 20 Hz pulses from October–May, whereas single pulses and very short series (2–5) of pulses were recorded most often in the summer. In the North Pacific, fin whale calls were detected seasonally off Oahu on recordings from 1978 to 1981. Long patterned series were most common in winter months, particularly December to March, and were not recorded in June, July, or November (Thompson and Friedl 1982).

There is no current estimate of the number of fin whales in the North Pacific (Angliss *et al.*, 2003) although Zerbini *et al.* (2006) provide an estimate of 1142–2389 animals along the Aleutian Islands and propose that North Pacific fin whale populations are increasing. Fin whales were sighted

during all months of the year during aerial surveys around Kodiak Island (Wynne and Witteveen, 2005).

B. Humpback whales

We found humpback sounds to be quite rare in mid-summer but surprisingly common during the fall and winter when, presumably, mature animals would be on the wintering grounds. As with fin whales, humpback whales are sighted regularly in the coastal waters of the GoA in all months of the year (Baker *et al.*, 1985; Gabriele and Frankel, 2003; Witteveen *et al.*, 2004; Zerbini *et al.*, 2006). Our consistent recording of song from autumn through spring, with highest occurrence in winter months, indicates that even whales that are not on the traditional wintering/breeding grounds sing. This observation has been made for humpback whales in Australia where most acoustic monitoring of singing has occurred on the migration route rather than on breeding grounds (Cato, 1991), and in the western Atlantic (Clark and Clapham, 2004). The recording of song in the GoA is not new. Baker *et al.*, (1985) found that a small number of humpback whales remained in southeast (SE) Alaska into the winter months. They recorded at least one instance of song in late December in SE Alaska. Further, song is regularly heard in SE Alaska throughout the fall and winter (Gabriele and Frankel, 2003; J. Straley, private communication, 2004).

There are several hypotheses that explain our recordings of song throughout much of the year. The migration of humpbacks between wintering and summering grounds does not occur all at once; rather, animals may be continuously migrating between these areas. Numerous authors have suggested that humpback whales may migrate far offshore between wintering and summering areas (Calambokidis *et al.*, 1996; Forney and Barlow, 1998; Mate *et al.*, 1998; Charif *et al.*, 2001). Song has been recorded on migratory routes in the North Atlantic (Clapham and Mattila, 1990), North Pacific (Norris *et al.*, 1999); off both coasts of Australia (Cato, 1991; Cato *et al.*, 2001) and off Bermuda (Payne and McVay, 1971). The first two of these studies proposed that singing might help whales maintain contact during migration (Clapham and Mattila, 1990; Norris *et al.*, 1999). The third study refutes this idea noting that singing males migrate later than other age/sex classes and therefore migratory singing should not serve to maintain contact among animals (Cato, 1991).

Humpback whales moving northward may sing until they reach the feeding grounds. It is possible that they continue to sing throughout the summer but, because the animals are relatively close to shore, our instruments do not detect the songs. This seems unlikely, as numerous studies in the summer have recorded social and feeding vocalizations or no vocalizations at all, but never song, during this period (Thompson *et al.*, 1986; Cerchio and Dahlheim, 2001; Sharpe, 2001; Gabriele and Frankel, 2003). Both Payne and Guineé (1983) and McSweeney *et al.* (1989) suggested that humpback whales from different wintering areas might use singing on the feeding grounds to maintain song structure between wintering seasons and perhaps converge on a common song for the following season. Last, Norris *et al.* (1999) proposed that singing during migration might be a continuation of courtship activity. Off western Australia, the overall proportion of migrating whales that were singers off western Australia was low but was much greater on the southward (towards breeding ground) migration (0.12) than for the northward migration (0.05; Cato *et al.*, 2001). Further, as seasonal hormonal levels have been implicated in singing behavior, less song on the northward migration may indicate decreases in levels of hormones that may influence singing. Clark and Clapham (2004) have since suggested that spring singing on feeding grounds by humpback whales in the northeast Atlantic might be a display by males hoping to take advantage of extra-seasonal mating opportunities.

C. Blue whales

The GoA appears to represent the northern extent of the range of eastern North Pacific blue whales. Like the western North Pacific calls, eastern North Pacific calls are detected only from late July to mid-December (Stafford, 2003). These sounds are also recorded only in the latter half of the year in the eastern North Pacific (Stafford *et al.*, 2001). It may be that only a small proportion of the population migrates as far north as the GoA as calls are only recorded for about five months in the GoA. This is opposed to the eastern tropical

Pacific southern limits of the population, where calls are recorded year-round but most often from January to June (Stafford *et al.*, 1999a).

Western North Pacific blue whales are recorded in the Gulf from mid-July until mid-December. Elsewhere in the western North Pacific, they are recorded from June to February (Stafford *et al.*, 2001). The closest instruments on which these calls have been recorded previously have seasonal call occurrence that is similar to the GoA hydrophones: July–early January (Watkins *et al.*, 2000a, b; Stafford *et al.*, 2001). This suggests that western North Pacific blue whales may be moving eastward into the Gulf in late July and move back westerly in November and early December. Although Watkins *et al.* (2000a, b) did not distinguish between eastern and western call types, they found that vocalizing blue whales from the western North Pacific moved from the southeast westward and back again. A similar pattern for a western Pacific migratory route was proposed by Nemoto (1955). It may be that these animals undertake a more east-west migration as opposed to a north–south migration as is seen in eastern Pacific blue whales (Stafford *et al.*, 1999a).

D. Watkins' whale

The source of the Watkins' whale call is currently unknown, but the similarity in seasonal occurrence of this call type to blue whale calls suggests they may be somehow related. It seems to share characteristics with the third harmonic of the eastern North Pacific blue whale AB call (Cummins and Thompson 1971) in that it is a frequency-modulated sweep at around 51 Hz. However, the duration (approximately 6 s per call) and repetition rate (9 s between calls) and bandwidth (1.5 vs 4 Hz) are less than for the eastern North Pacific blue whale calls.

Watkins *et al.* (2004) used U.S. Navy arrays to track this call in the northeast Pacific for 12 years (1992–2004). They contend that a single individual produced the calls as it transited the northeast Pacific and GoA. Our data, although not as comprehensive as those presented by Watkins *et al.* (2004), support this notion to a certain extent. Our detections of these sounds correspond well with theirs for the period October 1999–June 2002. That is, Watkins' whale calls were detected on the GoA array immediately after deployment in 1999 on the northernmost instrument (54° N 145° W) and throughout the GoA until 24 December. The calls were recorded latest on the southeastern-most instrument (50° N 135° W), a finding that corresponds to the description of the 1999–2000 track by Watkins *et al.* (2004). In 2000, our first detections of Watkins' whale calls were on 25 August [versus 24 August for Watkins *et al.* (2004)] and our last on 19 December, after which time the animal apparently moved too far south of the GoA instruments to be detected. Finally, in 2001, only two GoA instruments were functioning when Watkins' whale calls were first located on Sound Surveillance System (SOSUS) (22 August), and these calls were detected on our instrument only from 31 October to 13 December. The SOSUS track turned south on 7 December and the animal was apparently out of range of our instruments by 13 December. When Watkins' whale calls were detected on

multiple GoA instruments, they were detected during the same day and were found on a maximum of three instruments at any one time. Under low ambient noise conditions, this call type should theoretically be detectable on at least two and as many as three instruments at once (Fig. 5). Arguments against it being a lone animal include the requirement for low ambient noise levels and the fact that in a few instances, this call type was recorded on the instrument furthest west in the hour following a detection at an instrument over 700 km away the hour before.

Watkins *et al.* (2004) conclude that the reason they were able to track this unique source was due mainly to the relatively low ambient noise levels near 50 Hz. Although this certainly played a large role in the aiding the detection, the results of our transmission loss modeling suggest that the low attenuation of this whale's calls may have been equally important in the detection of this sound source over long distances.

It is unlikely that the source of these sounds will ever be known, although it seems very likely, as argued by Watkins *et al.* (2004), that a large whale produces them. Based on the unique sound characteristics and tracks of this animal, it may be a hybrid. The timing of detections of this call type in the GoA is the same as that of both eastern and western North Pacific blue whales. Combined with the calls' resemblance to the third harmonic of eastern North Pacific blue whales, Watkins' whale may be at least partly blue whale and is potentially a blue-fin hybrid. Although to our knowledge, sounds from such animals have never been recorded, there are good genetic records of their existence in the North Pacific, including the Gulf of Alaska (reviewed by Bérubé and Aguilar, 1998).

E. Ecological interpretation

The presence of baleen whales on known whaling grounds in the North Pacific has been directly correlated with the presence of their prey (Nemoto, 1955, 1970; Nemoto and Kasuya, 1965; Berzin and Rovnin, 1966). Blue whale arrivals on the northern whaling grounds were directly correlated with the increase of euphausiids in the northern Pacific from July to August, and both were found in abundance through September (Nemoto, 1955). Likewise, the occurrence of fin whales was positively correlated with the presence of *Thysanoessa longipes* in the central GoA (Nemoto and Kasuya, 1965). Not only was there a positive relationship between whale abundance and prey abundance, but there was a converse negative relationship as well: the absence of whales in a known whaling area reflected the absence of a preferred prey type in that area (Nemoto, 1955; Nemoto and Kasuya, 1965).

The winter occurrence of fin and humpback whales in the GoA is probably due to food availability, with vocalizing animals perhaps trying to take advantage of mating opportunities on feeding grounds (Clark and Clapham, 2004). Unlike fin and humpback whales, blue whale vocalizations are recorded only during an abbreviated part of the year in the GoA, and animals may not overwinter here.

Although little is known of broad-scale abundance of zooplankton in the deep water GoA, densities of adult euphausiids *Euphausia pacifica* and *T. longipes* are greatest in the winter while *T. inermis* are most numerous in the early spring (Cooney, 1987). Unlike the relatively stenophagous blue whale, both humpback and fin whales are known to feed on many different species, often switching prey with season and latitude (Nemoto, 1955, 1970; Kawamura 1980, 1982; Weinrich *et al.*, 1997).

With very few exceptions, humpback whales do not feed on their low-latitude wintering grounds (reviewed in Baraff *et al.*, 1991). Whether the same is true for fin whales is unknown, although they have been seen feeding in the Gulf of California (Tershy *et al.*, 1990). Compared to humpback or blue whales, relatively little is known of fin whale annual migratory patterns or destinations. Nevertheless, whalers noted that both fin and humpback whales caught on migration toward feeding grounds had much thinner blubber layers (Ingebrigtsen, 1929; Dawbin, 1966; Kawamura, 1975). Therefore it may be advantageous for nonbreeding animals to remain in more productive northern waters throughout the winter. Eastern North Pacific blue whales, on the other hand, winter in areas of known year-round productivity, so leaving a known feeding ground for the possibility of reproductive opportunities does not come at the cost of fasting. It is unknown where western North Pacific blue whales or Watkins' whale winter, although acoustic tracks of Watkins' whale indicate a southward movement in the winter in the eastern North Pacific (Watkins *et al.*, 2004).

A comparison of acoustic data with sighting and whaling data show a mismatch for species such as fin and humpback whales in that their sounds were either not recorded, or recorded at very low levels during the seasons (primarily late spring to early fall) in which they are reported to occur in the GoA. In these seasons, then, acoustic data may not be relied upon to determine the presence/absence of a species unless the absence of calls in deep water indicates animals having moved up on the shelf to feed, or whales that are quiet while feeding. However, the recording of these animals' sounds in winter, when there is a paucity of sighting data, provides a valuable indication of their presence. Blue whales, on the other hand, produce the call types we studied here year-round, and the presence or absence of their calls here appears to be useful for determining the presence or absence of animals.

F. Detection distances

Detection distances for frequencies characteristic of the diagnostic calls of four types of whales, were modeled in order to estimate how far different species might be detected by the instruments on which they were recorded. For all cases, even the very low-frequency blue and fin whales, these distances were much smaller than expected. These results are only applicable to the parameters of this experiment, in particular the depths of both the animals, the receiving instruments, and the transmission path between them. Further, although the transmission loss paths showed few seasonal differences, we did not account for the presumed

increase in wintertime ambient noise due to storms. Adjusting for increased noise would further decrease our detection distances in the wintertime when many of the acoustic detections occurred.

Depth dependence of frequency-specific noise levels has been studied in the North Pacific basin (Wagstaff, 2005); these levels include noise contributions from long distance shipping propagating in the deep sound channel, where our instruments were moored. For all species and locations, the detection distances were largely determined by the choice of ambient noise levels. This overlap suggests that masking due to anthropogenic noise might be limiting the range over which these animals can communicate.

Ambient noise has been reported to be increasing by 3 dB per decade off central and southern California since the mid-1960s presumed to be largely due to increases in shipping (Andrew *et al.*, 2002; McDonald *et al.*, 2006). We used noise levels from the temperate northeast Pacific as these were the best available; however, they may not be accurate for the subarctic North Pacific. However, although distant from the biggest west coast ports (Los Angeles–Long Beach, CA), the GoA is on the great circle route between North America and Asia and in the mid-1970s ship densities in the northeast Pacific were highest in the GoA and west of 160°W (Ross, 2005). Current shipping densities from NOAA's Voluntary Observing Ship Program show moderate to high densities of ships in the subarctic North Pacific and GoA (Roberts, 2006) suggesting that the increases documented for the temperate eastern Pacific might also apply to the GoA.

Finally, based on transmission loss modeling, it appears that for our experimental configuration, constraints can be placed on how far different whale species might be heard. Most species are detected only when they are less than 50 km from our instruments. Better measurements of ambient noise, both anthropogenic and natural, are required to better determine detection ranges. Additionally, acoustic behaviors, including source levels and vocalizing depths, of the species in question need to be better understood. We did not vary the source level of whale calls in this experiment, although changing these values, which must certainly vary in nature, can result in large differences in detection distance (Cato, 1991). With better-known parameters, in particular ambient noise levels and detection thresholds, detection radii might eventually be used to estimate minimum abundances of vocal animals over known areas.

ACKNOWLEDGMENTS

This work was supported by ONR Grant No. N00014-03-1-0099 and N00014-03-1-0735 and JISAO Grant No. NA17RJ1232 and NPS Contract No. N00024-02-D-6602. Critiques from two anonymous reviewers greatly improved an earlier draft of this manuscript. Haru Matsumoto designed the moored hydrophones and Matt Fowler recovered and deployed them without which this work would not have been possible. This is PMEL Contribution No. 2900.

Andrew, R. K., Howe, B. M., Mercer, J. A., and Dzieciuch, M. A. (2002). "Ocean ambient sound: comparing the 1960s with the 1990s for a receiver off the California coast," *ARLO* 3, 65–70.

- Angliss, R. P., and Lodge, K. L. (2003). "Alaska marine mammal stock assessments, 2003," U.S. Dept. of Commerce, NOAA Technical Memorandum No. NMFS-AFSC-144, available online from http://www.nmfs.noaa.gov/prot_res/PR2/Stock_Assessment_Program/sars.html (accessed 20 August 2007).
- Au, W. W. L., Pack, A. A., Lammers, M. O., Herman, L. M., Deakos, M. H., and Andrews, K. (2006). "Acoustic properties of humpback whale songs," *J. Acoust. Soc. Am.* 120, 1103–1110.
- Baker, C. S., Herman, L. M., Perry, A., Lawton, W. S., Straley, J. M., and Straley, J. H. (1985). "Population characteristics and migration of summer and late-season humpback whales (*Megaptera novaeangliae*) in southeastern Alaska," *Marine Mammal Sci.* 1, 304–323.
- Baraff, L. S., Clapham, P. J., and Mattila, D. K. (1991). "Feeding behavior of a humpback whale in low-latitude waters," *Marine Mammal Sci.* 7, 197–202.
- Barlow, J. (1994). "The abundance of cetaceans in California waters. Part I: Ship surveys in summer and fall of 1991," *Fish. Bull.* 93, 1–14.
- Bérubé, M., and Aguilar, A. (1998). "A new hybrid between a blue whale, *Balaenoptera musculus*, and a fin whale, *B. physalus*: Frequency and implications of hybridization," *Marine Mammal Sci.* 14, 82–98.
- Berzin, A. A., and Rovnin, A. A. (1966). "Распределение и миграции китов в северо-восточной части Тихого Океана, в Беринговом и Чукотском морях. [Distribution and migrations of whales in the northeastern part of the Pacific, Chukchi and Bering Seas]," edited by K. I. Panin, *Soviet Research on Marine Mammals in the Far East. Izv. Tikhookean. Nauchno-Issled. Inst. Rybn. Khoz. Okeanogr.* 58 (in Russian, translated by the U.S. Joint Publications Research Service for the Bureau of Commercial Fisheries), available from the National Marine Mammal Laboratory, NMFS, NOAA, Seattle, Washington pp. 179–207.
- Brickley, P. J., and Thomas, A. C. (2004). "Satellite-measured seasonal and interannual chlorophyll variability in the Northeast Pacific and Coastal Gulf of Alaska," *Deep-Sea Res., Part II* 51, 229–245.
- Brownell, R. L., Clapham, P. J., Miyashita, T., Kasuya, T. (2001). "Conservation status of North Pacific right whales," *J. Cetacean Res. Manage.* 2, 269–286.
- Calambokidis, J., Steiger, G. H., Evenson, J. R., Flynn, K. R., Balcomb, K. C., Claridge, D. E., Bloedel, P., Straley, J. M., Baker, C. S., von Ziegeler, O., Dahlheim, M. E., Waite, J. M., Darling, J. D., Ellis, G., and Green, G. A. (1996). "Interchange and isolation of humpback whales off California and other North Pacific feeding grounds," *Marine Mammal Sci.* 12, 215–226.
- Cato, D. (1991). "Songs of humpback whales: The Australian perspective," *Memoirs of the Queensland Museum* 30, 277–290.
- Cato, D. H., Paterson, R. A., and Paterson, P. (2001). "Vocalisation rates of migrating humpback whales over 14 years," *Memoirs of the Queensland Museum* 47, 481–489.
- Cerchio, S., and Dahlheim, M. (2001). "Variation in feeding vocalizations of humpback whales *Megaptera novaeangliae* from southeast Alaska," *Bioacoustics* 11, 277–295.
- Charif, R. A., Clapham, P. J., and Clark, C. W. (2001). "Acoustic detections of singing humpback whales in deep waters off the British Isles," *Marine Mammal Sci.* 17, 751–768.
- Charif, R. A., Mellinger, D. K., Dunsmore, K. J., Fristrup, K. M., and Clark, C. W. (2002). "Estimated source levels of fin whale (*Balaenoptera physalus*) vocalizations: Adjustments for surface interference," *Marine Mammal Sci.* 18, 81–98.
- Chow, R. K., and Browning, D. G. (1983). "Low-frequency attenuation in the Northeast Pacific Subarctic transition zone," *J. Acoust. Soc. Am.* 74, 1635–1638.
- Chow, R. K., and Turner, R. G. (1982). "Attenuation of low-frequency sound in the Northeast Pacific Ocean," *J. Acoust. Soc. Am.* 72, 888–891.
- Clapham, P. J., and Mattila, D. K. (1990). "Humpback whale songs as indicators of migration routes," *Marine Mammal Sci.* 6, 155–160.
- Clark, C. W., and Charif, R. A. (1998). "Acoustic monitoring of large whales to the west of Britain and Ireland using bottom-mounted hydrophone arrays, October 1996–September 1997," Report No. 281, Joint Nature Conservation Committee, Aberdeen, Scotland.
- Clark, C. W., and Clapham, P. J. (2004). "Acoustic monitoring on a humpback whale (*Megaptera novaeangliae*) feeding ground shows continual singing into late spring," *Proc. R. Soc. London, Ser. B* 271, 1051–1057.
- Clark, C. W., and Ellison, W. T. (2004). "Potential use of low-frequency sounds by baleen whales for probing the environment: Evidence from models and empirical measurements," in *Advances in the Study of Echolocation in Bats and Dolphins*, edited by J. A. Thomas and R. A. Kastelein

- (Plenum, New York), pp. 564–589.
- Collins, M. D. (1993). "A split-step Padé solution for the parabolic equation method," *J. Acoust. Soc. Am.* **93**, 1736–1742.
- Collins, M. D. (1995). "User's guide for RAM versions 1.0 and 1.0p," Technical report, Naval Research Laboratory, Washington, D.C.
- Cooney, R. T. (1987). "Zooplankton," in *The Gulf of Alaska: Physical Environment and Biological Resources*, edited by D. W. Hood and S. T. Zimmerman (Minerals Management Service, Anchorage, Alaska), Technical Report No. MMS 86-0095.
- Croll, D. A., Clark, C. W., Acevedo, A., Tershy, B., Flores, S., Gedamke, J., and Urban, J. (2002). "Only male fin whales sing loud songs," *Nature (London)* **417**, 809.
- Cummings, W. C., and Thompson, P. O. (1971). "Underwater sounds from the blue whale, *Balaenoptera musculus*," *J. Acoust. Soc. Am.* **50**, 1193–1198.
- Dawbin, W. H. (1966). "The seasonal migratory cycle of humpback whales," in *Whales, Dolphins and Porpoises*, edited by K. S. Norris (University of California Press, Berkeley), pp. 145–170.
- Forney, K. A., and Barlow, J. (1998). "Seasonal patterns in the abundance and distribution of California cetaceans, 1991-1992," *Marine Mammal Sci.* **14**, 460–489.
- Fox, C. G., Matsumoto, H., and Lau, T. K. A. (2001). "Monitoring Pacific Ocean seismicity from an autonomous hydrophone array," *J. Geophys. Res.* **106**, 4183–4206.
- Gabriele, C., and Frankel, A. (2003). "The occurrence and significance of humpback whale songs in Glacier Bay, southeast Alaska," *Arctic Research of the United States* **16**, 42–47.
- Ingebrigtsen, A. (1929). "Whales caught in the North Atlantic and other seas," *Conseil Permanent International pour l'Exploration de la Mer. Rapports et Proces Verbaux des Reunions* **56**, 123–135.
- Kawamura, A. (1975). "A consideration on an available source of energy and its cost for locomotion in fin whales with special reference to the seasonal migrations," *Sci. Rep. Whales Res. Inst.* **27**, 61–79.
- Kawamura, A. (1980). "A review of food of Balaenopterid whales," *Sci. Rep. Whales Res. Inst.* **32**, 155–197.
- Kawamura, A. (1982). "Food habits and prey distributions of three rorqual species in the North Pacific Ocean," *Sci. Rep. Whales Res. Inst.* **34**, 59–91.
- Kibblewhite, A. C., Bedfordand, N. R., and Mitchell, S. K. (1977). "Regional dependence of low-frequency attenuation in the North Pacific Ocean," *J. Acoust. Soc. Am.* **61**, 1169–1177.
- Mate, B. R., Gisiner, R., and Mobley, J. (1998). "Local and migratory movements of Hawaiian humpback whales tracked by satellite telemetry," *Can. J. Zool.* **76**, 863–868.
- McDonald, M. A., and Moore, S. E. (2002). "Calls recorded from North Pacific right whales (*Eubalaena japonica*) in the eastern Bering Sea," *J. Cetacean Res. Manage.* **4**, 261–266.
- McDonald, M. A., Hildebrand, J. A., and Wiggins, S. (2006). "Increases in deep ocean ambient noise in the Northeast Pacific west of San Nicolas Island, California," *J. Acoust. Soc. Am.* **120**, 711–718.
- McSweeney, D. J., Chu, K. C., Dolphin, W. F., and Guinee, L. N. (1989). "North Pacific humpback whale songs: A comparison of southeast Alaskan feeding ground songs with Hawaiian wintering ground songs," *Marine Mammal Sci.* **5**, 139–148.
- Mellinger, D. K., Stafford, K. M., and Fox, C. G. (2004a). "Seasonal occurrence of sperm whale (*Physeter macrocephalus*) sounds in the Gulf of Alaska," *Marine Mammal Sci.* **20**, 48–62.
- Mellinger, D. K., Stafford, K. M., Moore, S. E., Munger, L., and Fox, C. G. (2004b). "Detection of North Pacific right whale (*Eubalaena japonica*) calls in the Gulf of Alaska," *Marine Mammal Sci.* **20**, 872–879.
- 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.
- Moore, S. E., Stafford, K. M., Dahlheim, M. E., Fox, C. G., Braham, H. W., Polovina, J. J., and Bain, D. E. (1998). "Seasonal variation in reception of fin whale calls at five geographic areas in the North Pacific," *Marine Mammal Sci.* **14**, 617–627.
- Moore, S. E., Stafford, K. M., Mellinger, D. K., and Hildebrand, J. A. (2006). "Listening for Large Whales in the Offshore Waters of Alaska," *BioScience* **56**, 49–55.
- Moore, S. E., Watkins, W. A., Daher, M. A., Davis, J. R. (2002). "Blue whale habitat associations in the northwest Pacific: analysis of remotely-sensed data using a geographic information system," *Oceanogr.* **15**, 20–25.
- National Marine Fisheries Service (NMFS). (2005). "Species listed under the Endangered Species Act of 1973," web site http://www.nmfs.noaa.gov/prot_res/species/ESA_species.html (accessed 20 August 2007).
- Nemoto, T. (1955). "Foods of baleen whales in the Northern Pacific," *Sci. Rep. Whales Res. Inst.* **10**, 33–89.
- Nemoto, T. (1970). "Feeding pattern of baleen whales in the ocean," in *Marine Food Chains*, edited by J. H. Steele (University of California Press, Berkeley, CA) pp. 241–252.
- Nemoto, T., and Kasuya, T. (1965). "Foods of baleen whales in the Gulf of Alaska and the North Pacific," *Sci. Rep. Whales Res. Inst.* **19**, 45–51.
- Nieukirk, S. L., Stafford, K. M., Mellinger, D. K., and Fox, C. G. (2004). "Low-frequency whale sounds recorded from the mid-Atlantic Ocean," *J. Acoust. Soc. Am.* **115**(4), 1832–1843.
- Nishiwaki, M. (1966). "Distribution and migration of the larger cetaceans in the North Pacific as shown by Japanese whaling results," in *Whales, Dolphins and Porpoises*, edited by K. S. Norris (University of California Press, Berkeley, CA), pp. 171–191.
- Norris, T. F., McDonald, M. A., and Barlow, J. (1999). "Acoustic detections of singing humpback whales (*Megaptera novaeangliae*) in the eastern North Pacific during their northbound migration," *J. Acoust. Soc. Am.* **106**, 506–514.
- Omura, H. (1955). "Whales in the northern part of the North Pacific," *Norsk Hvalfangst-Tidende* **44**, 323–345.
- Payne, R. S., and Guinee, L. N. (1983). "Humpback whale (*Megaptera novaeangliae*) songs as an indicator of stocks," *Communication and Behavior of Whales*, edited by R. S. Payne (AAA Selected Symposium 76, Westview Boulder, CO), pp. 333–358.
- Payne, R. S., and McVay, S. (1971). "Songs of humpback whales," *Science* **173**, 585–597.
- Rice, D. W., and Wolman, A. A. (1982). "Whale census in the Gulf of Alaska June to August 1980," *Rep. Int. Whal. Comm.* **32**, 491–497.
- Richardson, W. J., Greene, C. R., Jr., Malme, C. I., and Thomson, D. H. (1995). *Marine Mammals and Noise* (Academic, San Diego).
- Rivers, J. (1997). "Blue whale, *Balaenoptera musculus*, vocalizations from the waters off central California," *Marine Mammal Sci.* **13**, 186–195.
- Roberts, J. P. (2006). "Marine environment protection and biodiversity conservation: The application and future development of the IMO's particularly sensitive sea area concept," Ph.D. dissertation, submitted to the University of Wollongong, Australia.
- Ross, D. (2005). "Ship sources of ambient noise," *IEEE J. Ocean. Eng.* **30**, 257–261.
- Sharpe, F. (2001). "Social foraging of the southeast Alaskan humpback whale, *Megaptera novaeangliae*," Ph.D. dissertation, Simon Fraser University, Burnaby, British Columbia, Canada.
- Shelden, K. E. W., Moore, S. E., Waite, J. M., Wade, P. R., and Rugh, D. J. (2005). "Historic and current habitat use by North Pacific right whales *Eubalaena japonica* in the Bering Sea and Gulf of Alaska," *Mammal Res.* **35**, 129–155.
- Širović, A., Hildebrand, J. A., Wiggins, S. M. (2007). "Blue and fin whale call source levels and propagation range in the Southern Ocean," *J. Acoust. Soc. Am.* **122**, 1208–1215.
- Širović, A., Hildebrand, J. A., Wiggins, S. M., McDonald, M. A., Moore, S. E., and Thiele, D. (2004). "Seasonality of blue and fin whale calls and the influence of sea ice in the Western Antarctic Peninsula," *Deep-Sea Res., Part II* **51**, 2327–2344.
- Springer, A. M., Piatt, J. F., Shuntov, V. P., VanVliet, G. B., Vladimirov, V. L., Kuzin, A. E., and Perlov, A. S. (1999). "Marine birds and mammals of the Pacific subarctic gyres," *Prog. Oceanogr.* **43**, 443–487.
- Stafford, K. M. (2003). "Two types of blue whale calls recorded in the Gulf of Alaska," *Marine Mammal Sci.* **19**, 682–693.
- Stafford, K. M., Bohnenstiehl, D. R., Tolstoy, M., Chapp, E., Mellinger, D. K., and Moore, S. E. (2004). "Antarctic-type blue whale calls recorded at low latitudes in the Indian and the eastern Pacific Oceans," *Deep-Sea Res., Part I* **51**, 1337–1346.
- Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (1999a). "An acoustic link between blue whales in the Northeast Pacific and the eastern tropical Pacific," *Marine Mammal Sci.* **15**, 1258–1268.
- Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (1999b). "Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific," *J. Acoust. Soc. Am.* **106**, 3687–3698.
- Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (2001). "Geographic and seasonal variation of blue whale calls in the North Pacific," *J. Cetacean Res. Manage.* **3**, 65–76.
- Stewart, B. S., Karl, S. A., Yochem, P. K., Leatherwood, S., and Laake, J. L. (1987). "Aerial surveys for cetaceans in the former Akutan, Alaska, whal-

- ing grounds," *Arctic* **40**, 33–42.
- Teague, W. J., Carron, M. J., and Hogan, P. J. (1990). "A comparison between the Generalized Digital Environmental Model and Levitus climatologies," *J. Geophys. Res.* **95**, 7167–7183.
- Tershy, B. R., Breese, D., Strong, C. S. (1990). "Abundance, seasonal distribution and population composition of balaenopterid whales in the canal," de Ballenas, Gulf of California, Mexico. Reports of the Internationals Whaling Commission (special Issue), Vol. **12**, pp. 369–375.
- Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (2000). "Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations," *J. Acoust. Soc. Am.* **107**, 1286–1300.
- Thompson, P. O., and Friedl, W. A. (1982). "A long-term study of low frequency sound from several species of whales off Oahu, Hawaii," *Cetology* **45**, 1–19.
- Thompson, P. O., Cummings, W. C., and Ha, S. J. (1986). "Sounds, source levels and associated behavior of humpback whales, southeast Alaska," *J. Acoust. Soc. Am.* **80**, 735–740.
- Urick, R. J. (1983). *Principles of Underwater Sound*, 3rd Ed. (McGraw-Hill, New York).
- Waite, J. M., Wynne, K., and Mellinger, D. K. (2003). "Documented sighting of a North Pacific right whale in the Gulf of Alaska and post-sighting acoustic monitoring," *Northwestern Naturalist* **84**, 38–43.
- Watkins, W. A. (1981). "Activities and underwater sounds of fin whales," *Sci. Rep. Whales Res. Inst.* **33**, 83–117.
- Watkins, W. A., Daher, M. A., George, J. E., and Rodriguez, D. (2004). "Twelve years of tracking 52-Hz whale calls from a unique source in the North Pacific," *Deep-Sea Res., Part I* **51**, 1889–1901.
- Watkins, W. A., Daher, M. A., Reppucci, G. M., George, J. E., Martin, D. L., DiMarzio, N. A., and Gannon, D. P. (2000a). "Seasonality and distribution of whale calls in the North Pacific," *Oceanogr.* **13**, 62–67.
- Watkins, W. A., George, J. E., Daher, M. A., Mullin, K., Martin, D. L., Haga, S. H., and DiMarzio, N. A. (2000b). "Whale call data for the North Pacific November 1995 through July 1999: Occurrence of calling whales and source locations from SOSUS and other acoustic systems," Technical Report No. WHOI-00-02, Woods Hole Oceanographic Institution, Woods Hole, MA.
- Watkins, W. A., Tyack, P., and Moore, K. E. (1987). "The 20 Hz signals of finback whales (*Balaenoptera physalus*)," *J. Acoust. Soc. Am.* **82**, 1901–1912.
- Weinrich, M. A., Martin, M., Griffiths, R., Bove, J., and Schilling, M. (1997). "A shift in distribution of humpback whales, *Megaptera novaeangliae*, in response to prey in the southern Gulf of Maine," *Fish. Bull.* **95**, 826–836.
- Wenz, G. M. (1962). "Acoustic ambient noise in the ocean: Spectra and sources," *J. Acoust. Soc. Am.* **34**, 1936–1956.
- Whitney, F. A., and Freeland, H. J. (1999). "Variability in upper-ocean water properties in the NE Pacific Ocean," *Deep-Sea Res., Part II* **46**, 2351–2370.
- Witteveen, B. H., Straley, J. M., von Ziegesar, O., Steel, D., and Baker, C. S. (2004). "Abundance and mtDNA differentiation of humpback whales (*Megaptera novaeangliae*) in the Shumagin Islands, Alaska," *Can. J. Zool.* **82**, 1352–1359.
- Wynne, K. M., and Witteveen, B. H. (2005). "Opportunistic aerial sightings of large whales within Steller sea lion critical habitat in the Kodiak archipelago. Gulf Apex Predator-prey Study (GAP)," University of Alaska, Fairbanks, Kodiak, Alaska, Final Report No. FY2001–2003, NOAA Grant No. NA16FX1270., pp. 105–119.
- Zerbini, A. N., Waite, J. M., Laake, J. L., and Wade, P. R. (2006). "Abundance, trends and distribution of baleen whales in western Alaska and the central Aleutian Islands," *Deep-Sea Res., Part I* **53**, 1772–1790.
- Wagstaff, R. A. (2005). "An ambient Noise Model for the Northeast Pacific Ocean Basin," *IEEE J. Ocean. Eng.* **30**, 286–294.

Environmental inversion using high-resolution matched-field processing^{a)}

Cristiano Soares^{b)} and Sérgio M. Jesus

Institute for Systems and Robotics, Universidade do Algarve, Campus de Gambelas, PT-8005-139 Faro, Portugal

Emanuel Coelho^{c)}

NATO Undersea Research Centre, Viale San Bartolomeo 400, I-19138 La Spezia, Italy

(Received 5 March 2007; revised 14 September 2007; accepted 17 September 2007)

This paper considers the inversion of experimental field data collected with light receiving systems designed to meet operational requirements. Such operational requirements include system deployment in free drifting configurations and a limited number of acoustic receivers. A well-known consequence of a reduced spatial coverage is a poor sampling of the vertical structure of the acoustic field, leading to a severe ill-conditioning of the inverse problem and data to model cost function with a massive sidelobe structure having many local extrema. This causes difficulties to meta-heuristic global search methods, such as genetic algorithms, to converge to the true model parameters. In order to cope with this difficulty, broadband high-resolution processors are proposed for their ability to significantly attenuate sidelobes, as a contribution for improving convergence. A comparative study on simulated data shows that high-resolution methods did not outperform the conventional Bartlett processor for pinpointing the true environmental parameter when using exhaustive search. However, when a meta-heuristic technique is applied for exploring a large multidimensional search space, high-resolution methods clearly improved convergence, therefore reducing the inherent uncertainty on the final estimate. These findings are supported by the results obtained on experimental field data obtained during the Maritime Rapid Environmental Assessment 2003 sea trial. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2799476]

PACS number(s): 43.30.Wi, 43.30.Pc, 43.60.Cg, 43.60.Pt [AIT]

Pages: 3391–3404

I. INTRODUCTION

During the 1990s, the problem of simultaneously estimating multiple ocean parameters by means of inversion of acoustic data collected with vertical receiver arrays aroused considerable interest in the underwater acoustic community. Several studies with experimental field data have demonstrated the viability of environmental inversion based on matched-field processing (MFP) with multiple unknown parameters. MFP-based inversion techniques perform a comparison of the full pressure field (amplitude and phase) received at an array of hydrophones with computer generated field replicas, usually by means of a correlation.^{1,2}

MFP, originally proposed for source localization, was first formulated by Bucker³ as he used realistic environmental models, introduced the concept of ambiguity surface, and demonstrated that there was enough complexity of the wave field to allow inversion. The field complexity can be measured in terms of the number of contributing normal modes, which is directly related to the degree of uniqueness of the inverse problem's solution. The number of contributing normal modes varies with physical parameters such as fre-

quency and water depth, among others. The idea of using vertical array is to spatially sample that normal modes structure.

Then it was recognized that MFP could be also applied to environmental inversion problems, such as the inversion of the ocean water column⁴ and bottom properties.^{5–8} As, in general, direct inversion of the acoustic field is not possible, the inverse problem is usually posed as a nonlinear optimization problem aiming at the maximization of the match between the measured acoustic field and the replica field calculated for candidate parameter values. In most cases multiple unknown parameters enter the optimization problem, therefore resulting in large search spaces. Thus, exhaustive search is not a viable practice. Gradient methods are also not viable due to the existence of many local extrema. This requires employment of efficient global optimization methods such as genetic algorithms (GA) and simulated annealing (SA).

Collins *et al.* first proposed including environmental parameters in the search space in the context of range-depth source localization as an attempt to overcome model mismatch.⁹ In this study with synthetic data the optimization was carried out with SA. There have been a number of papers on experimental results on inversion of acoustic data for geometric and environmental parameters^{10–12} aimed at sup-

^{a)} Portions of this work were presented at the European Conference on Underwater Acoustics on June 2006, Carvoeiro, Portugal.

^{b)} Electronic mail: csoares@ualg.pt

^{c)} Current address: Naval Research Laboratory, Code 7322, Bldg. 1009, Room C128, Stennis Space Center, MS 39529.

porting source localization. Other experimental studies used global optimization methods for the estimation of ocean-bottom properties.^{5,6,13}

In most of these studies, part of the success of MFP-based inversion techniques is explained by the fact that most acoustic experiments were carried out under highly controlled conditions, employing acoustic reception systems with a large number of receivers, moored arrays, and low frequency acoustic projectors. In other words, acoustic systems traditionally employed are research directed apparatus, bulky and difficult to operate for their deployment requirements, and are therefore not suitable for operational use.

Current developments of receiver systems go in the sense of reducing their overall size along with the length of the array itself and the number of receivers with the objective of reducing the cost and deployment requirements of these systems. The point is that if a sparse array is used then higher order modes are undersampled, i.e., the spatial Nyquist criterion is not taken into account, and MFP cannot effectively take advantage of that field complexity. The result is that the ambiguity surface, or hypersurface in the case of multiple unknown parameters, will show many sidelobes spread over the search space comparable with the main peak at the true solution, leading to a severely ill-conditioned problem with a large number of local extrema.¹⁴ When dealing with real data, the inherent model mismatch and the presence of noise create a situation where there is no assurance of existence of an optimum solution in coincidence with (or even close to) the true model parameters. An additional concern arises when the optimization problem is solved with aid of meta-heuristic methods such as a genetic algorithm (GA) even in the absence of noise and model mismatch. The large number of local extrema associated with the typically large search space is a major difficulty factor to this class of search methods in attaining convergence to the true model parameters.

This problem leads to an important discussion in MF approaches, which is on the ability of the processor to attenuate sidelobes. In the past, much effort has gone into developing processor techniques with increased sidelobe attenuation capabilities. One of the main topics was the debate on incoherent and coherent processors, where it was claimed that using coherent processors would allow for increased sidelobe attenuation in comparison to the incoherent counterparts.¹⁵⁻²⁰ Another possibility would be the employment of high-resolution processors. However, this possibility has not been significantly considered in the past due to the generalized notion that those methods have weak probabilities of successful application with experimental data due to their high sensitivity to model mismatch. This paper proposes broadband and high-resolution MF processors for environmental inversion of acoustic data collected with the Acoustic Oceanographic Buoy (AOB),^{21,22} a light receiving system with a sparse vertical array deployed in a free-drifting configuration, where the signals were transmitted by a towed acoustic source. Here, the application of high-resolution processors to environmental inversion is motivated by their possibility to significantly improve the convergence of global search algorithms due to their increased ability to attenuate sidelobes. Simulation results show that in the case of an ex-

haustive search, conventional processing is more capable of correctly pinpointing the maximum at the true parameter value than the proposed high-resolution methods. However, both synthetic and experimental inversion results obtained with a GA show that the high-resolution methods can significantly improve convergence to the global solution of the inverse problem.

This paper is organized as follows: Section II develops broadband and high-resolution matched-field processors; Sec. III presents a synthetic study aimed at understanding the difficulties in applying high-resolution processors and comparing them with the conventional processor; Sec. IV gives a description of the MREA'03 sea trial, and presents experimental results obtained with the proposed matched-field processors; finally, Sec. V draws final conclusions.

II. MATCHED-FIELD PROCESSORS FOR PARAMETER ESTIMATION

In order to cope with the difficulty that arises from using a sparse array to collect acoustic data in conjunction with meta-heuristic search methods, the following proposes various broadband (BB) matched-field processors. There are at least two issues that can contribute to alleviate the ill-conditioning of the inversion problem: One is to efficiently use the spectral components of the acoustic field by exploiting field coherence across the spectral band, which has been claimed in the literature as a means of exploiting additional information contained in the acoustic field; the other is the application of matched-field processors based on a BB data model exploiting that cross-frequency coherence. The following matched-field processors are considered herein: A BB Bartlett processor,^{2,1} a BB minimum-variance (MV) processor;^{23,2,1} and a subspace based method, the BB Multiple Signal Classification (MUSIC) processor.²⁴ The MV and the MUSIC processors are high-resolution methods, with an increased ability for attenuating sidelobes in comparison to the Bartlett processor.

A. The broadband data model

The broadband data model for the acoustic data received at an L -receiver array is written as a concatenation of K narrow-band signals $\underline{Y}(\omega_k)$ at discrete frequencies of interest ω_k :

$$\underline{Y} = [\underline{Y}^T(\omega_1), \dots, \underline{Y}^T(\omega_k), \dots, \underline{Y}^T(\omega_K)]^T = \mathbf{H}(\underline{\theta})\tilde{\underline{S}} + \underline{N} \quad (1)$$

in order to introduce, as much as possible, a common framework for the narrow-band and broadband cases (see Ref. 20 for a detailed discussion). This data model allows for accounting for the field coherence across frequencies. The vector $\underline{\theta}$ represents the channel parameters and matrix $\mathbf{H}(\underline{\theta})$ is the channel response matrix given as

$$\mathbf{H}(\underline{\theta}) = \begin{bmatrix} \underline{H}(\omega_1, \underline{\theta}) & \cdots & \underline{0}_{k-1} & \cdots & \underline{0}_{K-2} \\ \underline{0}_1 & \cdots & \underline{H}(\omega_k, \underline{\theta}) & \cdots & \underline{0}_1 \\ \underline{0}_{K-2} & \cdots & \underline{0}_{K-k} & \cdots & \underline{H}(\omega_K, \underline{\theta}) \end{bmatrix}, \quad (2)$$

where the $\underline{H}(\omega_k, \underline{\theta})$ is an L -vector representing the channel response at frequency $\omega_k, k=1, \dots, K$. $\underline{0}_k$ is a vector with kL zeros. This channel matrix is analogous to that used in classical array processing models for multiple emitters. In the present case, each column is relative to a frequency ω_k , however, the channel vectors do not overlap across the columns, in order to keep frequencies separated. The channel matrix has KL rows and K columns. The vector $\tilde{\underline{S}}$ has entries $S(\omega_k)\alpha(\omega_k)$, i.e., the source spectrum multiplied by a random perturbation factor at each frequency $\omega_k \in [\omega_1, \omega_K]$. The random perturbation factor $\alpha(\omega_k)$ appears as an attempt to account for unmodeled ocean inhomogeneities.²⁰ The vector \underline{N} represents the noise, which is assumed Gaussian zero mean, and follows the same notation as \underline{Y} in Eq. (1). Let

$$\mathbf{C}_{YY} = E\{\underline{Y}\underline{Y}^H\} = \mathbf{H}\mathbf{C}_{SS}\mathbf{H}^H + \sigma_N^2\mathbf{I} \quad (3)$$

be a generic definition of the spectral density matrix (SDM) for \underline{Y} defined in Eq. (1), where \mathbf{C}_{SS} is the signal matrix given by $E\{\tilde{\underline{S}}\tilde{\underline{S}}^H\}$, and σ_N^2 the noise variance. The dimensions of the SDM \mathbf{C}_{YY} are $KL \times KL$ consisting of $L \times L$ cross-frequency SDMs $\mathbf{C}_{YY}(\omega_{k_1}, \omega_{k_2})$. The SDMs for $k_1 \neq k_2$ are noiseless according to Eq. (3) since it is assumed that the noise is uncorrelated both across space and frequency. Concerning the signal component, if the signal receptions are fully coherent, then it just happens that $\mathbf{C}_{SS} = \underline{S}\underline{S}^H$, which has rank equal one. On the other hand, if the emitted wave form is a random signal, then $\mathbf{C}_{SS} = \text{diag}[\sigma_S^2(\omega_1), \dots, \sigma_S^2(\omega_k), \dots, \sigma_S^2(\omega_K)]$, with $\sigma_S^2(\omega_k) = E\{\alpha^*(\omega_k)\alpha(\omega_k)S^*(\omega_k)S(\omega_k)\}$. In that case the rank of the signal matrix is equal to K . Note that for this case the SDM \mathbf{C}_{YY} consists only of block matrices in the diagonal. The intermediate case is that where the rank of the signal matrix can vary between 1 and K , representing partial frequency cross correlation. This model is the most generic in the framework of a full broadband data model. At this point we stress the relevance of the rank of the signal matrix: From the signals' point of view, the ocean represents a system with a response that may have features of random nature. In other words, a sequence of deterministic emissions is generally received as a random sequence. The degree of randomness seen at the receivers may depend not only on ocean inhomogeneities such as sea surface roughness, but on small motion of the receivers. There may be a significant contribution related to the variability in the geometry of the experimental setup caused by drifts both of the emitter and the receivers. In terms of the BB data model, it should be noted that the channel response is assumed to be deterministic, but in practice there are both channel random features and parameter variability over the observation window, which is to be accounted for by the introduction of the random perturbation factor. These phenomena may have an impact on the coherence across the spectral band and therefore on the rank of the \mathbf{C}_{SS} signal matrix. Next, the three above-mentioned processors will be derived using the BB data model.

B. The BB Bartlett processor

Conventional or Bartlett matched-field processors are the most popular in underwater acoustic estimation problems, since they have been used in virtually every study on MFP. The frequency domain Bartlett processor, also called linear processor, performs matched-field beamforming by weighting the output of the array elements at different frequencies and summing over all elements:

$$P_B(\underline{\theta}) = E\{\text{tr}[\mathbf{w}^H(\underline{\theta})\underline{Y}(\underline{\theta}_0)\underline{Y}^H(\underline{\theta}_0)\mathbf{w}(\underline{\theta})]\}, \quad (4)$$

where \mathbf{w} is a weighting matrix with K columns. Note that it is assumed that the acoustic field is zero mean without loss of generality. Replacing with Eq. (3) and by performing a few ordinary algebra steps to maximize this criterion with respect to $\mathbf{w}(\underline{\theta})$ under the constraint $\text{tr}[\mathbf{w}^H(\underline{\theta})\mathbf{w}(\underline{\theta})]=1$ the following function is obtained:

$$P_B(\underline{\theta}) = \frac{\text{tr}[\mathbf{H}^H(\underline{\theta})\mathbf{C}_{YY}\mathbf{H}(\underline{\theta})\mathbf{C}_{SS}]}{\text{tr}[\mathbf{H}^H(\underline{\theta})\mathbf{H}(\underline{\theta})\mathbf{C}_{SS}]} \quad (5)$$

This is the BB Bartlett processor for generic assumptions on the emitted signal component in terms of the cross-frequency structure. Other functions can be obtained by working out assumptions on \mathbf{C}_{SS} comprehending either uncorrelated or fully correlated frequency components.

C. The BB minimum-variance processor

The Bartlett processor generally has important limitations in terms of sidelobe attenuation. This might become a major difficulty in multiparameter estimation problems, when several unknown parameters are considered. As an attempt to alleviate such limitation Capon²³ proposed a processor commonly known as Minimum Variance Distortionless Response (MVDR) processor. The derivation of the broadband MV processor is well documented in the literature and follows a similar notation as that for the above-presented BB Bartlett processor resulting as

$$P(\underline{\theta}) = \frac{\text{tr}[\mathbf{H}^H(\underline{\theta})\mathbf{H}(\underline{\theta})\mathbf{C}_{SS}]}{\text{tr}[\mathbf{H}^H(\underline{\theta})\mathbf{C}_{YY}^{-1}\mathbf{H}(\underline{\theta})\mathbf{C}_{SS}]} \quad (6)$$

With regard to calculations, the MV processor presents the need to invert the SDM \mathbf{C}_{YY} , which can be done in a straightforward fashion provided that the SDM is of rank KL . In practice, this requires the number of snapshots of the received signal Y to be equal or larger than KL for calculating the sample SDM. Otherwise, it may be necessary to diagonal overload the SDM, as suggested in Ref. 25.

D. The BB MUSIC processor

The BB data model has been discussed in Sec. II A in the context of channel variability and ocean inhomogeneities, raising the question of the rank of the signal matrix \mathbf{C}_{SS} , which is equivalent to the signal subspace dimension. Historically, the subspace approach has been reported in the framework of classical beamforming for direction-of-arrival estimation and detection of emitters where the signal subspace dimension is the number of independent emitters detected. The present case consists of a single emitter radiating

at several frequencies. In this context the dimension of the signal subspace is related to the degree of spectral coherence of the acoustic field at the discrete frequencies of interest—hence a measure of the cross correlation of those spectral components, while the number of frequencies considered is always known. In general the SDM defined in Eq. (3) can be expressed in terms of the eigendecomposition

$$\mathbf{C}_{YY} = \mathbf{U}_S \mathbf{\Lambda}_S \mathbf{U}_S^H + \sigma_N^2 \mathbf{U}_N \mathbf{U}_N^H, \quad (7)$$

where the data space is separated into signal and noise subspaces. This is an ordinary eigenfactorization with the fact that the eigenvalues and eigenvectors appear separated, with the subscripts S and N denoting signal subspace and noise subspace, respectively.

The idea would be to hypothesize a geometric solution for the eigenproblem, in particular, concerning the signal subspace represented by $\mathbf{\Lambda}_S$ and \mathbf{U}_S . However, since the data model assumes an arbitrary signal subspace dimension varying from 1 to K , the best that can be asserted is that the span of the signal subspace is the same as that of the columns of

$$\frac{\mathbf{H}(\underline{\theta}_0) \mathbf{C}_{SS}^{1/2}}{\sqrt{\text{tr}[\mathbf{H}(\underline{\theta}_0) \mathbf{C}_{SS} \mathbf{H}^H(\underline{\theta}_0)]}}. \quad (8)$$

The dimension of the signal subspace is equal to the rank of \mathbf{C}_{SS} . This defines the signal subspace in agreement with Eq. (3), but according to Schmidt²⁴ the signal subspace can also be defined by its orthogonal complement—the noise subspace. This is acceptable due to the orthogonality between the columns of \mathbf{U}_S and \mathbf{U}_N in Eq. (7), i.e., $\mathbf{U}_S \mathbf{U}_N^H = \mathbf{0}$. Thus, as the span of \mathbf{U}_S is that of Eq. (8), the condition

$$\mathbf{U}_N^H \frac{\mathbf{H}(\underline{\theta}_0) \mathbf{C}_{SS}^{1/2}}{\sqrt{\text{tr}[\mathbf{H}(\underline{\theta}_0) \mathbf{C}_{SS} \mathbf{H}^H(\underline{\theta}_0)]}} = \mathbf{0} \quad (9)$$

is verified. The eigenvectors of the SDM \mathbf{C}_{YY} are separated into signal and noise eigenvectors as in Eq. (7), and the so-called orthogonal projector onto the noise subspace is given as $\mathbf{\Pi}^\perp = \mathbf{U}_N \mathbf{U}_N^H$. The MUSIC processor is defined as

$$P_{\text{MUSIC}}(\theta) = \frac{\text{tr}[\mathbf{H}^H(\theta) \mathbf{H}(\theta) \mathbf{C}_{SS}]}{\text{tr}[\mathbf{H}^H(\theta) \mathbf{\Pi}^\perp \mathbf{H}(\theta) \mathbf{C}_{SS}]}, \quad (10)$$

such that the solution parameter occurs at the maximum of $P_{\text{MUSIC}}(\theta)$. The degree of the solution uniqueness will certainly depend on the dimension of the signal subspace since the orthogonality in Eq. (9) works as a constraint of the solutions satisfying that condition. The smaller the signal subspace dimension the larger the dimensionality of that constraint, thus, reinforcing the solution uniqueness. The dimensionality of the noise subspace will in general be high. In theory, estimates of an arbitrary accuracy can be obtained if the observation time is sufficiently long, if the signal-to-noise ratio (SNR) is adequate, and if the signal model is sufficiently accurate. The main limitations of this method are the failure to correctly estimate the parameter with a low number of observations and a poor SNR. This method has been credited as being highly sensitive to model mismatch.

Finally, it should be noted that in practice only a sample SDM $\hat{\mathbf{C}}_{YY}$ is available. Thus, in Eqs. (5) and (6) \mathbf{C}_{YY} must be replaced by $\hat{\mathbf{C}}_{YY}$, and in Eq. (10) $\mathbf{\Pi}^\perp$ must be replaced by $\hat{\mathbf{\Pi}}^\perp$.

E. Estimating the signal matrix

Earlier, three matched-field processors based on the broadband data model were developed. However, the development was carried out assuming full knowledge of the signal matrix. In practice, knowledge of the emitted signal is often not available, or such knowledge may be useless due to unmodeled ocean inhomogeneities or variability in the channel response. This leads to the requirement of estimating the signal matrix \mathbf{C}_{SS} , which is analogous to deconvolution.^{26,27} Classical deconvolution assumes full knowledge of the source location and environmental parameters, which is not the case in environmental estimation problems.

The estimation of the signal matrix can be based on the signal subspace. Let $\mathbf{C}_{XX} = \mathbf{H} \mathbf{C}_{SS} \mathbf{H}^H$ be the signal component of the SDM defined in Eq. (3). This can be estimated together with σ_N^2 . Using the eigenvalue representation $\lambda_1 \geq \dots \geq \lambda_M$ and the orthonormal eigenvectors \underline{u}_i , ($i=1, \dots, M$) of $\mathbf{C}_{YY}(\underline{\theta}_0)$ spanning the signal subspace, and assuming that $\lambda_M > \lambda_{M+1} = \dots = \lambda_{KL} = \sigma_N^2$, one can write

$$\mathbf{C}_{XX} = \sum_{i=1}^M (\lambda_i - \sigma_N^2) \underline{u}_i \underline{u}_i^H,$$

where M is the dimension of the signal subspace. Optimum estimates of the eigenvalues λ_i and the eigenvectors \underline{u}_i can be obtained from the sample SDM $\hat{\mathbf{C}}_{YY}(\underline{\theta}_0)$, and an optimum estimate of σ_N^2 can be obtained by²⁸

$$\hat{\sigma}_N^2 = \frac{\text{tr} \hat{\mathbf{C}}_{YY} - \sum_{i=1}^M \hat{\lambda}_i}{KL - M}, \quad (11)$$

which is equivalent to the arithmetic mean of the $KL-M$ smallest eigenvalues $\lambda_i, i=M+1, \dots, K$ of $\hat{\mathbf{C}}_{YY}$. Now the estimate $\hat{\sigma}_N^2$ can be used in Eq. (11) to estimate \mathbf{C}_{XX} . Finally, the estimate of \mathbf{C}_{SS} proceeds by filtering out the channel response:

$$\hat{\mathbf{C}}_{SS} = \mathbf{H}^+(\hat{\underline{\theta}}_0) \hat{\mathbf{C}}_{XX} [\mathbf{H}^+(\hat{\underline{\theta}}_0)]^H. \quad (12)$$

In Eq. (12) one problem persists: Generally $\underline{\theta}_0$ is unknown, and the deconvolution algorithm cannot be completed. In the framework of parameter estimation one can replace $\underline{\theta}_0$ with $\underline{\theta}$, making $\hat{\mathbf{C}}_{SS}$ dependent on $\underline{\theta}$, and then replacing it in the processor expressions obtained earlier. It has been observed with synthetic data that the lack of knowledge on the emitted wave form will lead to a drawback in the parameter estimation performance. Knowledge on the structure of the emitted wave form can be seen as *a priori* information entering the parameter estimation algorithm.

III. SIMULATIONS

In Sec. II, three matched-field processors based on a broadband data model were developed. The Bartlett proces-

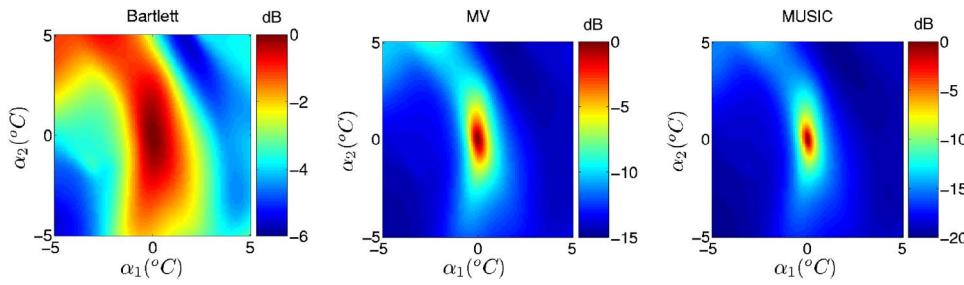


FIG. 1. (Color online) The behavior of the broadband processors for the coherent case assuming an unknown signal matrix.

processor is simply based on correlations—its implementation is straightforward. The MV and MUSIC high-resolution processors go beyond simple correlations and their computation requires additional steps. The computation of the MV processor involves the inversion of the SDM matrix, which has been noted as a difficulty in the MFP literature. The MUSIC processor involves splitting the data into signal and noise subspaces, whose correct estimation is fundamental for its success. So far, high-resolution methods have not been applied to experimental data with the purpose of performing environmental inversions. The objective of the following is to perform a simulation study in order to compare the three proposed processors covering issues such as sidelobe structure; performance for different data model assumptions; and the influence on the genetic algorithm’s performance. Although all three processors will be analyzed with the same depth, the Bartlett processor will be seen as the reference in terms of performance and some more focus will be on the high-resolution methods, since these somewhat constitute a novelty for this application.

The synthetic data are generated using an environmental model for a shallow water scenario similar to that of the North Elba site. The forward problem is solved using the normal modes propagation model C-SNAP.²⁹

A. Portraying MF processors as cost functions

A matched-field processor can be seen as a function of the hypothetical parameter vector θ , and is usually called cost function in the context of inverse problems. Concerning the behavior of a processor, assuming absence of noise and model mismatch of any type, one of the key characteristics is the ratio between the maximum value of the processor and the sidelobes, which has traditionally been considered an important issue in the context of a processor’s robustness against noise. Herein the importance of that issue is reinforced in the context of the optimization problem with aid of a genetic algorithm. The main interest is to illustrate how the matched-field processors obtained in Sec. II compare in terms of sidelobe attenuation and resolution, and how the assumptions of known or unknown wave form, or the assumption of coherent or incoherent signals impact on these characteristics. This can be carried out numerically by calculating each cost function as a function of physical parameters of interest for a given scenario example. The source was supposed to be at a 6 km range and at a 60 m depth, and receivers were at depths 15, 60, and 75 m ($L=3$). The acoustic field was considered for frequencies 400, 450, and 500 Hz ($K=3$). The spectral density matrix was computed

using Eq. (3). The noise power σ_N^2 was set as the mean of the first K eigenvalues λ_k of the \mathbf{C}_{XX} matrix (the SDM of the signal component) in order to obtain autofrequency SDMs with the same SNR, in both coherent and incoherent cases. The noise is assumed uncorrelated both across space and frequency. Note that under the assumption of coherent signals the eigenvalues of \mathbf{C}_{XX} $\lambda_1 > 0$ and $\lambda_2 = \dots = \lambda_K = 0$, and for the incoherent case, in general, $\lambda_k > 0$. Note also that for computing the SDM, $\mathbf{C}_{SS} = \mathbf{I}$ for the coherent case, and $\mathbf{C}_{SS} = \mathbf{I}$ for the incoherent case. Some more remarks are necessary before proceeding: (a) In Sec. II, the emitted wave form will always be represented by second-order statistics, i.e., by matrix \mathbf{C}_{SS} ; (b) in Sec. II A, for computing the cost functions, and, in particular, for estimating the signal matrix or for estimating the noise subspace, it is assumed that the dimension of the signal subspace is known, which is 1 in the coherent case and K in the incoherent case.

The cost functions were computed as a function of two coefficients α_1 and α_2 used to parametrize the temperature profile (see the following details) with $[\alpha_1 \alpha_2]^T = [00]^T$ as true parameter values. Figure 1 shows the matched-field response of the three BB processors computed for the case assuming coherent spectral components and unknown signal structure. The three plots respectively correspond to implementations of Eqs. (5), (6), and (10), together with Eq. (12) for estimating the signal matrix. Observing Fig. 1, the plot on the left contains a very smooth function as is typical for the Bartlett processor, with a variation between minimum and maximum values of 6 dB. The plot in the middle corresponds to the coherent MV processor, which is clearly superior to the Bartlett processor in terms of sidelobe attenuation, with values ranging by about 15 dB. Finally, a plot corresponding to the MUSIC processor was computed. This processor has the best sidelobe attenuation performance of all, with values ranging between -20 and 0 dB. The reader might ask how this was done if this processor approaches ∞ as the parameter vector θ approaches the true value, under the conditions used for generating the synthetic data. It is possible to portray the MUSIC processor such that its maximum is 1 by

$$P_{\text{MUSIC},1} = \frac{\gamma}{\gamma + \frac{1}{P_{\text{MUSIC}}}}. \quad (13)$$

It can be seen that $P_{\text{MUSIC},1} \rightarrow 1$ as $P_{\text{MUSIC}} \rightarrow \infty$. This modification has two effects: (1) The maximum value that can be attained is a known finite value; and (2) implementation allows for smoothing the processor. Small values for γ will produce a peaky function, while large values for γ will pro-

TABLE I. Peak-to-surface average ratio obtained for the different processors.

	Coherent known signal	Coherent unknown signal	Incoherent known signal	Incoherent unknown signal
Bartlett	2.23	1.86	1.64	1.58
MV	17.9	14.5	6.08	5.41
MUSIC	53.0	41.9	41.9	37.2

duce a smooth function. In this study, γ has been always set to 0.01.

Table I summarizes the results obtained in terms of peak-to-surface average ratio for different combinations of coherent/incoherent and known/unknown signal matrix. This measure is the ratio between the surface maximum and its average MF response. It is easy to conclude that there is an increasing discriminating potential when presenting the developed methods in that sequence. The coherent MV processor shows the highest effectiveness in comparison with its incoherent counterpart. For the other two processors coherent processing with unknown signal gives similar performance to incoherent processing with known signal.

B. Error performance

The next issue to be investigated is how the coherent processors perform against the number of signal snapshots N for a given SNR. Here the parameter to be estimated is α_1 , whose true value is 0. Once again the three frequencies/three receivers case is taken with a SNR=0 dB. Figure 2 shows plots with computations of the RMSE as a function of the number of snapshots N based on 100 estimates of the parameter. In no case is the Cramer-Rao lower bound attained. Figure 2(a) shows the RMSE with known signal matrix and signal subspace dimension. The MUSIC and Bartlett processors perform similarly for a known signal matrix. The MV processor has poor performance for low number of snapshots and recovers comparatively to the others as the number of snapshots increases. Finally, in the case of unknown signal matrix and signal subspace dimension, Fig. 2(b) shows how the low number of snapshots can impact on the performance of subspace-based methods, in particular, on the separation of the subspaces, in a conjunction with low SNR and low number of snapshots. For the other two processors, working with unknown signal slightly increases the RMSE.

The difficulties seen with the high-resolution methods are related to poor estimates of the eigenvalues λ_i . The data model assumes that the eigenvalues associated with the noise subspace are all equal. However, if N is finite, those will be different with probability 1. The MV processor requires inversion of the SDM, whose accuracy depends on the eigenvalues' estimates. Since it weights the eigenvector associated with the smallest eigenvalue most heavily, and this is the least stable vector because it has the least energy and must be orthogonal to all others, a small N is certainly a source of relatively poor performance. Concerning the MUSIC processor, the problem arises when the signal space is to be split into the signal and noise subspaces. This important issue is

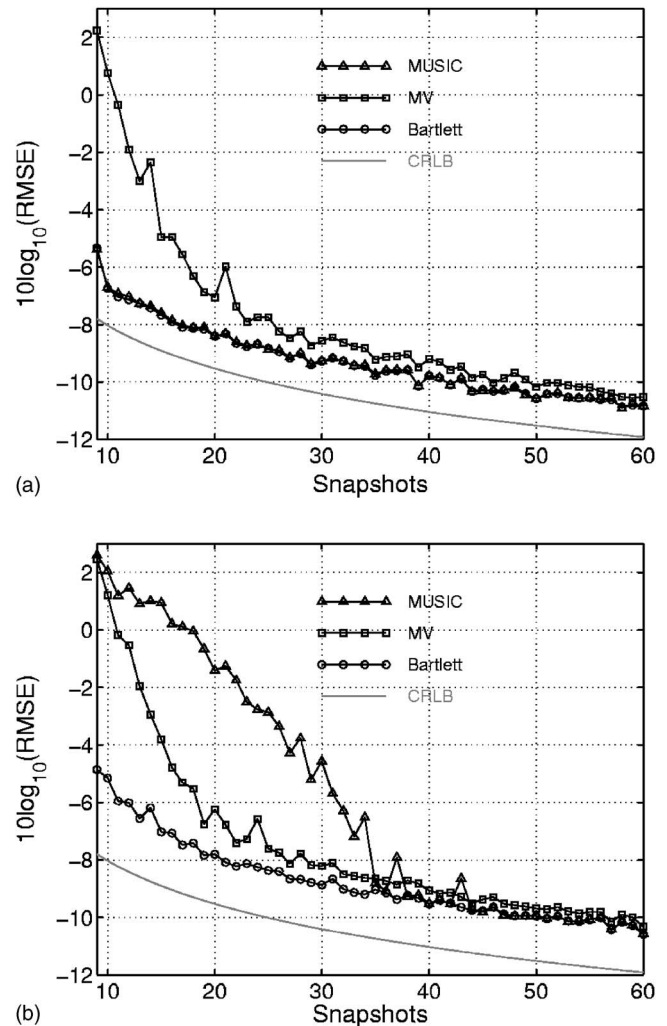


FIG. 2. RMSE as a function of the number of snapshots for the three processors and the coherent model Cramer-Rao lower bound under comparison: (a) With known signal matrix and (b) with unknown signal matrix and signal subspace dimension.

illustrated in Fig. 3, which shows in (a) the computed eigenvalues for two particular cases, one using 9 signal realizations, and the other using 11. The signal subspace dimension is estimated using the MDL criterion.³⁰⁻³² The former case yields a signal subspace with dimension eight, although a deterministic signal component is used for data generation. Note that the smallest eigenvalue has a very high ratio to its predecessor. The latter case, $N=11$, yielded a signal subspace with dimension one—in that case the ratios between contiguous higher-order eigenvalues are reduced. Figure 3(b) shows the average eigenspectrum when the number of realizations varies from 9 to 20. For each case 100 realizations of the eigenspectrum were computed and averaged. On average, the eigenspectrum tends to become flattened as $N \rightarrow \infty$. Finally, Fig. 3(c) shows the average order estimate obtained using the same data as in Fig. 3(b), applying the MDL information criterion. For the minimal number of signal realizations (9) the average order obtained is about 3; for 10 realizations it is about 2; for 11 realizations or more it estimates on average the correct value, which is 1. This exercise illustrates the potential impact of the number of signal realizations on methods relying on the eigendecomposition of the data.

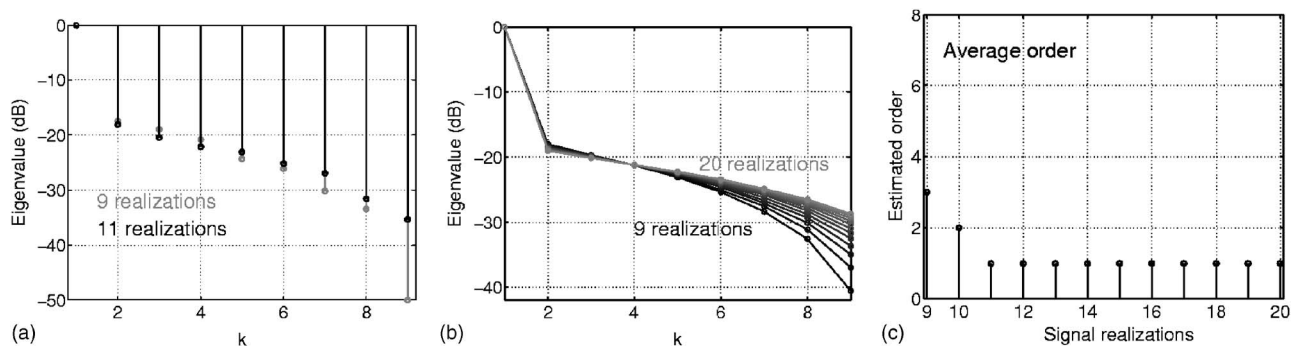


FIG. 3. Eigenspectra for finite number of signal observations: (a) Comparison of two eigen-spectra using $N=9$ (gray) and $N=11$ black; (b) average eigenspectrum for a varying number of signal realizations; and (c) average order estimation for a varying number of signal realizations.

C. Global search

In Sec. III A it is shown that all three proposed MF processors significantly differ in their ability of attenuating sidelobes. The next case study is to conclude about the comparative performance of the processors when the environmental estimation problem includes multiple unknown parameters. This problem is usually solved with the aid of a global search method such as a genetic algorithm. The idea is to find out whether high-resolution methods can improve the convergence of meta-heuristic search methods due to their reduced sidelobe structure. Sidelobes competing with the main peak may be seen as false attractors that cause difficulties to any meta-heuristic search method in converging to the solution maximizing the cost function.

The data are generated and inverted 50 times using a genetic algorithm. The inversion search space regarded the water column and the seafloor properties, but array tilt was also included. The signal matrix and the signal subspaces dimension were known. The general conditions for synthetic data generation are the same as those used earlier with SNR of 0 dB and 16 snapshots.

A posteriori distributions can provide insight into the performance of the environmental inversion process. These distributions emphasize the variability of each parameter over the search interval, which is intimately related to the ambiguity pattern of the cost function used and the sensitivity to each parameter. Figure 4 shows the *a posteriori* distributions obtained for the three methods. To obtain these distributions the individuals of the last generation of all independent populations are merged and histograms are computed from the parameter vectors represented by those individuals.

The idea of showing all these distributions is to obtain a global comparison of the three processors in terms of convergence rather than performing a detailed analysis. The MUSIC processor clearly has the narrowest distributions. In fact, that processor contributed to improving the convergence of the genetic algorithm, supporting the belief that a massive sidelobe structure causes difficulties in terms of the population convergence. The MV processor appears to be more uncertain, which is rather attributed to the problem of inverting the SDM with a small N than to the sidelobe structure. Finally, the distributions obtained with the Bartlett processor are significantly more spread out over the search interval

than the others, which is attributed to its massive sidelobe structure causing difficulties for the search algorithm in converging to the true solution. From the *a posteriori* distributions, model estimates based on the distribution peak, called *Maximum A Posteriori* (MAP) estimates, can be obtained. The MUSIC processor produced more reliable MAP estimates, since its parameter distributions are the most compact, and all parameters, except sediment upper speed and subbottom density, have *a posteriori* distributions with a peak close to the true parameter value (indicated by the gray asterisk).

IV. ENVIRONMENTAL INVERSION OF EXPERIMENTAL DATA

A. The MREA'03 sea trial

The Maritime Rapid Environmental Assessment 2003 (MREA'03) sea trial took place from 26 May to 27 June 2003, in the Ligurian Sea, with target areas North and South of Elba Island. This paper considers only the acoustic experiment held on 21 June 2003³³ whose area of operation was North of the Elba Island as shown in Fig. 5.

1. The deployment geometry

On 21 June, the AOB was deployed on a free drift configuration with very favorable weather conditions in an area of mild bottom range-dependency, attaining a variability of 20 m over some acoustic tracks. The experimental setup consisted of a towed acoustic source and a free drifting vertical line array with receivers at nominal depths of 15, 60, 75, and 90 m. Figure 5(b) shows the bathymetry in the interior of the white box depicted in the map of Fig. 5(a) together with the source ship navigation and AOB drift estimated from GPS recordings. The acoustic buoy was deployed at 09:01 GMT and recovered at about 15:16 GMT. During this time the buoy drifted about 1.7 km away to the Southeast of the point of deployment, at approximate average displacement of 4.5 m/min (white dashed line).

The acoustic source was deployed immediately after the acoustic buoy and towed by the RV Alliance to West where it was stalled between times 09:53 and 11:05 GMT. Then RV Alliance steadily moved to the East and performed the geometry shown in Fig. 5(b) (white solid line) until source recovery. Figure 6(a) shows the GPS estimated range be-

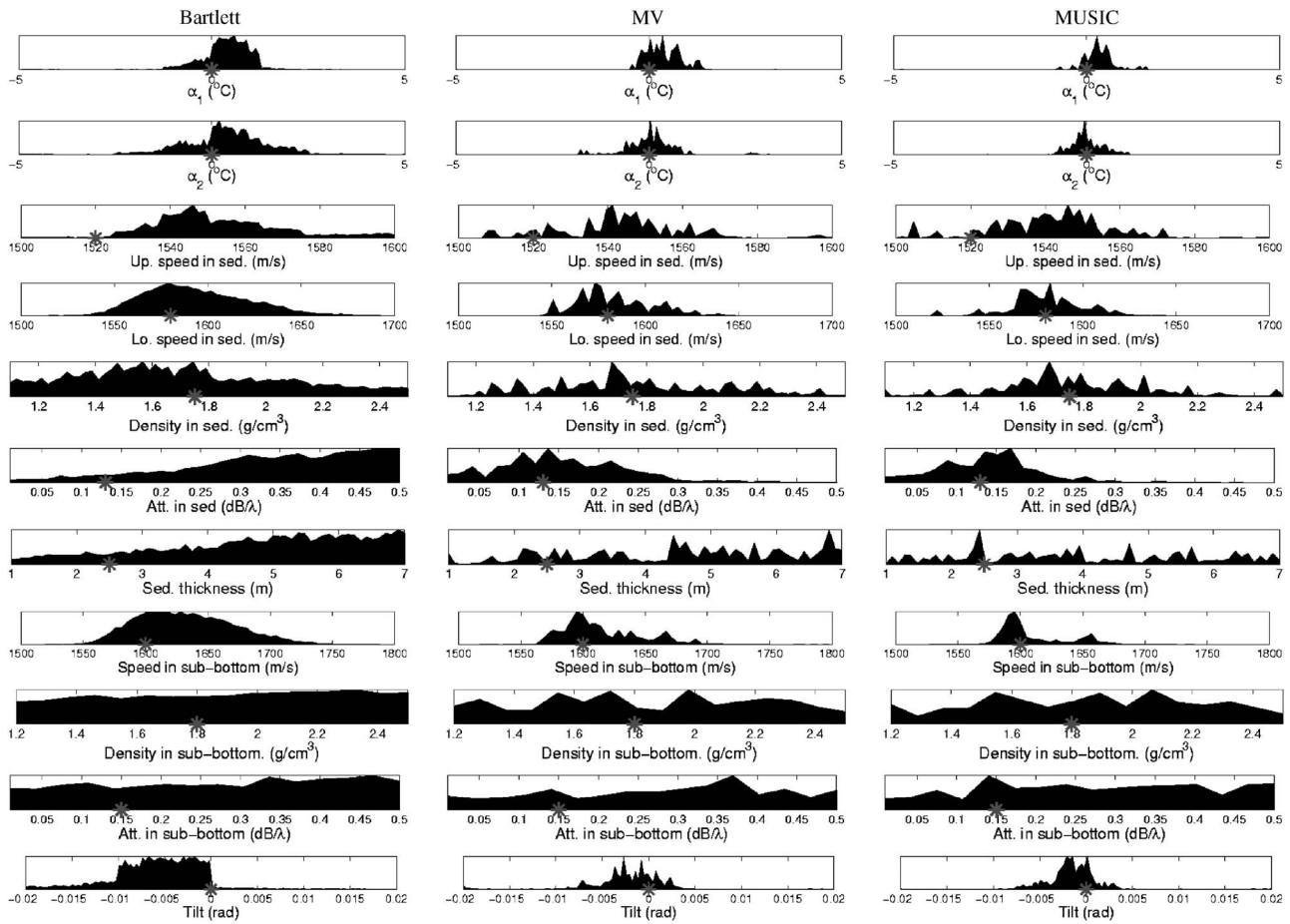


FIG. 4. *A posteriori* probability distributions for each parameter based on the last generation of 50 independent populations. Each column corresponds to processors entering the comparison. The gray asterisks indicate the correct parameter value.

tween source and receiver and Fig. 6(b) the source depth. The acoustic source was deployed at a variable depth, between 54 and 106 m, depending on ship speed.

2. Acoustic signals

The emitted wave forms consisted of 2-s LFM chirps emitted in two frequency bands. The *A1* and *A1double* chirps are in the band 500–800 Hz. The signals differ in terms of repetition interval and duty cycle: *A1* lasts for 2 s and has a repetition interval of 8 s hence a duty cycle of 25%; *A1double* lasts for 4 s and has a repetition interval of 10 s hence a duty cycle of 40%. The objective of emitting the *A1double* wave form was to increase the number of signal realizations in a given observation interval, which may significantly impact on the performance of some matched-field processors. The *A2* chirp is in the band 900–1200 Hz. Table II shows the emission schedule indicating the periods during which each wave form was emitted. The signals were received at a vertical array containing four hydrophones at nominal depths of 15, 60, 75, and 90 m. Figure 7 shows an example of receptions of the *A2* chirps at time 11:45 GMT collected at the third receiver. The data are extremely clean without blanks and interruptions, which results from the AOB’s local storage capability. However, it was found that the acoustic data collected by the deepest hydrophone was very noisy most of the time, possibly due to deployment

issues that were not well understood. For this reason it was decided to not consider that hydrophone in the present work. A final remark is that a channel fading effect that significantly reduces the signal received on the top most hydrophone was noticed, probably due to the effect of the thermocline.

3. Environmental data measurements

Regarding the inversion of the acoustic data collected on 21 June, 95 conductivity-temperature-depth (CTD) measurements taken during the days 16, 17, and 19 June taken at the positions marked by the black circles in Fig. 5 are considered in this study. No CTD measurements were performed during that acoustic experiment since RV Alliance was towing the source. There is a significant difference in scale between the acoustic (white box) and the oceanographic survey, as the latter was set up for different purposes. Those CTD measurements are being taken as an attempt to cope with the difficulty in sampling the ocean volume, both in time and space, and obtain representative *a priori* oceanographic data as an input to the acoustic inversion problem. It is of concern to what extent these historical data collected several days before may be representing the oceanography of the target day. Figure 8 shows the measured temperature profiles with two empirical orthogonal functions (EOF) representing more than 80% of the water column variability. The water column

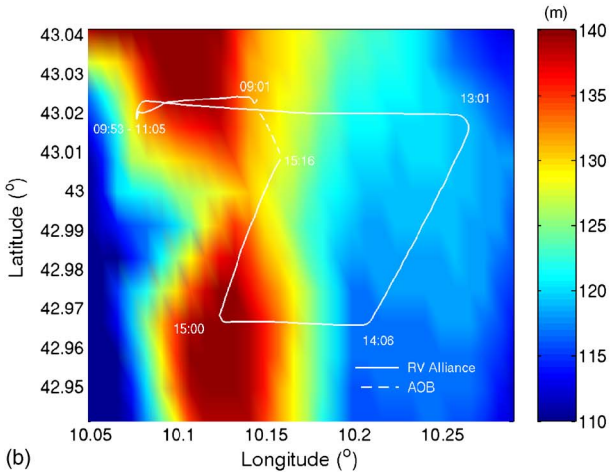
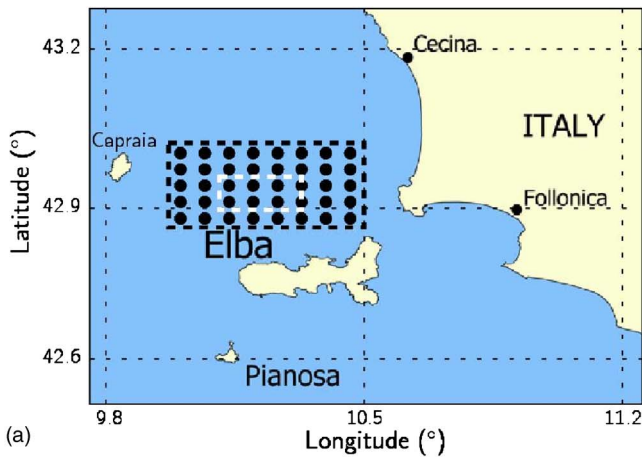


FIG. 5. (Color online) The Maritime Rapid Environmental Assessment 2003 (MREA'03) experimental area: (a) Black circles indicate the sampling grid setup for the CTD measurements used in this study, and the dashed white box limits the area where the acoustic experiment of 21 June took place and (b) GPS estimated source ship navigation (white solid curve) and AOB drift (white dashed curve) during the deployment of 21 June.

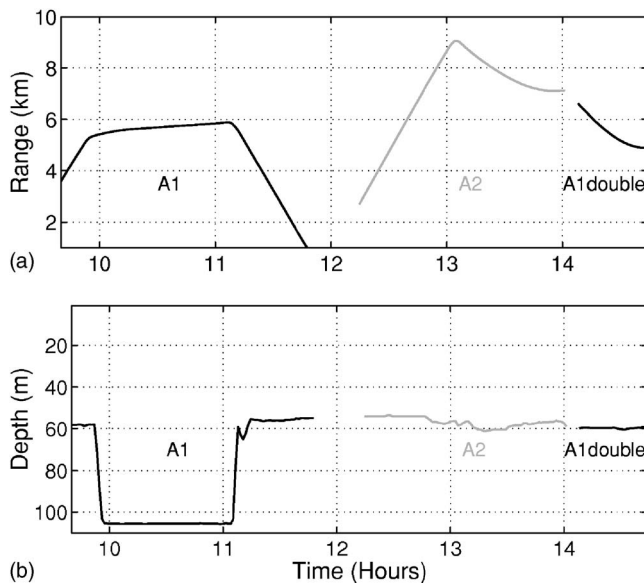


FIG. 6. Source range (a) and depth (b) measured during the deployment of 21 June. The curves are broken indicating change of the emitted wave form. A1, A2, and A1double denote the wave forms emitted in each interval.

TABLE II. Signal emission schedule on 21 June. The times are in GMT.

	A1	A2	A1double
Start	09:40	12:14	14:07
End	11:47	14:01	14:44

temperature is then modeled as a sum of the mean temperature (thick curve) and the two EOFs weighted by associated EOF coefficients α_n . It is useful to measure the coefficients for the historical data available in order to obtain hints on their range of variation. In the present case α_1 varied in the interval from -15 to 15 , and α_2 varied in the interval from -5 to 5 temperature profiles considered.

4. The environmental model

One of the tasks with the largest impact on the final result is the choice of an adequate environmental model to represent the propagation conditions of the experiment. This choice is generally the result of a compromise between a detailed, accurate, and parameter full model and a light model ensuring a rapid convergence during the processing. The baseline computer model adopted for the MREA'03 was built based on the segmentation of archival bathymetric information along the source-receiver cross sections at different times. As shown in Fig. 5 the bathymetry in the experimental area is accurately known. The water depth at the AOB deployment site was approximately 120 m and the maximum depth at the emitting source was 140 m. The baseline geoaoustic properties were drawn from previous studies in that area.^{10,34} The baseline model consists of an ocean layer overlying a sediment layer and a bottom half space with the bathymetry assumed range dependent, as shown in Fig. 9. The sound-speed profile was calculated using the Mackenzie formula with the mean temperature and mean salinity profiles as inputs (see Fig. 8).

For the purposes of the inversion the forward model was divided into four parameter subsets—water column temperature, sediment, subbottom, and geometric parameters. The temperature in the water column is parametrized by the two EF coefficients as discussed earlier.

B. Results

The following reports on environmental inversions of the experimental acoustic data for water column and seafloor properties using the broadband processors proposed in Sec. II, and the comparison of their estimation performance.

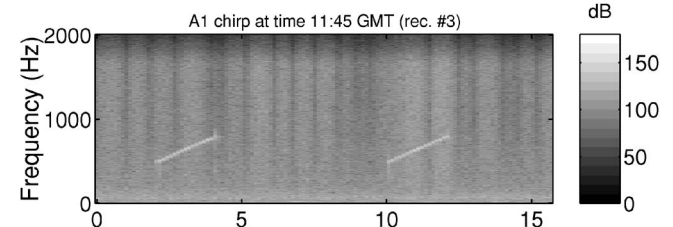


FIG. 7. Example of an A1 chirp received on the AOB.

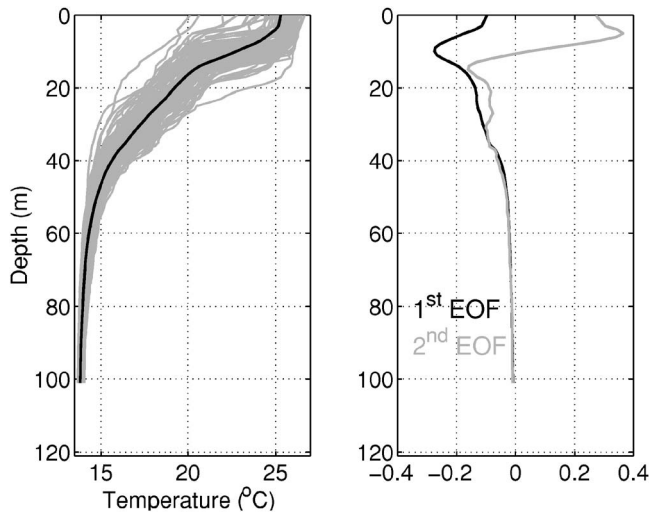


FIG. 8. CTD-based data used for temperature estimation taken during 16, 17, and 19 June Temperature profiles with mean profile in solid black (left) and representative empirical orthogonal functions (EOF) computed from the temperature profiles (right).

1. Frequency clustering

The proposed BB matched-field processors are based on the $KL \times KL$ SDM C_{YY} measuring cross correlations of the acoustic field across space and frequency. Both the MV and MUSIC processors require C_{YY} to be full-rank, i.e., $N \geq KL$. If an observation window of 80 s is taken, then the MREA'03 data set provides $N=10$ for the A1 and A2 intervals, and $N=16$ for the A1double interval. Given a number of receivers $L=3$ and $N=10$, one can choose a number of frequencies $K=3$ in order to assure that full-rank SDMs are obtained in all emission intervals. In order to use an increased number of frequencies while assuring that the SDM is full rank one can use an alternative matched-field processor output given as

$$P^{N_g}(\underline{\theta}) = \frac{1}{N_g} \sum_{n=1}^{N_g} P(\underline{\theta}, \underline{\omega}_n). \quad (14)$$

This is an incoherent average over N_g coherent frequency clusters $\underline{\omega}_n$, where $P(\underline{\theta}, \underline{\omega}_n)$ is a given matched-field processor. The question is how to choose the frequency clus-

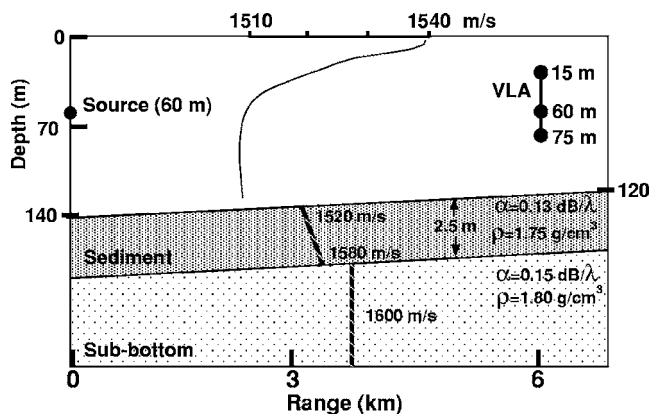


FIG. 9. Baseline model for the MREA'03 sea trial. All parameters except water depth are range independent.

TABLE III. GA settings for environmental inversion.

Parameter	Setting
Generations	30
Population size	200
Independent populations	1
Mutation probability	0.004
Crossover probability	0.9
Number of crossover points	4

ters. As the idea is to use coherent processors one can carry out an optimization aimed at finding SDMs $C_{YY}(\underline{\omega}_n)$ with the most coherent frequencies. Ideally, one would cluster perfectly coherent frequencies, which results in a signal subspace with dimension equal 1. Thus, it appears natural to implement an optimization scheme based on the minimization of the signal subspace dimension, using an information criterion as cost function. Since these criteria are not reliable for $N \approx KL$ it was decided to choose N_g cluster with the largest λ_1/λ_2 —the ratio between the two largest eigenvalues of $C_{YY}(\underline{\omega}_n)$ is a simplified measure of the signal's coherence.

This is a preprocessing step that performs a selection of frequency combinations based on a coherence criterion. In this study a number of clusters $N_g=7$ will be used, and the optimization was carried out using a frequency resolution of 4 Hz. In order to assure spectral diversity, frequencies in a cluster are separated by at least 52 Hz.

2. Data processing procedure

Several steps are performed until the inversion is complete:

- (1) Frequency selection based on the λ_1/λ_2 optimization criterion.
- (2) Acoustic field inversion for water column and seafloor properties, and geometric nuisance parameters.
- (3) Inversion validation by means of source localization with large search bounds using the estimated environmental models.
- (4) Reconstruction of physical parameters of interest using only environmental estimates validated in step (3).

In step (2) the unknown parameters are divided into water column (α_1 and α_2 EOF coefficients), sediment (upper and lower compressional speeds, density, attenuation, and thickness), and subbottom (compressional speed, density, and attenuation). Additionally, geometric parameters, array tilt, and receiver depth are included. These parameters are regarded as nuisance parameters, since there is no interest in their estimates once the inversion is finished. The inversion is posed as an optimization problem solved with aid of a genetic algorithm (GA).³⁵ The parameter vector is coded into a 68-bit chain, which results in a search space size approximately equal to 2.95×10^{20} . The GA settings are summarized in Table III. Since there is a new time bin every 80 s only a single population is used for each inversion, which is sufficient to achieve the main objective of comparing the proposed MF processors' inversion performance.

TABLE IV. Rates of successful localization (%) for the different processors and different wave forms.

Processor	A1	A2	A1double	Overall
Bartlett	22.7	38.3	51.9	32.7
MV	22.7	29.6	85.2	35.1
MUSIC	39.0	53.1	70.4	48.3

Step (3) is to validate the model estimates obtained in step (2) by means of range-depth source localization. This step is based on the accurate knowledge on source range and depth available and on the fact that source position is on top of the parameter hierarchy. It is assumed that if the environmental estimates are not accurate then the source cannot be properly located. Performing source localization with large search bounds should give an indication on the quality of a given environmental estimate.

Finally, step (4) is to produce the final environmental estimates using only those estimates validated in step (3).

3. Environmental inversion: Comparison of three MF processors

Here, the three BB processors will be applied to experimental field data. The whole data set collected on 21 June will be inverted with each processor. In Sec. III C, inversions on synthetic data with a GA indicated that high-resolution processors may contribute to improve the convergence to the true solution. The comparison performed here may also serve the purpose of understanding how these processors behave in a real situation with the inherent model mismatch. The maximum source-receiver range is 9 km, which is more than 70 water depths, possibly giving rise to environmental mismatch.

The inversion is carried out assuming an unknown signal matrix C_{SS} . Since the frequencies are optimized in the sense of clustering those with highest coherence, it is assumed that the signal subspace dimension is always one, although source and receiver are moving most of the time. The processing is, therefore, considered broadband coherent with unknown wave form.

No ground truth measurements are available for evaluating the inversion performance achieved. Thus, one can refer directly to step (3), the validation step, and analyze those results. Source localization along time was performed within search bounds from 1 to 10 km in range, and from 1 to 110 m in depth. The source is admitted as correctly localized if the error both in range and depth is simultaneously less than 5% of the search interval amplitude. In other words the maximum must fall in a 0.9-km-long and 11-m-wide rectangle centered on the true location. Table IV shows the rate of successful localization achieved with each processor. The localization rate was computed for each emission interval due to the different signal characteristics, and then summarized on the rightmost column as an overall result. Considering the overall results, the MUSIC processor clearly outperforms the other two processors, as the localization rate increases with the processors' resolution. A more detailed analysis of the result allows for making the following re-

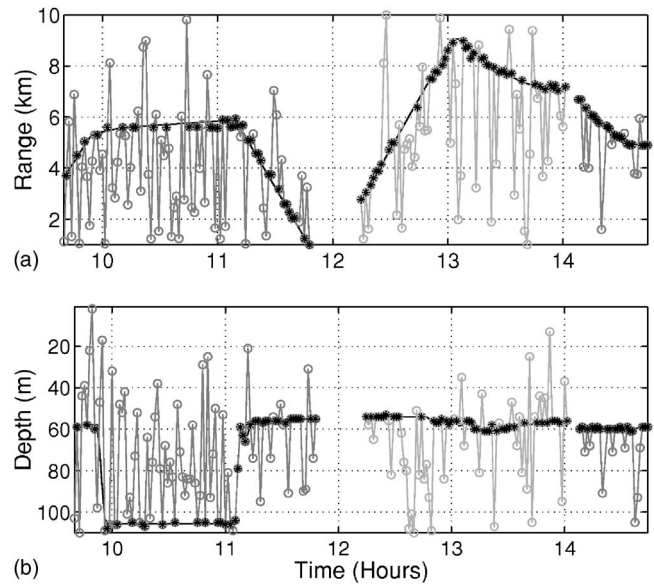


FIG. 10. Source localization obtained with the MUSIC processor. Source range (a) and source depth (b). True location is given by the black curve in the background. The gray curves with circles are the source localization results. The black asterisks indicate the successful localizations.

marks: (i) The *A1* interval was the most difficult. Close inspection shows that the localization rate is clearly lower with the source stalled at 105 m (during time 10:00 to 11:00) in the three cases. (ii) The MV processor performed remarkably well during the *A1double* interval. It appears that this processor performed particularly well with the number of signal realizations $N=16$ provided by that data portion, while showing difficulties in the other data portions providing $N=10$ (which is consistent with the synthetic study). (iii) The source was located both in range and depth at ranges up to 9 km, as indicated by the black asterisks in Fig. 10.

Figure 11 shows the environmental inversion results obtained with the MUSIC processor. The reason for showing the run performed with the MUSIC processor was its overall superiority in terms of source localization, and the belief that this means that the environmental estimations are also of superior quality in comparison with those obtained with the other processors. The circles filled with an asterisk correspond to inversions with successful localization.

It can be seen in plot (a) that from time 10:00 to time 11:00 the estimate of the α_1 EOF coefficient, a leading parameter, varies over the entire search interval. When the source moves and goes up to approximately midwater column, the variability substantially reduces, with the estimates becoming confined in the interval -10 to 10 . During *A2* the estimates continue in the interval -10 to 10 at the beginning but then during the remaining part the estimates are in the upper half of the search interval between 0 and 15 . Finally, the α_1 estimates in the *A1double* period are clearly in the interval -10 to 10 . The second EOF coefficient was searched in an interval with amplitude less than half of that for the first one. The EOFs used to model the water column differ significantly only in the first third of the water column, while they coincide for the remaining depth. On the other hand only one receiver is in place in the first third of the water

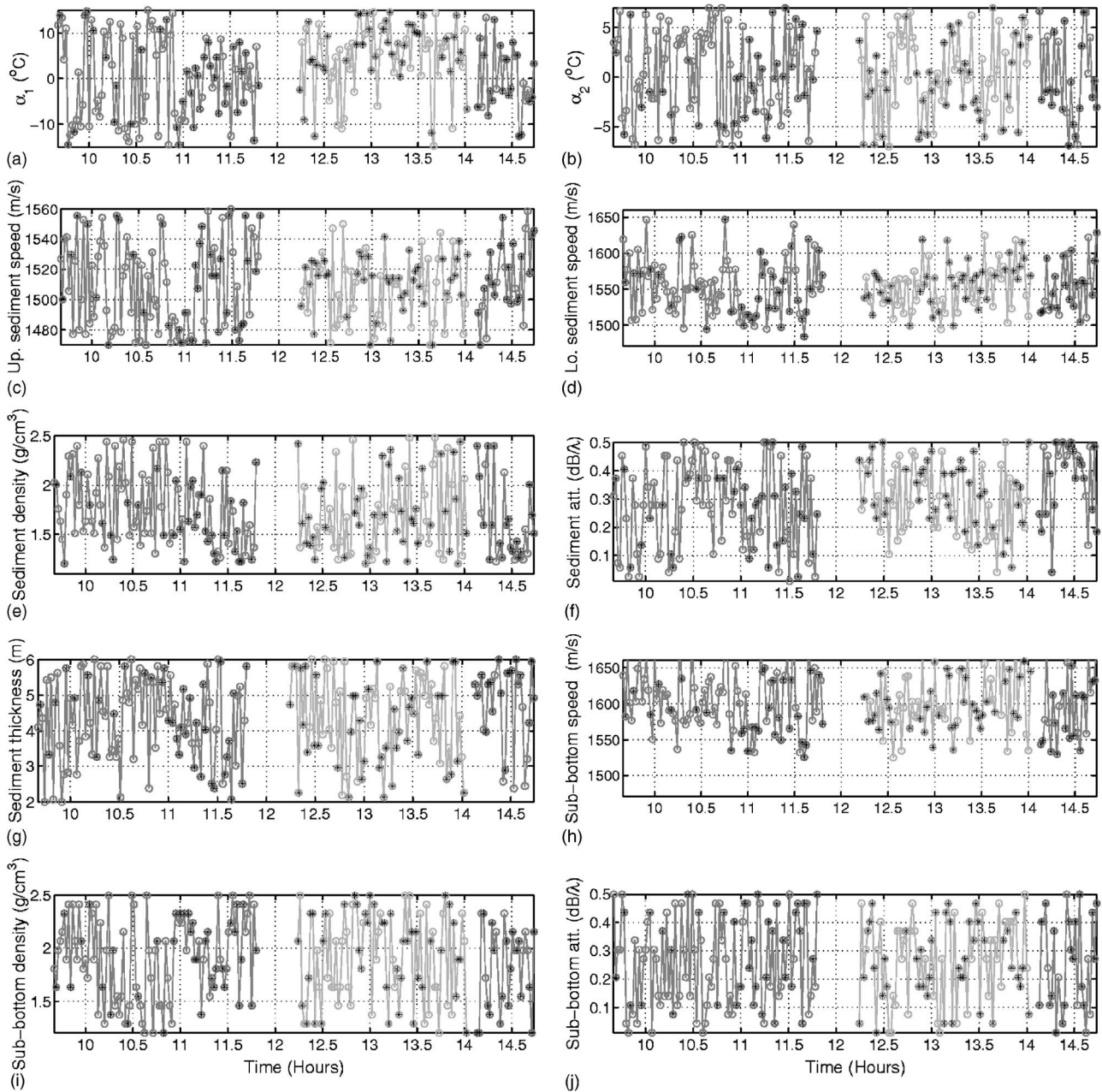


FIG. 11. Model parameter estimates obtained via acoustic data inversion using the BB MUSIC processor. Water column [(a) and (b)]; sediment [(c)–(g)]; and subbottom [(h)–(j)]. The black asterisks indicate model estimates allowing for successful source localization in the validation step.

column. In Ref. 36 it is shown that these circumstances lead to an ambiguous cost function regarding the α_i coefficients, resulting in a poor solution uniqueness.

Concerning the seafloor parameter estimates, most of them appear to be unstable, making it nearly impossible to draw some value from the plots of Fig. 11. Nevertheless, the compressional speeds in sediment and subbottom appear to be fairly restricted to subintervals in the search interval with aid of the source localization step. In order to obtain a single estimate for each of the seafloor parameters, *a posteriori* distributions based on the individuals of the last generation of each inversion allowing for correct source localization (43 out of 81 inversions) were computed considering only the *A2 emission period* (see Fig. 12). Sediment and subbottom compressional speeds, and sediment attenuation are relatively

compact, each with a peak close to the baseline value. Also, sediment density and thickness show an outstanding distribution peak. The sediment attenuation distribution, although compact, is concentrated at the upper search bound far from the baseline value. Table V contains the MAP estimates for the seafloor parameters, together with the baseline values and a measure of the estimation reliability, which is the standard deviation of the *a posteriori* distribution divided by the search interval length. The MAP estimates of the compressional speeds are in fair agreement with the baseline values. Sediment density and thickness are also credible.

Finally, it can be remarked that during the *A2 emission period*, the MUSIC processor produced the most reliable environmental estimates. During the *A1double emission period* the most reliable environmental estimates were produced by

TABLE V. Baseline seafloor parameters, parameter MAP estimates on 43 GA populations, and a reliability measure.

Mode parameter	Baseline	MAP	Reliability
Sediment			
Up comp speed (m/s)	1520	1527	0.22
Lo comp speed (m/s)	1580	1570	0.17
Density (g/cm ³)	1.75	1.48	0.27
Attenuation (dB/λ)	0.13	0.44	0.20
Thickness (m)	2.5	4.92	0.27
Subbottom			
Comp speed (m/s)	1600	1618	0.14
Density (g/m ³)	1.80	1.29	0.31
Attenuation (dB/λ)	0.15	0.40	0.29

the MV processor. This supports the close relation between the rate of source localization and the quality of the environmental estimation, and the choice of source localization as a validation tool.

V. CONCLUSIONS

Acoustic data were collected using a prototype of a light acoustic receiving system, an AOB, developed to meet operational requirements such that it consisted of a vertical line array with only three operating receivers. One of the objectives of these acoustic data was to evaluate the performance of newly adapted field inversion methods for the estimation of water column and geoacoustic properties using such a light acoustic receiving system.

The application of genetic algorithms for the estimation of model parameters has been shown to be very effective on several occasions. However, when a sparse receiving array is used the inversion problem becomes heavily ill-conditioned. Traditionally, the difficulty of solving an ill-conditioned problem has been associated with model mismatch and noise, which may lead to the situation where the optimum is not in coincidence with (or even close to) the true model parameters.

This paper associates this difficulty with the application of a genetic algorithm to carry out the optimization, specifically due to the existence of many local extrema in concurrence with the main peak, together with the typically large search space of a multiple environmental estimation problem. In order to cope with this difficulty broadband matched-field processors, specifically, high-resolution processors, are proposed for their increased ability in attenuating sidelobes.

A Bartlett, a minimum-variance, and MUSIC processor, all truly broadband rather than a superposition of multiple frequencies since they are based on a broadband data model, were first compared in terms of estimation performance with synthetic data. It was concluded that in the case of an exhaustive search of the unknown parameter, the high-resolution processors were unable to outperform the Bartlett processor. The main difficulty of the application of the high-resolution processors is related to a reduced number of signal realizations to compute the sample spectral density matrix (SDM). In the case of a multiparameter inversion problem the comparison shows that the high-resolution methods

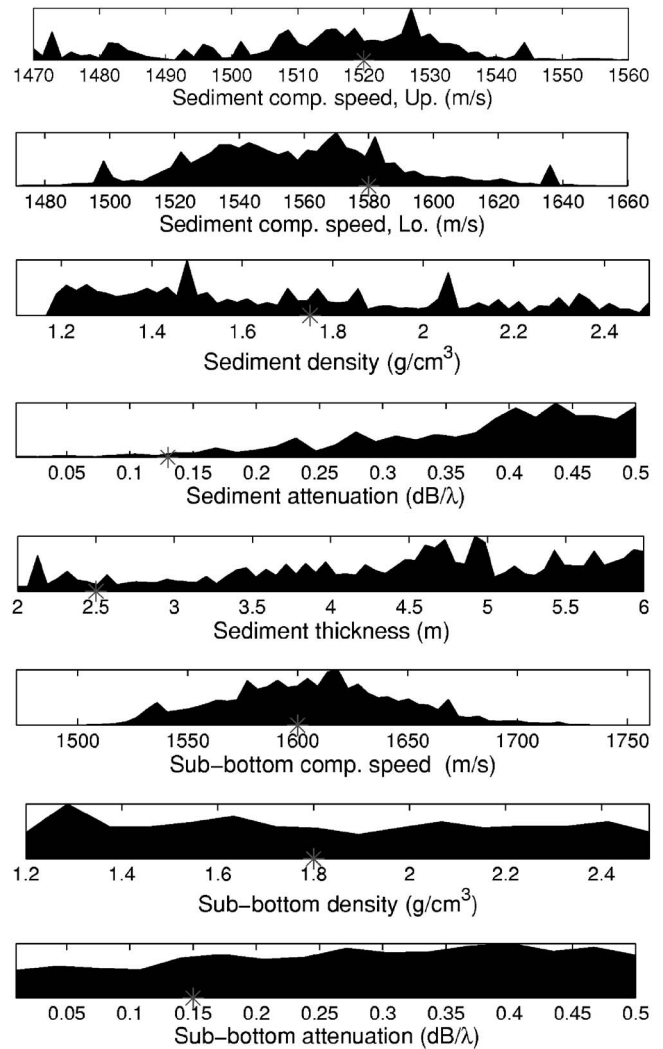


FIG. 12. *A posteriori* probability distributions for the seafloor parameters based on the last generation of the GA. Only inversions validated by means of source localization during the A2 period are considered. The gray asterisk indicates the baseline value of the parameter.

clearly contribute to improve the convergence of the genetic algorithm to the true model parameters, therefore reducing the inherent uncertainty.

The inversion results with the experimental data are consistent with those obtained with the synthetic data in terms of processor performance comparison. The inversion algorithm included a source localization step with large search bounds aimed at discarding model estimates upon wrong localization result. As no concurrent environmental ground truth data were available the processor performance was evaluated by means of the rate of correct source localization. Based on that criteria, the MUSIC processor achieved the best overall performance. Note that this processor was applied assuming that the dimension of the signal subspace was one, since a coherence optimization step was performed before. This contributed to avoiding a major difficulty with this processor—the estimation of the signal subspace dimension. The minimum-variance processor achieved an exceptional performance in an emission interval where a larger number of signal realizations were available. With this data set successful source localization results were obtained for ranges up to 9 km.

The EOF coefficients, used to parametrize the water column, were estimated with some uncertainty, although well restricted to subintervals over some periods. A single estimate of the seafloor properties was obtained by means of the maximization of the *a posteriori* distributions based on the last populations of the genetic algorithm. This result is essentially in line with the baseline model values for compressional speeds in the sediment and subbottom, and density in the sediment. The estimate of the sediment thickness is also well determined. Finally, a strong relation between uncertainty and source localization rate was observed (Fig. 12).

The dimension of the search space appears to be a major impairment for consistently obtaining valid model estimates (by means of correct source localization) over time since, in principle, this cannot be attributed to model mismatch or additive noise.

ACKNOWLEDGMENTS

The authors would like to thank the NATO Undersea Research Centre for the organization of the MREA'03 sea trial. This work was financed by FCT, Portugal, under fellowship SFRH/BD/12656/2003 and NUACE project, Contract No. POSI/CPS/47824/2002, and the development of the AOB prototype was funded by the Portuguese Ministry of Defense under the LOCAPASS project.

¹A. Tolstoy, *Matched Field Processing for Underwater Acoustics* (World Scientific, Singapore, 1993).
²A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," *IEEE J. Ocean. Eng.* **18**, 401–424 (1993).
³H. P. Buckner, "Use of calculated sound fields and matched-detection to locate sound source in shallow water," *J. Acoust. Soc. Am.* **59**, 368–373 (1976).
⁴A. Tolstoy, O. Diachok, and N. L. Frazer, "Acoustic tomography via matched field processing," *J. Acoust. Soc. Am.* **89**, 1119–1127 (1991).
⁵M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean-bottom properties," *J. Acoust. Soc. Am.* **92**, 2770–2783 (1992).
⁶P. Gerstoft, "Inversion of acoustic data using a combination of genetic algorithms and the Gauss-Newton approach," *J. Acoust. Soc. Am.* **97**, 2181–2190 (1995).
⁷S. E. Dosso, M. L. Jeremy, J. M. Ozard, and N. R. Chapman, "Estimation of ocean-bottom properties by matched-field inversion of acoustic field data," *IEEE J. Ocean. Eng.* **OE-18**, 232–239 (1993).
⁸C. E. Lindsay and N. R. Chapman, "Estimation of ocean-bottom properties by matched-field inversion of acoustic field data," *IEEE J. Ocean. Eng.* **OE-18**, 224–231 (1993).
⁹M. D. Collins and W. A. Kuperman, "Focalization: Environmental focusing and source localization," *J. Acoust. Soc. Am.* **90**, 1410–1422 (1991).
¹⁰D. F. Gingras and P. Gerstoft, "Inversion for geometric parameters in shallow water: Experimental results," *J. Acoust. Soc. Am.* **97**, 3589–3598 (1995).
¹¹P. Gerstoft and D. Gingras, "Parameter estimation using multi-frequency range dependent acoustic data in shallow water," *J. Acoust. Soc. Am.* **99**, 2839–2850 (1996).
¹²C. Soares, M. Siderius, and S. M. Jesus, "Source localization in a time-varying ocean waveguide," *J. Acoust. Soc. Am.* **112**, 1879–1889 (2002).
¹³M. Snellen, D. G. Simons, M. Siderius, J. Sellschopp, and P. L. Nielsen, "An evaluation of the accuracy of shallow water matched field inversion results," *J. Acoust. Soc. Am.* **109**, 514–527 (2001).
¹⁴R. M. Hamson and R. M. Heitmeyer, "An analytical study of the effects of environmental and system parameters on source localization in shallow

water by matched-field processing of a vertical array," *J. Acoust. Soc. Am.* **86**, 1950–1959 (1989).
¹⁵A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, "Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem," *J. Acoust. Soc. Am.* **80**, 571–587 (1998).
¹⁶A. Tolstoy, "Computational aspects of matched field processing in underwater acoustics," in *Computational Acoustics*, edited by D. Lee, A. Cakmak, and R. Vichnevetsky (North-Holland, Amsterdam, 1990), Vol. **3**, pp. 303–310.
¹⁷Z.-H. Michalopoulou, "Matched-field processing for broad-band source localization," *IEEE J. Ocean. Eng.* **21**, 384–392 (1996).
¹⁸Z.-H. Michalopoulou, "Source tracking in the Hudson Canyon experiment," *J. Comput. Acoust.* **4**, 371–383 (1996).
¹⁹G. J. Orris, M. Nicholas, and J. S. Perkins, "The matched-phase coherent multi-frequency matched field processor," *J. Acoust. Soc. Am.* **107**, 2563–2575 (2000).
²⁰C. Soares and S. M. Jesus, "Broadband matched field processing: Coherent and incoherent approaches," *J. Acoust. Soc. Am.* **113**, 2587–2598 (2003).
²¹A. Silva, F. Zabel, and C. Martins, "Acoustic oceanographic buoy: A telemetry system that meets rapid environmental assessment requirements," *Sea Technol.* **47**, 15–20 (2006).
²²S. M. Jesus, C. Soares, E. Coelho, and P. Picco, "An experimental demonstration of blind ocean acoustic tomography," *J. Acoust. Soc. Am.* **3**, 1420–1431 (2006).
²³J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE* **57**, 1408–1418 (1969).
²⁴R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. dissertation, Stanford University, Stanford, CA, 1982.
²⁵K. Hsu and A. B. Baggeroer, "Application of the maximum-likelihood method (mlm) for sonic velocity logging," *Geophysics* **51**, 780–787 (1986).
²⁶P. C. Mignerey and S. Finette, "Multichannel deconvolution of an acoustic transient in an oceanic waveguide," *J. Acoust. Soc. Am.* **92**, 351–364 (1992).
²⁷S. Finette, P. C. Mignerey, J. F. Smith, and C. D. Richmond, "Broadband source signature extraction using a vertical array," *J. Acoust. Soc. Am.* **94**, 309–318 (1993).
²⁸J. F. Boehme, *Advances in Spectrum Analysis and Array Processing*, (Prentice Hall, Englewood Cliffs, NJ, 1991), Vol. **2**, Chap. 1, pp. 1–63.
²⁹C. M. Ferla, M. B. Porter, and F. B. Jensen, "C-SNAP: Coupled SACLANTCEN normal mode propagation loss model," Memorandum SM-274, SACLANTCEN Undersea Research Center, La Spezia, Italy, 1993.
³⁰G. Schwartz, "Estimating the dimension of a model," *Ann. Stat.* **6**, 461–464 (1978).
³¹J. Rissanen, "Modeling by shortest data description," *Automatica* **14**, 465–471 (1978).
³²M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-33**, 387–392 (1985).
³³S. Jesus, A. Silva, and C. Soares, "Acoustic Oceanographic Buoy test during the MREA'03 sea trial," Internal Rep. 04/03, SiPLAB/CINTAL, Universidade do Algarve, Faro, Portugal, November 2003.
³⁴F. B. Jensen, "Comparison of transmission loss data for different shallow-water areas with theoretical results provided by a three-fluid normal-mode propagation mode," in *Sound Propagation in Shallow Water*, edited by O. F. Hastrup and O. V. Oleson (SACLANT Undersea Research Centre, La Spezia, Italy, 1974), Vol. **II**, pp. 79–92, SACLANTCEN document CP-14.
³⁵T. Fassbender, "Erweiterte genetische algorithmen zur globalen optimierung multi-modaler funktionen (Extended genetic algorithms for global optimization of multi-modal functions)." Diplomarbeit, Ruhr-Universität, Bochum, 1995.
³⁶C. Soares, S. M. Jesus, and E. Coelho, "Acoustic oceanographic buoy testing during the maritime rapid environmental assessment 2003 sea trial," in *Proceedings of the European Conference on Underwater Acoustics 2004*, edited by D. Simons, Delft, The Netherlands, pp. 271–279.

Resonant generation of surface acoustic waves between moving and stationary piezoelectric crystals

Vladimir N. Khudik^{a)} and Constantine E. Theodosiou

Department of Physics and Astronomy, University of Toledo, Toledo, Ohio 43606, USA

(Received 2 March 2007; revised 4 September 2007; accepted 22 September 2007)

The propagation of surface acoustic waves in a system composed of two piezoelectric crystals moving with respect to each other and separated by a vacuum gap is considered. The waves are localized on different sides of the gap and coupled only through the electrostatic interaction. It is shown that when the velocity of the relative motion of crystals is close to some value, there occurs a wave instability resulting in a resonant generation of these surface waves. The rate of growth of Bleustein–Gulyaev waves in piezoelectric crystals of $6mm$ symmetry class is determined analytically. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2799503]

PACS number(s): 43.35.Pt [PEB]

Pages: 3405–3408

I. INTRODUCTION

Surface acoustic waves^{1,2} can only damp when they move along the surface of the piezoelectric crystal which is free of the action of external mechanical stresses and electric fields. The situation is different when these waves propagate along surfaces of moving and stationary crystals. The composed system is far from thermodynamic equilibrium, and the waves can become unstable. When the velocity of the moving crystal is close to some value, the wave amplitudes grow exponentially even though the crystals are separated by a vacuum gap and surface acoustic waves can interact with each other only through the electric field. In this unstable system, the energy of growing oscillations is drawn from the kinetic energy of the relative motion of crystals.

The mechanism of this instability is of a general nature and can be qualitatively described as follows. Let us consider for simplicity the system depicted in Fig. 1: The first stationary crystal is in the region $x_2 < 0$, and the second crystal (moving with the velocity V along the x_1 axis) is in the region $x_2 > d$, where d is the length of the gap between crystals. The gap length is assumed to be large so that the surface acoustic waves on different sides of the gap only weakly interact with each other. Let one wave propagate along the x_1 axis with the phase velocity V_{s1} with respect to the stationary crystal and the other wave propagate in the opposite direction with the phase velocity V_{s2} with respect to the moving crystal, see Fig. 1. When the velocity V of the relative motion of the crystals is equal to the sum $V_{s1} + |V_{s2}|$, the second wave moves with respect to the stationary crystal with the same velocity as the first wave and both waves remain in the same phase all the time. The electric field in the first wave acts as a resonant force on oscillations of the second wave and vice versa. Such an interaction of the waves moving in opposite directions results in the exponential growth of both waves.

The general dispersion equation of surface acoustic waves is quite complicated; it becomes somewhat simpler in

the case of Bleustein–Gulyaev (BG) waves^{3–7} in which the mechanical displacement is normal to the sagittal plane. In this paper, we consider the wave interaction and find the rate of the wave growth for BG waves propagating in piezoelectric crystals of $6mm$ symmetry class. We assume that both crystals are from the same material and their axes of symmetry aligned with the x_3 axis (which is normal to the sagittal plane x_1 – x_2 shown in Fig. 1). Since all velocities (V , V_{s1} and V_{s2}) are very small compared to the light velocity, we use the electrostatic approximation for describing BG waves in both stationary and moving crystals.

II. DISPERSION EQUATION FOR COUPLED SURFACE ACOUSTIC WAVES

In BG surface acoustic waves, the components of the mechanical displacement in the sagittal plane are equal to zero so that equations of motion^{2,8} reduce to one equation for the u_3 component. In the stationary crystal (i.e., in the region $x_2 < 0$), we have

$$\rho \frac{\partial^2 u_3}{\partial t^2} = \frac{\partial}{\partial x_1} \sigma_{31} + \frac{\partial}{\partial x_2} \sigma_{32}, \quad (1)$$

where ρ is the material mass density,

$$\sigma_{3k} = c_{44}^E \frac{\partial u_3}{\partial x_k} + e_{15} \frac{\partial \varphi}{\partial x_k} \quad (k = 1, 2) \quad (2)$$

are the components of the stress tensor, c_{44}^E is the elastic coefficient, and e_{15} is the piezoelectric coefficient. The electric potential φ satisfies the electrostatic equation

$$\frac{\partial D_1}{\partial x_1} + \frac{\partial D_2}{\partial x_2} = 0, \quad (3)$$

where

$$D_k = e_{15} \frac{\partial u_3}{\partial x_k} - \varepsilon_{11}^S \frac{\partial \varphi}{\partial x_k} \quad (k = 1, 2) \quad (4)$$

are the components of the electric induction, and ε_{11}^S is the dielectric constant. The surface wave is localized near the surface of the crystal, so that $u_3 \rightarrow 0$ when $x_2 \rightarrow -\infty$. The component σ_{32} of the stress tensor must vanish at $x_2 = 0$, and

^{a)}Author to whom correspondence should be addressed. Electronic mail: vkhudik@physics.utoledo.edu

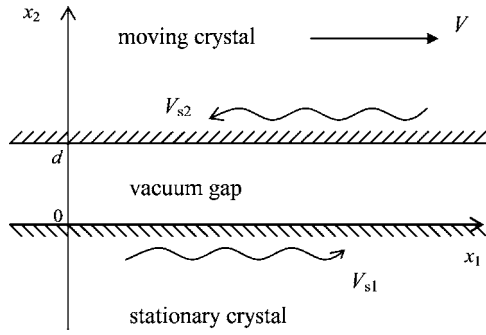


FIG. 1. The system composed of the moving and stationary crystals. Surface acoustic waves on different sides of the gap propagate in opposite directions.

the electric potential φ and the D_2 component of the electric induction must be continuous across the plane $x_2=0$.

In the moving crystal (i.e., in the region $x_2 > d$), Eq. (1) for the u_3 component of the displacement should be modified as

$$\rho \left(\frac{\partial}{\partial t} + V \frac{\partial}{\partial x_1} \right)^2 u_3 = \frac{\partial}{\partial x_1} \sigma_{31} + \frac{\partial}{\partial x_2} \sigma_{32}, \quad (5)$$

and Eq. (3) for the electric potential remains the same. In this crystal, $u_3 \rightarrow 0$ when $x_2 \rightarrow +\infty$, $\sigma_{32}=0$ at $x_2=d$ and φ and D_2 are continuous across the plane $x_2=d$.

We assume that u_3 and φ are proportional to $\exp(-i\omega t + ikx_1)$, where ω and k are the wave frequency and the wave vector. Then, the solution of Eqs. (1)–(4) localized near the surface of the stationary crystal is

$$u_3 = u_{3s} \exp(-i\omega t + ikx_1 + \kappa x_2),$$

$$\varphi = ru_3 + (\varphi_s - ru_{3s}) \exp(-i\omega t + ikx_1 + \kappa x_2), \quad (6)$$

where u_{3s} and φ_s are the mechanical displacement and the electric potential at $x_2=0$, $r \equiv e_{15}/\varepsilon_{11}^S$, and $\kappa = (k^2 - \omega^2/c_T^2)^{1/2}$. In the last formula, c_T is the velocity of the transverse bulk wave

$$c_T^2 = \bar{c}_{44}/\rho, \quad \bar{c}_{44} \equiv c_{44}^E + e_{15}^2/\varepsilon_{11}^S. \quad (7)$$

The solution of Eqs. (3) and (5) localized near the surface of the moving crystal is

$$u_3 = u_{3m} \exp(-i\omega t + ikx_1 - \kappa' x_2),$$

$$\varphi = ru_3 + (\varphi_m - ru_{3m}) \exp(-i\omega t + ikx_1 - \kappa' x_2), \quad (8)$$

where u_{3m} and φ_m are the displacement and the potential at $x_2=d$, and $\kappa' = [k^2 - (\omega - kV)^2/c_T^2]^{1/2}$.

Substituting Eq. (6) in the condition $\sigma_{32}=0$ at $x_2=0$, one can express u_{3s} in terms of φ_s

$$u_{3s} = - \frac{\varepsilon_{11}^S e_{15}}{\varepsilon_{11}^S c_{44}^E \kappa + e_{15}^2 (\kappa - k)} k \varphi_s. \quad (9)$$

In view of Eq. (9), the D_2 component of the electric induction at $x_2=0$ can be expressed in terms of the potential as

$$D_2 = -k\varphi_s \varepsilon_0 \bar{\varepsilon}_*(\omega/k), \quad (10)$$

where

$$\bar{\varepsilon}_*(v) = (\varepsilon_{11}^S/\varepsilon_0) [1 - K^2(1 - v^2/c_T^2)^{-1/2}] \quad (11)$$

is the effective permittivity function for the stationary crystal, $v = \omega/k$ is the phase velocity of the wave, ε_0 is the permittivity of free space, and K^2 is the piezoelectric coupling constant, $K^2 = e_{15}^2/(\varepsilon_{11}^S c_{44}^E + e_{15}^2)$. In the same way, the electric induction at $x_2=d$ in the moving crystal can be written in the form

$$D_2 = k\varphi_m \varepsilon_0 \bar{\varepsilon}_*(\omega/k - V), \quad (12)$$

where $\bar{\varepsilon}_*$ is the effective permittivity function for the moving crystal. In the case under consideration (in which both crystals are of the same material and have the same orientation of axes of symmetry), $\bar{\varepsilon}_*(\omega/k - V) = \varepsilon_*(\omega/k - V)$.

In the gap between crystals (i.e., in the region $0 < x_2 < d$), the potential φ satisfies the Laplace equation

$$\Delta \varphi = 0, \quad (13)$$

whose solution is

$$\varphi = [\varphi_1 \exp(-kx_2) + \varphi_2 \exp(kx_2 - kd)] \exp(-i\omega t + ikx_1), \quad (14)$$

where φ_1 and φ_2 are constants of integration. In view of Eqs. (10), (12), and (14), the continuity conditions for D_2 and φ can be written as

$$-k\varphi_s \varepsilon_0 \bar{\varepsilon}_*(\omega/k) = \varepsilon_0 (k\varphi_1 - k\varphi_2 e^{-kd}), \quad x_2 = 0,$$

$$\varphi_s = \varphi_1 + \varphi_2 e^{-kd}, \quad x_2 = 0, \quad (15)$$

$$k\varphi_m \varepsilon_0 \bar{\varepsilon}_*(\omega/k - V) = \varepsilon_0 (k\varphi_1 e^{-kd} - k\varphi_2), \quad x_2 = d,$$

$$\varphi_m = \varphi_1 e^{-kd} + \varphi_2, \quad x_2 = d. \quad (16)$$

Finally, from Eqs. (15) and (16) one obtains the dispersion equation for the surface wave which propagates near the gap between moving and stationary crystals

$$\frac{[\varepsilon_*(\omega/k) + 1][\bar{\varepsilon}_*(\omega/k - V) + 1]}{[\varepsilon_*(\omega/k) - 1][\bar{\varepsilon}_*(\omega/k - V) - 1]} = e^{-2kd}. \quad (17)$$

When this equation has a solution with $\text{Im } \omega > 0$, the wave is unstable and its amplitude grows exponentially with time.

III. INTERACTION OF SURFACE ACOUSTIC WAVES

In the short wave limit, $kd \rightarrow \infty$ ($e^{-2kd} \rightarrow 0$), Eq. (17) reduces to the dispersion equations for two independent waves

$$\varepsilon_*(\omega/k) = -1,$$

$$\bar{\varepsilon}_*(\omega/k - V) = -1, \quad (18)$$

which are localized near the surfaces of the stationary and moving crystals. In general case, these waves propagate with phase velocities

$$\omega/k = V_{s1}, \quad V_{s1} = \pm V_{BG},$$

$$\omega/k - V = V_{s2}, \quad V_{s2} = \pm V_{BG}, \quad (19)$$

where V_{BG} is the velocity of BG surface acoustic wave

$$V_{BG} = c_T \sqrt{1 - K^4 / (1 + \varepsilon_{11}^S / \varepsilon_0)^2}. \quad (20)$$

In the case when $kd \gg 1$, the electric potential is exponentially small in the middle of the gap [$\varphi \sim \exp(-kd/2)$ at $x_1 = d/2$, see Eq. (14)] so that the wave propagating in the crystals is composed of two weakly coupled surface waves. This allows us to simplify the dispersion Eq. (17). Assuming that conditions (19) are almost satisfied, i.e., assuming that $\omega/k = V_{s1} + \delta\omega/k$ and $V = V_{s1} - V_{s2} + \delta V$ ($\delta\omega \ll \omega$, $k\delta V \ll \omega$), we can present the effective permittivity functions as

$$\varepsilon_*(\omega/k) \approx -1 + \varepsilon'_*(V_{s1})\delta\omega/k,$$

$$\bar{\varepsilon}_*(\omega/k - V) \approx -1 + \bar{\varepsilon}'_*(V_{s2})(\delta\omega/k - \delta V),$$

where $\varepsilon'_*(v) = d\varepsilon_*(v)/dv$ and $\bar{\varepsilon}'_*(v) = d\bar{\varepsilon}_*(v)/dv$. After that, Eq. (17) reduces to the quadratic equation

$$\frac{\varepsilon'_*(V_{s1})(\delta\omega/k)[\bar{\varepsilon}'_*(V_{s2})(\delta\omega/k - \delta V)]}{(-2)(-2)} \approx e^{-2kd}, \quad (21)$$

whose roots are given by the formula

$$\delta\omega = \frac{1}{2}k\delta V \pm k \sqrt{\frac{1}{4}\delta V^2 + \frac{4}{\varepsilon'_*(V_{s1})\bar{\varepsilon}'_*(V_{s2})}e^{-2kd}}. \quad (22)$$

Note that $\varepsilon_*(v)$ is an even function of v . Therefore when waves move in opposite directions, $V_{s1} = +V_{BG}$ and $V_{s2} = -V_{BG}$, then $\varepsilon'_*(V_{s1}) = -\bar{\varepsilon}'_*(V_{s2}) = \varepsilon'_*(V_{BG})$, where $\varepsilon'_*(V_{BG}) = (1 + \varepsilon_{11}^S / \varepsilon_0)^3 V_{BG} / [(\varepsilon_{11}^S / \varepsilon_0) K^4 c_T^2]$. The expression under the square root sign on the right-hand side of Eq. (22) is negative and the waves exponentially grow when $|\delta V| < 4e^{-kd} / \varepsilon'_*(V_{BG})$. At the resonant velocity $V = 2V_{BG}$, the growth rate is

$$\gamma = \frac{2k}{\varepsilon'_*(V_{BG})}e^{-kd} = \frac{c_T}{d}(kd)e^{-kd} \frac{2(\varepsilon_{11}^S / \varepsilon_0)c_T K^4}{(1 + \varepsilon_{11}^S / \varepsilon_0)^3 V_{BG}}. \quad (23)$$

Near the resonance, the weakly coupled waves on different sides of the gap have equal amplitudes but different phases. Using Eqs. (15) and (16), one can find the ratio φ_m to φ_s

$$\frac{\varphi_m}{\varphi_s} = -\frac{\varepsilon_*(\omega/k) + 1}{\varepsilon_*(\omega/k - V) - 1}e^{kd} \approx \frac{\delta\omega}{\gamma} = e^{i\theta}, \quad (24)$$

where $\sin \theta = [1 - k^2 \delta V^2 / (2\gamma)^2]^{1/2}$. When $\delta V = 0$, the shift θ in wave phases is exactly equal to $\pi/2$.

Note that when both waves move in the same direction, $V_{s1} = +V_{BG}$ and $V_{s2} = +V_{BG}$ (or $V_{s1} = -V_{BG}$ and $V_{s2} = -V_{BG}$), then $\varepsilon'_*(V_{s1}) = \bar{\varepsilon}'_*(V_{s2})$. All terms under the square root sign in the right-hand side of Eq. (22) are positive, the wave amplitudes do not change in time, and the shift θ in wave phases is equal to zero. So, the propagation of the waves in the opposite directions is a key condition for the resonant wave generation.

In the case when $kd \sim 1$, the waves on different sides of the gap are strongly coupled and the dispersion Eq. (17) can be solved numerically, see Fig. 2. As one can see from this figure, there is also a branch of oscillations, which corresponds to unstable waves with the nonresonant interaction (for these waves, $\text{Re } \omega/k \approx V - V_{BG}$ when $V > 2V_{BG}$, and the difference between ω/k and V_{BG} increases with an increase of V).

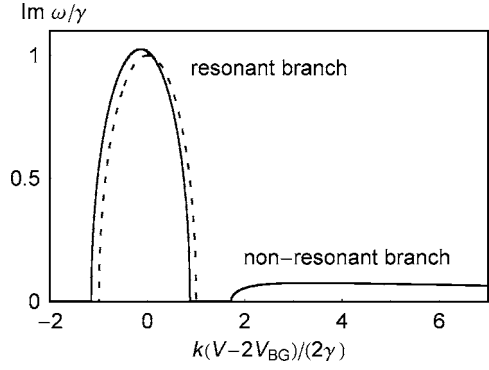


FIG. 2. The renormalized growth rate $\text{Im } \omega/\gamma$ [with γ given by Eq. (23)] as a function of the renormalized deviation of the velocity of the relative motion of crystals V from the resonant value $2V_{BG}$ for strongly coupled waves: $kd = 2.5$, $\varepsilon_{11}^S / \varepsilon_0 = 10$, and $K^2 < 1$ (solid curve). The dashed curve is the renormalized growth rate for weakly coupled waves ($kd \gg 1$) given by Eq. (22). In the resonant branch, the growth rate for strongly coupled waves slightly differs from that for weakly coupled waves. In the nonresonant branch, it slowly decreases with an increase of the velocity of the relative motion of crystals.

Our numerical analysis of Eq. (17) also shows that the growth rate (as a function of k and V) reaches its maximum value

$$\gamma_{\max} \approx \frac{0.16}{(\varepsilon_{11}^S / \varepsilon_0)} K^4 \frac{c_T}{d} \quad (25)$$

at $kd \approx 2 / (1 + \varepsilon_{11}^S / \varepsilon_0)$ and $V \approx (2 - 0.1K^4)c_T$. (The accuracy of these formulas is about several percent for $K^2 < 0.5$.)

We have considered the system in which the stationary crystal is in the region $x_2 < 0$, and the moving one is in the region $x_2 > d$. The similar results can be obtained for BG waves when the system is composed of the moving and stationary piezoelectric plates of the thickness δ , which are in regions $-\delta < x_2 < 0$ and $d < x_2 < d + \delta$ and whose axes of symmetry are normal to the sagittal plane. (Since the general expression for the effective permittivity function is lengthy, it is not presented here.) When the plate thickness is small ($\delta \ll k^{-1}$), there is no difference between bulk and surface waves. In this case the effective permittivity function is given by

$$\varepsilon_*(v, k) \approx 1 + k\delta \left[\frac{\varepsilon_{11}^S}{\varepsilon_0} - \frac{\varepsilon_0 c_{44}^E}{\varepsilon_{11}^S c_{44}} + \frac{e_{15}^2}{\varepsilon_0 (c_{44}^E - \rho v^2)} \right]. \quad (26)$$

By solving equation $\varepsilon_*(v, k) = -1$, one finds the phase velocity of independent BG waves propagating along thin piezoelectric plates, $V_s^2 \approx c_{44}^E / \rho = c_T^2 (1 - K^2)$. For weakly coupled waves ($kd \gg 1$), the growth rate at the resonant velocity $V = 2V_s$ is

$$\gamma \approx \frac{2k}{\varepsilon'_*(V_s, k)}e^{-kd} \approx \frac{kd}{4} \frac{c_T^2}{V_s d} (k\delta)e^{-kd} (\varepsilon_{11}^S / \varepsilon_0) K^2, \quad (27)$$

where $\varepsilon'_*(V_s, k) \equiv \partial \varepsilon_*(v, k) / \partial v|_{v=V_s}$.

IV. GENERALIZATION AND CONCLUDING REMARKS

Let us emphasize one more time that the resonant generation of surface acoustic waves is of quite general nature: it also occurs when crystals have different thickness or are

from different materials, or when surface acoustic waves on different sides of the gap are of different polarization. The only necessary condition is that both these waves should create the electric field in the gap through which they interact with each other.

In the general case, the rate of the wave growth can be found as follows. First, one can determine the effective permittivity function for the stationary crystal, $\varepsilon_*(\omega/k, k) = -D_n/(k\varphi_s)$, where D_n is the component of the electric induction normal to the surface of the crystal and φ_s is the electric potential on this surface. Although the effective permittivity can depend also on the wave number k , it is always an even function of the phase velocity $v = \omega/k$ (because the frequency ω enters the basic equations for piezoelectric crystals only through the inertia term $\rho \partial^2 \mathbf{u} / \partial t^2 = -\omega^2 \rho \mathbf{u}$). Then, the effective permittivity function $\bar{\varepsilon}_*(\omega/k - V, k)$ is determined for the moving crystal. Similarly, it is always an even function of $\omega/k - V$. (When crystals and wave polarizations are the same, then $\bar{\varepsilon}_*(\omega/k - V, k) = \varepsilon_*(\omega/k - V, k)$.) After that, one approaches dispersion Eq. (17). For short waves, $kd \gg 1$, the wave frequency is given by general formula (22) when the velocity of the relative motion of the crystals is close to the resonant value. Since derivatives $\varepsilon'_*(V_{s1}, k)$ and $\bar{\varepsilon}'_*(V_{s2}, k)$ are odd functions of phase velocities of the surface acoustic waves, one can choose the direction of V_{s1} and V_{s2} in such a way that $\varepsilon'_*(V_{s1}, k)\bar{\varepsilon}'_*(V_{s2}, k) < 0$. Surface acoustic waves propagating with these velocities are unstable and their growth rate at the resonant velocity of the relative motion of crystals $V = V_{s1} - V_{s2}$ can be obtained from the formula $\gamma = 2k \exp(-kd) / |\varepsilon'_*(V_{s1}, k)\bar{\varepsilon}'_*(V_{s2}, k)|^{1/2}$. The growth rate reaches (as a function of k and V) its maximum at $k \sim d^{-1}$ (when waves are strongly coupled). The maximum value can be found numerically from the dispersion Eq. (17).

Note that in the general case, functions $\varepsilon_*(\omega/k, k)$ and $\bar{\varepsilon}_*(\omega/k - V, k)$ can be quite complicated and, apparently, the dispersion Eq. (17) may also have solutions with $\text{Im } \omega > 0$ which correspond to unstable waves with the nonresonant interaction (similar to those considered in the previous section).

Phase velocities of surface waves (found from equations $\varepsilon_*(V_{s1}, k) = -1$ and $\bar{\varepsilon}_*(V_{s2}, k) = -1$) in the general case depend on the wave vector k . It is interesting to note that since the phase velocity of bending waves in thin piezoelectric plates can be considerably less than the characteristic velocity of the sound in the piezoelectric material, $V_s \sim c_T k(k\delta) \ll c_T$ where δ is the plate thickness, these waves can be generated at comparatively small velocities of the relative motion of the plates, $V \approx 2V_s \ll c_T$.

Note that for experimental observation of the described physical effect, one can create the relative motion of the crystals in a rotor-stator system where both the rotor and stator are coated with piezoelectric layers.

It is remarkable that the wave generation occurs in the system where the piezoelectric crystals do not have a mechanical contact and the surface acoustic waves propagating in the stationary and moving crystals interact with each other only through the electric field.

ACKNOWLEDGMENT

The authors are grateful to Yu. A. Tsidulko for fruitful discussions.

- ¹L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskii, *Electrodynamics of Continuous Media* (Butterworth-Heinemann, Oxford, 1984), pp. 67–74.
- ²G. A. Maugin, *Continuum Mechanics of Electromagnetic Solids* (North-Holland, Amsterdam, 1988).
- ³J. L. Bleustein, "A new surface wave in piezoelectric materials," *Appl. Phys. Lett.* **13**, 412–413 (1968).
- ⁴Yu. V. Gulyaev, "Electroacoustic surface waves in solids," *JETP Lett.* **9**, 37–38 (1969).
- ⁵G. Koerber and R. F. Vogel, "Generalized Bleustein modes," *IEEE Trans. Sonics Ultrason.* **19**, 3–8 (1972).
- ⁶V. M. Bright and W. D. Hunt, "Light diffraction by Bleustein-Gulyaev surface acoustic waves in piezoelectric cubic crystals," *J. Appl. Phys.* **68**, 1985–1992 (1990).
- ⁷F. S. Hickernell, "Shear horizontal BG surface acoustic waves on piezoelectrics: A historical note," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **52**, 809–811 (2005).
- ⁸R. R. Ramos and J. A. Otero, "Wave propagation in a piezoelectric layer," *J. Appl. Phys.* **81**, 7242–7247 (1997).

Entrainment and stimulated emission of ultrasonic piezoelectric auto-oscillators

Richard L. Weaver^{a)} and Oleg I. Lobkis

¹*Department of Physics, University of Illinois, 1110 W Green Street, Urbana, Illinois 61801, USA*

Alexey Yamilov

²*Department of Physics, University of Missouri-Rolla, 1870 Miner Circle, Rolla, Missouri 65409, USA*

(Received 24 February 2007; revised 27 September 2007; accepted 28 September 2007)

Theoretical modeling and laboratory tests are conducted for nonlinear auto-oscillating piezoelectric ultrasonic devices coupled to reverberant elastic bodies. The devices are shown to exhibit behavior familiar from the theory of coupled auto-oscillators. In particular, these spontaneously emitting devices adjust their limit-cycle frequency to the spectrum of the body. It is further shown that the auto-oscillations can be entrained by an applied field; an incident wave at a frequency close to the frequency of the natural limit cycle entrains the oscillator. Special attention is paid to the phase of entrainment. Depending on details, the phase is such that the oscillator can be in a state of stimulated emission: the incident field amplifies the ultrasonic power emitted by the oscillator. These behaviors are essential to eventual design of an ultrasonic system that would consist of a number of such devices all synchronized to their mutual field, a system that would be an analog to a laser. A prototype *uaser* is constructed. © 2007 Acoustical Society of America.. [DOI: 10.1121/1.2800315]

PACS number(s): 43.35.Yb, 43.58.Wc, 43.38.Ew [TDM]

Pages: 3409–3418

I. INTRODUCTION

Auto-oscillators are common in acoustics. Simple autonomous nonlinear mechanical systems, with steady forcing, can have stable limit cycles. Examples are the bowing of a violin string¹ vortex shedding by a steady flow over an obstacle² or through a clarinet reed. Otoacoustic emissions³ are of this class, as are thermo-acoustic engines.⁴ Unstable feedback loops^{5–7} occur in room acoustics, hearing aids⁸ and vibration control.⁹ The intriguing case in which two or more mechanical limit cycle oscillators are coupled through a linear acoustic medium has also received attention.^{10–12} Here we study the case of an ultrasonic transducer as an element in a nonlinear auto-oscillating electronic circuit, its behavior with and without an acoustic medium, and its behavior when an incident acoustic field is applied. As might be expected, it is found that an incident acoustic field can entrain such an oscillator to its frequency. More significantly, it is further found that entrainment can occur with a phase that corresponds to an increased rate of ultrasonic energy emission from the transducer, i.e., to stimulated emission. In the absence of any applied field, a set of two or three such transducers is demonstrated to synchronize itself to the complex mutual radiation field, and to do so with phases that correspond to stimulated emission. Such a system may be thought of as an analog to a laser.

Lasing is not necessarily a quantum phenomenon. Classical laser designs^{13–19} have been proposed but little has yet been realized in the laboratory. Here we report construction of nonlinear electronic oscillators coupled piezoelectrically to an elastic body and capable of both spontaneous and

stimulated emission. Such systems may have both pedagogic and research value. Because acoustic waves with their longer wavelengths and longer time scales permit probes and controls to a degree not possible in optics, these systems may permit experiments that complement those possible with lasers.

The classical laser designs of Borenstein and Lamb^{13,14} and Kobelev *et al.*¹⁶ are composed of incoherently excited Duffing oscillators. The oscillators emit spontaneously in a trivial fashion. Theoretical arguments indicate that, when they go into resonance with each other with the right phase relation as enforced by nonlinear interactions with their mutual radiation field, they also emit by stimulated emission. An acoustic version has been reported by Bredikhin *et al.*¹⁷ who present theory and measurements on a system of impulsively excited nonlinear bubbles. While ringing down, the bubbles sometimes synchronize by means of their nonlinearity and their mutual radiation field. The system differs from the present design based on limit cycle oscillators.

Zavtrak and co-workers^{18,19} have suggested that bubbles or other particles in a fluid could be pumped by an applied coherent harmonic electric or acoustic field. (This differs from the usual notion of a laser as converting incoherent pump excitation to coherent radiation.) They suggest that the particles would bunch spatially under the influence of their radiation forces, leading to a coherent re-emission of sound in a direction imposed by their radiation field and the modes of their cavity. This acoustic version of a free-electron laser or gyrotron²⁰ has not been constructed.

Independently, recent years have seen considerable interest in dynamic synchronization^{10,21–28} in which sets of distinct coupled auto-oscillating limit cycles synchronize to each other. The phenomenon occurs in disparate circumstances, including firefly flashes, brain waves, esophageal

^{a)}Author to whom correspondence should be addressed. Electronic mail: r-weaver@uiuc.edu

waves, bridges with crowds of pedestrians, and chemical oscillations. It occurs between lasers, thermo-acoustic engines, Josephson junctions, metronomes and pendulum clocks. The chief mathematical model for studying the synchronization of large numbers of auto-oscillators is that of Kuramoto, and its generalizations.^{23–27} After arguing that the state of a limit cycle oscillator is well represented in terms of its phase ξ , and that the phase of each of the oscillators is weakly coupled to the phases of the others, Kuramoto²⁹ derived a set of N coupled Adler equations.

$$d\xi_n/dt = \omega_n + (1/N) \sum_{m=1}^N A \sin(\xi_m - \xi_n).$$

Kuramoto showed that this model, in the thermodynamic limit $N \rightarrow \infty$, exhibits a phase transition in which a macroscopic number of oscillators synchronize to each other if A is sufficiently large compared to the deviations among the many ω_n . Sundry generalizations have been discussed also, such as time delays and randomness in the couplings. Resemblance between this phase transition and the onset of coherence among the atoms of a laser has been noted.²² In the event that coupling is mediated by a complex wave-bearing body, as in the experiments to be reported here and in large arrays of THz Josephson junctions^{22,30} or the random laser,³¹ the coupling is neither uniform as in the Kuramoto model above, nor completely random as in the model of Daido.²⁵ Thus such systems constitute a so-far unexplored model.

Synchronization of a set of oscillators is not sufficient for emulation of a laser. The sine-qua-non of a laser is *stimulated emission*, in which a wave incident upon an excited oscillator is re-emitted with unchanged phase and increased amplitude. This is a classical phenomenon,^{13–19} and is describable without quantum mechanics. A classical linear oscillator will exhibit stimulated absorption or stimulated emission depending on the phase of the oscillation relative to that of the incident field. Thus an incoherent array of excited linear oscillators will show no net stimulated emission. If, however, all or most of the oscillators can be induced to have the same frequency and the correct phases, the set will exhibit stimulated emission. Only to the extent that oscillators and incident fields have the same frequency and the correct phase difference will the energy in the oscillators be transferred efficiently to the wave field. As described above, dynamic synchronization^{22–30} suggests that nonlinear auto-oscillators can be entrained to their mutual radiation field; they can synchronize to a common frequency with fixed phase differences. If that phase difference is such that each oscillator does work at a rate greater than it does without an incident field, i.e., if there is stimulated emission, the system will be a laser analog.²²

Here we present the elements of one such acoustic system. A piezoelectric auto-oscillator design is described in the next section. A mathematical model is derived that shows potential for entrainment, synchronization and stimulated emission. Section III then presents a laboratory version of the model and makes quantitative observations of entrainment and stimulated emission under the influence of a pre-

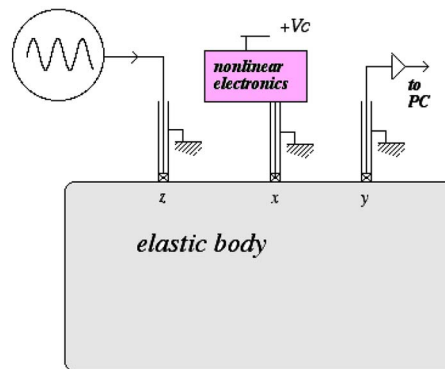


FIG. 1. (Color online) A reverberant elastodynamic body, large compared to a wavelength, is driven at position x by a piezoelectric transducer that is part of a circuit with an attracting limit cycle. The acoustic state of the body is monitored by a separate receiver at y . An optional third transducer at z is driven by an applied harmonic signal at prescribed frequency and amplitude.

scribed applied acoustic wave. Finally, it is demonstrated (Sec. IV) that two or three piezoelectric auto-oscillators can synchronize autonomously (i.e., without an externally prescribed incident wave) to their mutual ultrasonic field. We observe stimulated emission and super radiance, and term the system a *uaser*.

II. AUTO-OSCILLATION, ENTRAINMENT AND STIMULATED EMISSION

Figure 1 shows a piezoelectric transducer attached at x to an elastic body. Like most ultrasonic transducers it is reciprocal; it both radiates and receives ultrasound. It has an internal electronic impedance that is nominally capacitive, with additional small contributions from the mechanics. This transducer is part of a nonlinear electronic circuit (see Fig. 2) with a limit cycle at a frequency and amplitude that depends on circuit parameters. A second transducer at y monitors the acoustic state of the body. An optional third transducer at z applies a continuous harmonic wave at a prescribed frequency and amplitude. Additional limit-cycle oscillators (not pictured) may also be attached.

The auto-oscillator (also called a limit-cycle oscillator) will be influenced by its acoustic environment. Passive mechanical impedance felt by the piezoelectric element in con-

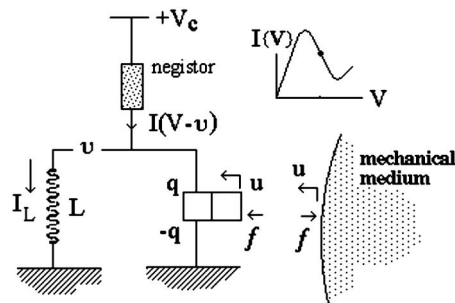


FIG. 2. Diagram for a piezoelectric auto-oscillator. The current I through the negistor depends on the potential across it in the pictured nonlinear fashion $I(V_c - v)$ characteristic of the device. The static supply voltage V_c is adjusted so that, at static equilibrium, $v=0$ and the $I(V)$ relation is at its inflection point. At this point current varies with v like $\epsilon C(1 - \beta v^2)v$, with $\epsilon > 0$. Further details of the negistor are described in Sec. III.

tact with the acoustic medium will manifest as electrical impedance and thus influence the auto-oscillation. Furthermore, any acoustic waves incident upon the piezoelectric element from elsewhere will generate additional currents in the circuit and so influence the oscillation. Thus we have an acoustic realization of systems studied in the literature,^{22–30} consisting of forced and coupled limit-cycle oscillators. Of particular interest in this realization is that the interactions are by means of a complex wave bearing medium, and thus potentially irregular functions of time or frequency and position. An additional provocative question not found in that literature arises as well: whether the coupling enhances or diminishes the acoustic power flowing from the oscillator and thus whether a set of such oscillators may be a model for a laser.

In this section we propose an analytic model for this system and predict the frequency and rate of power emission at x in the absence of any source at z . We also predict circumstances, with arguments similar to those of Refs. 23–28, under which the oscillator at x will be entrained by an applied field from z . We further describe circumstances under which that entrainment corresponds, not only to locking, but also to stimulated emission. (The terms *synchronization*, *locking* and *entrainment* are defined for the present purposes in the footnote.²³) This section establishes the possibilities then illustrated in Secs. III and IV. Sections III and IV can, however, be read independently.

The acoustic state of the body is monitored by a separate receiver at y . An optional third transducer at z is driven by an applied harmonic signal at prescribed frequency and amplitude.

A. Linear transceiver

A passive linear reciprocal transducer may be represented in acoustics by a linear two-port network (see Ref. 32 for details), with potential drop v and charge q on one port and mechanical force f and mechanical displacement u on the other. These inputs and outputs are related by a matrix.

$$\begin{pmatrix} v \\ f \end{pmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{pmatrix} q \\ u \end{pmatrix}; \quad (1a)$$

$$\begin{pmatrix} q \\ u \end{pmatrix} = \begin{bmatrix} T_{22}/D & -T_{12}/D \\ -T_{21}/D & T_{11}/D \end{bmatrix} \begin{pmatrix} v \\ f \end{pmatrix}, \quad (1b)$$

where $D=T_{11}T_{22}-T_{12}T_{21}$ is $[T]$'s determinant. T_{11} may be interpreted as the inverse of the transducer's electric capacitance when clamped ($u=0$). T_{22}/D may be interpreted as the capacitance C of the device when there is no force on it ($f=0$). T_{22} is the shortcircuited mechanical stiffness of the device. Reciprocity demands $T_{12}=T_{21}$. All coefficients are in general complex functions of frequency; $T^*(\omega)=T(-\omega)$, and may in principle be measured.

The rate at which the force and electric potential are doing work on the transducer is

$$P = (vf) \times \left(\frac{\partial q/\partial t}{\partial u/\partial t} \right) = (qu)[T]^T \times \left(\frac{\partial q/\partial t}{\partial u/\partial t} \right).$$

The symbol \times represents, in the time domain, a simple multiplication; other juxtapositions represent time-domain convolutions. In the frequency domain these juxtapositions are simple multiplications. If q and u are harmonic: $q = Q \exp(i\omega t) + \text{c.c.}$; $u = U \exp(i\omega t) + \text{c.c.}$, we may conclude that the time average power flow into the transducer is

$$P = -i\omega(QU)[T(\omega)^T - T(\omega)^*] \begin{pmatrix} Q^* \\ U^* \end{pmatrix}$$

with $[T(\omega)] = \int \exp(-i\omega t)[T(t)]dt = [T(\omega)]^T$. A passive transducer must have $P \geq 0$. We therefore conclude that $[T(\omega)]$'s imaginary part must be positive (negative) semidefinite at positive (negative) frequency.

B. Linear piezoelectric transducer in a nonlinear electronic circuit

The transducer is inserted into the circuit of Fig. 2 in which the negistor has a nonlinear current voltage relation as pictured. At and near the inflection point in this curve the negistor has what is commonly termed negative differential electric resistance. When the negistor is placed in a standard *RLC* circuit, we expect oscillations with negative damping, oscillations that grow exponentially with time until such amplitude as the negative differential resistance $I(V)$ relation fails. The circuit can then be expected to settle into a limit cycle for which average resistance is zero. If the static bias voltage V_c is set at the inflection point, then fluctuations v of the voltage across the negistor correspond to fluctuating current $I(V_c - v) = I(V_c) + \varepsilon C(1 - \beta v^2)v$. The slope at the inflection, $-\varepsilon C$, is the negative differential conductance. (The factor C has been inserted for later convenience.) Curvature at the inflection point is characterized by β . It governs the amplitude of the limit cycle.

Current conservation demands, where I_L is the current in the inductor,

$$I(V_c - v) = dq/dt + I_L.$$

This has a steady solution at $v=q=0$, $I_L = I(V_c)$. Recalling $dI_L/dt = v/L$, differentiating with respect to time gives

$$-I'(V_c - v)dv/dt = d^2q/dt^2 + v/L. \quad (2)$$

Substituting for d^2q/dt^2 by (1b) gives

$$d^2v/dt^2 - \varepsilon(1 - 3\beta v^2)dv/dt + (LC)^{-1}v = (T_{12}/T_{22})d^2f/dt^2. \quad (3)$$

This equation governs the fluctuations in voltage and could be useful if the force f were prescribed. In particular if $f=0$, corresponding to the transducer being mechanically free, (3) becomes a van der Pol equation.³³

Equation (1b) implies a relation between f , u and v ;

$$f = T_{21}/T_{11}v + D/T_{11}u. \quad (4)$$

So, if u is prescribed there is an alternate form

$$d^2v/dt^2 - \varepsilon(1 - 3\beta v^2)dv/dt + [(LC)^{-1} - \{T_{12}T_{21}/T_{22}T_{11}\}d^2/dt^2]v = (DT_{12}/T_{22}T_{11})d^2u/dt^2. \quad (5)$$

C. Spontaneous emission

Equations (3)–(5) are not complete unless we specify u or f , or a relation between them. A case of primary interest is that of free radiation into an otherwise passive acoustic medium, for which: $u = -Gf$. (The minus sign is required due to the usual definition of elastodynamic Greens functions and to the opposite orientation of f and u in Fig. 2). Here G is the $\hat{n}\hat{n}$ component (\hat{n} being the unit vector in the direction of f or u) of the ultrasonic Greens' dyadic at position x . [The Greens dyadic $G_{ij}(x, y, \omega)$ describes the complex amplitude of the linear elastodynamic displacement response of the solid in direction i at position x due to a unit harmonic force at y in direction j : so the quantity G used here is $n_j n_i G_{ij}(x, x, \omega)$]. Thus, from Eq. (4)

$$[1 + DG/T_{11}]f = T_{21}/T_{11}v \quad (6)$$

and

$$d^2v/dt^2 - \varepsilon(1 - 3\beta v^2)dv/dt + [(LC)^{-1} - \{T_{12}T_{21}/T_{22}(T_{11} + DG)\}d^2/dt^2]v = 0. \quad (7)$$

Solutions of nonlinear dynamical equations such as Eq. (7) are difficult. This is especially so if the complicated time delays and temporal convolutions in T and G are fully expressed. For the present purposes we will presume that G may be evaluated at the frequency of chief interest, that G is only weakly dependent on frequency and can be replaced by a constant. The elements of $[T]$ also vary with frequency but generally in a much smoother fashion, so replacing them by constants is readily justified. Approximating Eq. (7) as a phase oscillator simplifies it further.^{22–28} This requires that the oscillator be assumed to stay on or near its limit cycle, at a frequency Ω , with a slowly varying real amplitude and phase V , and ξ :

$$v = V(t)\exp(i\Omega t + i\xi(t)) + c.c. \quad (8)$$

On assuming $dV/dt \ll \Omega V$; $d\xi/dt \ll \Omega$, and $|\omega - \Omega| \ll \Omega$ we derive (this also requires an assumption of $\varepsilon \ll \Omega$.)

$$[2i\Omega dV/dt - \Omega^2 V + \omega^2 V - 2V\Omega d\xi/dt - i\varepsilon\Omega(1 - 6\beta V^2)V + i\mu V] = 0 \quad (9)$$

whose real and imaginary parts are

$$dV/dt - (1/2)\varepsilon(1 - 6\beta V^2)V + (1/2\Omega)\mu V = 0$$

$$d\xi/dt = (\omega^2 - \Omega^2)/2\Omega \approx (\omega - \Omega), \quad (10)$$

where ω and μ are given by

$$\omega = (\text{Re}[(LC)^{-1} + \{T_{12}T_{21}\Omega^2/T_{22}[T_{11} + DG]\}])^{1/2}$$

$$\mu = \text{Im}[(LC)^{-1} + \{T_{12}T_{21}\Omega^2/T_{22}[T_{11} + DG]\}] \quad (11)$$

A steady solution implies $\Omega = \omega$, and $V = [(\varepsilon\Omega - \mu)/6\varepsilon\Omega\beta]^{1/2}$.

When the device is unattached to the solid body, as enforced mathematically by taking $f=0$, equivalently $G=\infty$, the frequency of the auto oscillation is the real part of $(LC)^{-1/2}$. When $G \neq \infty$, the frequency is changed by the acoustics to a value Ω that is a solution of the implicit equation [11(a) with $\Omega = \omega$.] In a reverberant body G is an irregular function of ω , so there may be multiple solutions.

The power radiated from the device into the mechanical medium is

$$\Pi = -f \times du/dt = f \times dG/dt \quad (12)$$

which is, by Eq. (6),

$$\begin{aligned} \Pi &= [1 + DG/T_{11}]^{-1}T_{21}/T_{11}v \times dG/dt[1 + DG/T_{11}]^{-1} \\ &\quad \times T_{21}/T_{11}v \\ &= -2\Omega \text{Im} G(\Omega)|T_{11} + DG|^{-2}|T_{21}|^2V^2 \\ &= -2\Omega \text{Im} G(\Omega)|T_{11} + DG|^{-2}|T_{21}|^2[(\varepsilon\Omega - \mu)/6\varepsilon\Omega\beta] \end{aligned} \quad (13)$$

[The first equality in Eq. (13) arises from a time averaging.] This is the power of spontaneous emission.

Section III shows measurements on a system like that modeled here, with observations of the frequency of auto oscillation with and without $G=\infty$.

D. Entrainment

An incident wave field u^{inc} can modify the state of the nonlinear oscillator. The nonlinear oscillator augments the incident field with its own radiation $-Gf$, so instead of $u = -Gf$ we have $u = u^{\text{inc}} - Gf$. The equation governing the oscillator is still (3), but now in lieu of (6),

$$f = T_{21}/T_{11}v + D/T_{11}[u^{\text{inc}} - Gf] \quad (14)$$

so

$$\begin{aligned} d^2v/dt^2 - \varepsilon(1 - 3\beta v^2)dv/dt \\ + [(LC)^{-1} - \{T_{12}T_{21}/T_{22}(T_{11} + DG)\}d^2/dt^2]v \\ = (DT_{12}/T_{22}(T_{11} + DG))d^2u^{\text{inc}}/dt^2 \end{aligned} \quad (15)$$

Equation (15) is identical to Eq. (7), except that now it has an external forcing term from the incident field. It is a van-der Pol oscillator with forcing. That such oscillators can be entrained to the frequency of the forcing is well known.^{14,23–28} We take the incident field at x , u^{inc} to be of the form $U \exp(i\Omega t) + c.c.$ with independent parameter Ω and without loss of generality real positive U . The approximations used above now give

$$\begin{aligned} dV/dt - (1/2)\varepsilon(1 - 6\beta V^2)V + (1/2\Omega)\mu V \\ = - (U\Omega/2)\text{Im}\{DT_{12} \exp(-i\xi)/T_{22}(T_{11} + DG)\} \end{aligned} \quad (16)$$

and

$$\begin{aligned} d\xi/dt = (\omega - \Omega) + (U/2V)\Omega \text{Re}\{T_{12} \exp(-i\xi)/C \\ \times (T_{11} + DG)\}. \end{aligned} \quad (17)$$

This is similar to Adler's equation.^{23,27,28}

For sufficiently small detuning,

$$|\omega - \Omega| < \Omega(U/2V)|T_{12}/C(T_{11} + DG)|, \quad (18)$$

the nonlinear differential equation [Eq. (17)] has stationary solutions at two distinct values of ξ . Each solution corresponds to entrainment of the oscillator to the incident field at frequency Ω . Only the solution with $\text{Im}\{T_{12} \exp(-i\xi)/C(T_{11} + DG)\} < 0$ is stable.

E. Power flow and stimulated emission from an entrained oscillator

A general expression for the phase at stable entrainment is complicated, but in the absence of detuning ($\omega = \Omega$), it simplifies

$$\exp(-i\xi) = -i(DT_{12}/T_{22}(T_{11} + DG))^* / |DT_{12}/T_{22}(T_{11} + DG)|. \quad (19)$$

Examination of Eq. (16) shows that the stable solution for ξ corresponds to a positive right hand side, i.e., the stable phase at entrainment is such that the amplitude V of the oscillations around the inflection point is increased relative to its value in the absence of an incident field. Increased V corresponds to increased energy dissipation in the negistor. That conclusion is independent of the parameters $[T]$ of the two-port network, and independent of the degree of detuning. Thus this would appear to be *stimulated absorption*, and behaviors observed in the laboratory (see Sec. III) would appear to be not represented in this model.

That conclusion is incorrect. While the flow of energy out of the nonlinear circuit into the transducer is, apparently, decreased by the incident acoustic field, the flow of energy out of the transducer into the medium differs from that by dissipation in the transducer. Thus the model may yet describe stimulated emission, but only if the transducer is dissipative and only if power dissipation within the transducer is lessened by the presence of the incident field.

Thus we are led to ask about the radiation of acoustic energy into the mechanical medium. The power radiated from the device is $\Pi = -f \times du/dt = f \times \{-du^{\text{inc}}/dt + dG/dt\}$ which is [by Eq. (14)]

$$\begin{aligned} \Pi &= [T_{11} + DG]^{-1}(T_{21}v + Du^{\text{inc}}) \\ &\quad \times dG/dt [T_{11} + DG]^{-1}(T_{21}v + Du^{\text{inc}}) \\ &\quad - [T_{11} + DG]^{-1}(T_{21}v + Du^{\text{inc}}) \times \{du^{\text{inc}}/dt\}. \end{aligned} \quad (20)$$

The terms in v^2 are almost identical to the spontaneous emission rate, differing only in that V is now slightly different. The terms $u^{\text{inc}2}$ are independent of V , and presumably due to passive losses on scattering off the dissipative parts of $[T]$. This presumption is supported by a short calculation that shows, for the case $T = \text{real}$, that the term in $u^{\text{inc}2}$ vanishes. The cross terms, in $V \times U$, resemble stimulated emission and absorption. These terms are

$$\begin{aligned} \Pi_{UV} &= |T_{11} + DG|^{-2} [T_{21}v \times (dG/dt)Du^{\text{inc}} \\ &\quad + Du^{\text{inc}} \times (dG/dt)T_{21}v - (T_{11} + DG)^* T_{21}v \times du^{\text{inc}}/dt] \end{aligned} \quad (21)$$

On time averaging over one cycle, we recover

$$\Pi_{UV} = 2\Omega UV |T_{11} + DG|^{-2} \text{Im}[T_{21}^* \exp(-i\xi)(T_{11} + DG^*)].$$

A substitution of the expression (19) for ξ (valid in the absence of detuning) gives

$$\begin{aligned} \Pi_{UV} &= -2\Omega UV |T_{22}| / |T_{11} + DG| |DT_{12}| \\ &\quad \text{Re}[T_{21}^{*2} (D/T_{22})^* [(T_{11} + DG^*) / (T_{11} + DG)^*]] \end{aligned} \quad (22)$$

If T is real, this is manifestly negative. [We recall that $D/T_{22} = C$ is the free capacitance and its real part is presumably positive.] We therefore conclude, as in the paragraph following Eq. (18), that real T implies that the UV part of the power flow into the acoustic medium is negative, i.e., that an entrained oscillator absorbs energy from the field incident upon it.

If the elements of $[T]$ are complex the conclusion can differ. The expressions (20)–(22) are complicated, but permit some simplifications. For the case in which T is dominated by positive real parts of the elements on its diagonal, D and T_{11} are very nearly real; the ratio $[(T_{11} + DG^*) / (T_{11} + DG)^*]$ is unity; $(D/T_{22}) \sim T_{11}$ is real and positive. We find Π_{UV} is positive if T_{21} is imaginary, *regardless of G*. Thus stimulated emission depends chiefly on the phase of T_{12} . If we suppose instead that T is dominated by the real part of T_{11} , with all other elements being complex and of equal small order, then $D \sim T_{11}T_{22}$ and

$$\begin{aligned} \Pi_{UV} &= -2\Omega UV |T_{22}| / |T_{11} + DG| |DT_{12}| \\ &\quad \text{Re}[T_{21}^{*2} T_{11} (1 + T_{22}G^*) / (1 + T_{22}^*G^*)]. \end{aligned} \quad (23)$$

In this case the sign of Π_{UV} depends on the phase of T_{12} and also, if the magnitude of $T_{22}G$ is of order unity or larger, on the phase of T_{22} .

We conclude that at least for some sets of parameters the model exhibits stimulated emission. For sufficiently weak detuning, and for transducers with lossy electromechanical coupling $[T]$, we expect an incident field to stimulate emission. These conclusions have been corroborated by direct numerical simulations (not presented) of Eq. (15). In particular we observe spontaneous emission and entrainment; we also see the predicted stimulated emission and its quantitative accord with theory.

III. LABORATORY STUDIES

We have investigated these predictions in the laboratory by constructing the “negistor,” or “lambda diode,”³⁴ illustrated in Figs. 2 and 3, and incorporating it into the circuit of Figs. 1 and 2. We find that the circuit auto-oscillates in a periodic nearly harmonic limit cycle whose frequency is tunable by varying inductance L or by adding additional capacitance. (The parameter ε describing the strength of the negative resistance is also tunable, by adding additional resistance.) Here, and in all cases below, the spectral width of the auto-oscillation is finer than our precision of 1–10 Hz. Linewidth may be governed by background noise as is the Schawlow–Townes theory for the linewidth of a laser³⁵ or a Larsen feedback circuit.⁶ In the following subsection, the circuit of Fig. 3 is shown to oscillate in a limit cycle with a frequency ω that depends, as predicted in Eq. (11), on the complex mechanical compliance represented by G . The next

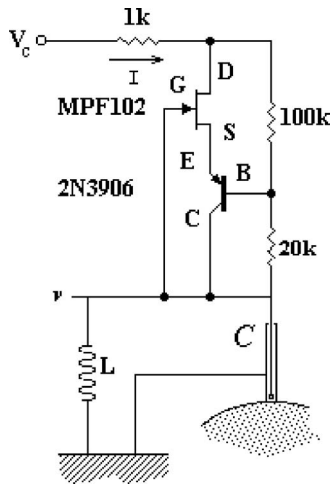


FIG. 3. A lambda diode, or negistor, is constructed from two transistors. It has a current voltage relation as shown in the inset to Fig. 2. When it is incorporated into a circuit with an inductor L ($\sim 220 \mu\text{H}$) and a capacitance C as provided by the piezoelectric transducer and its cables ($\sim 200 \text{ pf}$), the circuit auto-oscillates at a frequency of about $\omega = 1/\sqrt{LC}$.

subsection demonstrates that the circuit will be pulled or entrained by an applied field depending on the degree of detuning and the amplitude of the applied field, as predicted in Eq. (18). Finally it is shown in III C that the applied field can enhance or diminish the acoustic power from the auto-oscillator, as predicted in Sec. II D.

A. Free acoustic oscillation, spontaneous emission

When the piezoelectric transducer with its nonlinear electronics is attached to an elastic body, the frequency of oscillation shifts, as shown in Fig. 4. Frequency changes can be different depending on the position of attachment, the material, the size of the solid body, or the presence of oil couplant. Frequencies always increase when the transducer is placed on the body, reflecting an increase in effective stiff-

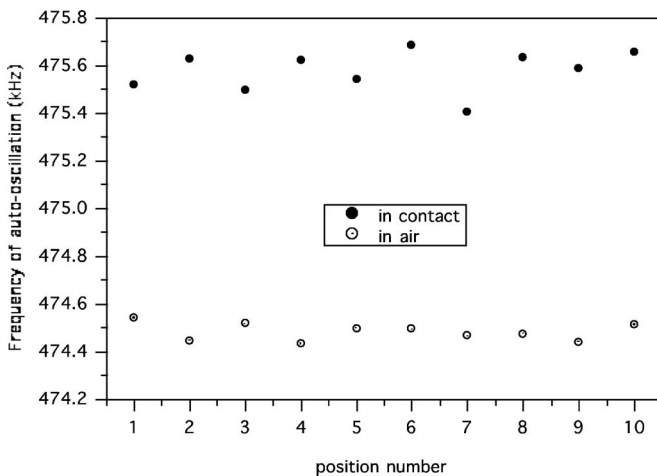


FIG. 4. Frequency of auto oscillation varies by about 1 kHz as the transducer is alternately in contact (filled circles) and out of contact (open circles) with the aluminum block. Slight variations in the frequency of the noncontact case are ascribed to stray capacitances due to the operator's hands. Variations in frequency among the filled circles are ascribed to variations in coupling strength (reattachment of a transducer is notoriously resistant to precise reproducibility) or to variations in local Greens function.

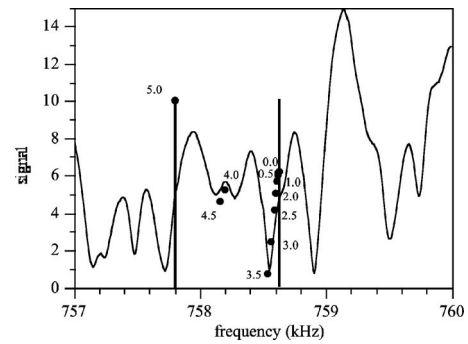


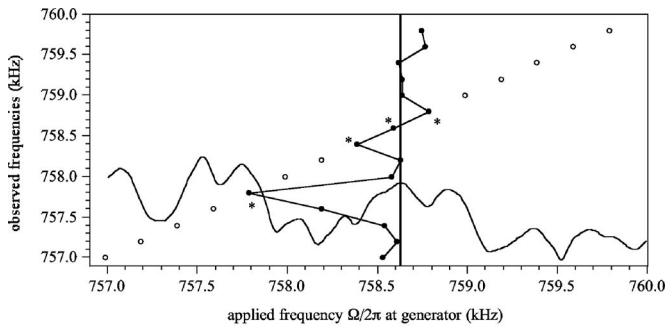
FIG. 5. The nonlinear oscillator's frequency (isolated points) is pulled, monotonically and discontinuously, as the signal strength from the generator at y is increased from 0 to 5 V in 0.5 V steps. The continuous curve is the transfer function $|h(\omega)|$ between monitor and generator. The bold vertical lines mark the positions of the spectral lines (one at 757.79 due to the generator, another at 758.63 due to the uninfluenced auto-oscillator) at vanishing voltage from the generator. The isolated filled circles are labeled with the corresponding values of g in volts; their positions indicate the pulled frequency and the signal strength from the auto-oscillator.

ness. No change in frequency is attendant upon additional grounding of the transducer case, so changes may be ascribed to the mechanics, not the electronics. Changes are small compared to natural frequency of the order of 500 kHz.

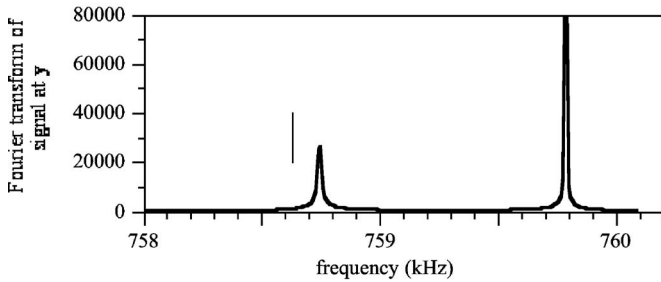
Theoretical frequency is given by Eq. (11). From the observed magnitude of the frequency changes in Fig. 4 as G is alternated between ∞ and the G of the elastic body in different places, and based upon statistically identical data from steel where G is smaller than it is in aluminum, we estimate $\text{Re}(T_{12}T_{21}/T_{22}T_{11})$ to be of the order 0.004, and DG to be less than T_{11} . It is worth noting that in a system with high modal overlap (meaning level spacings are much less than absorption widths) such as the one used in generating Fig. 4, G has fluctuations from place to place or frequency to frequency that are small; $|\delta G|/|G| \ll 1$.³⁶ G is very nearly equal to the value it would take in an infinite half space. For this reason the frequency of auto-oscillation ought to depend only weakly on position x .

B. Entrainment to an applied field

Entrainment is studied by adding a prescribed incident field through the transducer at point z in Fig. 1. Conditions are sought under which the nonlinear electronic circuit adopts the frequency of the applied field. Figure 5 shows (continuous curve) a short section of the spectrum of the ultrasonic transfer function between the generator and the monitor, $|h_{zy}|$ on a 70 mm aluminum cube with enhanced dissipation. (On ignoring scattering by the transducers one can calculate this transfer function in terms of other quantities that have been defined here: $h_{zy} \approx T_{12}G_{zy}T_{21}/(1 + T_{22}G_{zz})(1 + T_{22}G_{yy}) \sim T_{12}^2G_{zy}$.) The bold vertical lines indicate the frequency of the generator at z (757.79 kHz) and that of the undisturbed oscillator in contact with the solid at x (758.63 kHz). The several isolated points indicate the frequency and rms amplitude received at y , due to the oscillator, at each of 11 equally spaced generator amplitudes from 0.0 to 5.0 V. It may be seen that the oscillator frequency is pulled²³ towards that of the generator, and that the degree of



(a)



(b)

FIG. 6. (a) A study of auto-oscillator frequency *versus* the frequency of an entraining field with a fixed harmonic source amplitude $g=5$ V. The natural frequency (i.e., in the absence of an applied field) of the auto-oscillator at x is 759.63 kHz, as indicated by the vertical line. The irregular smooth curve shows the (scaled to fit) transfer function $|h_{xz}(\omega)|$ between the source of the entraining field and the position of the auto-oscillator. Open circles indicate the frequency of the entraining field as specified by the operator. Closed circles indicate the frequency of the auto-oscillator as it is pulled towards that of the entraining field. The structure of the entrainment is complex, and the pulling is nonmonotonic and discontinuous, influenced as it is by the complicated function $h(\omega)$. Entrainment occurs for the points labeled *. (b) An example of the spectra (of the signals received at y) used to construct Fig. 6(a). The strong peak at 759.79 kHz corresponds to a direct signal from the generator at z to the monitor at y . The weaker peak at 758.73 kHz indicates the frequency chosen by the auto-oscillator. It has been pulled, from its original value of 758.62 indicated by the light vertical line segment. Were the auto-oscillator to be entrained we would see only the peak at 759.79. This plot corresponds to the uppermost points in Fig. 6(a).

pulling is monotonic in generator strength. This behavior is familiar in the Adler equation from the literature on entrainment of a single auto-oscillator.²⁸ Less familiar are the occasional discontinuities. The two discontinuities, between 3.5 and 4.0 V, and between 4.5 and 5.0 V where locking ensues, appear to correspond to features in the transfer function h of the solid body. They are related to the random reverberant nature of the wave propagation and are not found in the standard (e.g., Refs. 23 and 24) model with frequency-independent coupling.

Figure 6(a) shows the behavior of entrainment as the frequency of the applied continuous harmonic signal from the generator is varied, at a fixed amplitude of 5.0 V. Frequency is varied in 0.2 kHz steps from 756.99 to 759.79 kHz. The natural frequency of the auto oscillator without applied field is indicated by the vertical line at 758.63 kHz. In the absence of full entrainment the auto-oscillator frequency is pulled towards that of the generator. The auto-oscillator is entrained to the applied frequency if the applied frequency is close to the natural frequency, and if the signal from the generator at z as received at the auto-

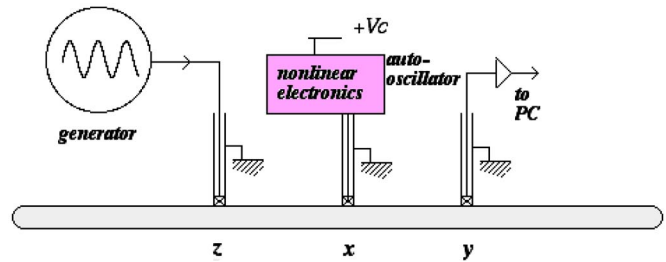


FIG. 7. (Color online) The system is attached to an aluminum rod of 15 cm length and 3.16 mm diameter.

oscillator at x is strong enough, i.e., if the transfer function is large. Entrainment is a complicated nonmonotonic function of applied frequency. Figure 6(b) shows the data used to determine the uppermost points in 6(a), the spectrum at y for the case $\Omega/2\pi=759.79$ kHz at which the oscillator was pulled to $\omega/2\pi=758.73$ kHz.

C. Stimulated emission and absorption

We also study the energy in the system as a function of the amplitude of the applied field, and do so with focus on the case of no de-tuning, i.e., for $\Omega=\omega$. Being unable to measure the power inflow at x , we instead measured the signal strength at y and interpreted it as a measure of the (square root of the) energy in the cavity. This is an imperfect measure of acoustic energy unless the system has low modal overlap and the frequency of interest is near one of the modes, so that only one natural mode is excited. Thus we chose a thin rod with low modal density and modest absorption as pictured in Fig. 7.

We also recognize that the total energy in the rod is proportional to the sum of the power flows from the nonlinear circuit at x and the generator at z . The energy radiated into the rod may be decomposed in two terms, each equal to the time average of the dynamic force at a transducer times the material velocity at the same point. This quantity at x , it was shown above, has a term in V^2 and a term in UV , where U is the field incident from the prescribed forcing at z and proportional to the continuous wave signal input strength “ g ” from the generator. The work done at z , however, includes not only a term in g^2 , but also a term in gV due to the field incident from x . Thus the ultrasonic energy radiated into the body includes not only the intuitive g^2 and V^2 terms, and the gV term described in Sec. II D, but another term scaling like gV that was not discussed there. Analysis shows that the excess stimulated emission of energy at z is of the same order as that at x , but with a phase depending on the phase of G_{zx} . It is not necessarily positive, even if Π_{UV} is positive. Nevertheless, for the case of low modal overlap, the two terms can be shown to have the same sign and the total energy in the cavity can be a proxy for the power flow from the nonlinear circuit.

Figure 8 shows the rms of the continuous harmonic signal observed at the monitor at y for various levels g of the applied input at z . Figure 8(b) shows the difference of the squares of the curves in 8(a), thus corresponding to the energy of spontaneous emission (the value at $g=0$) plus the UV part linear in g that corresponds to stimulated emission.

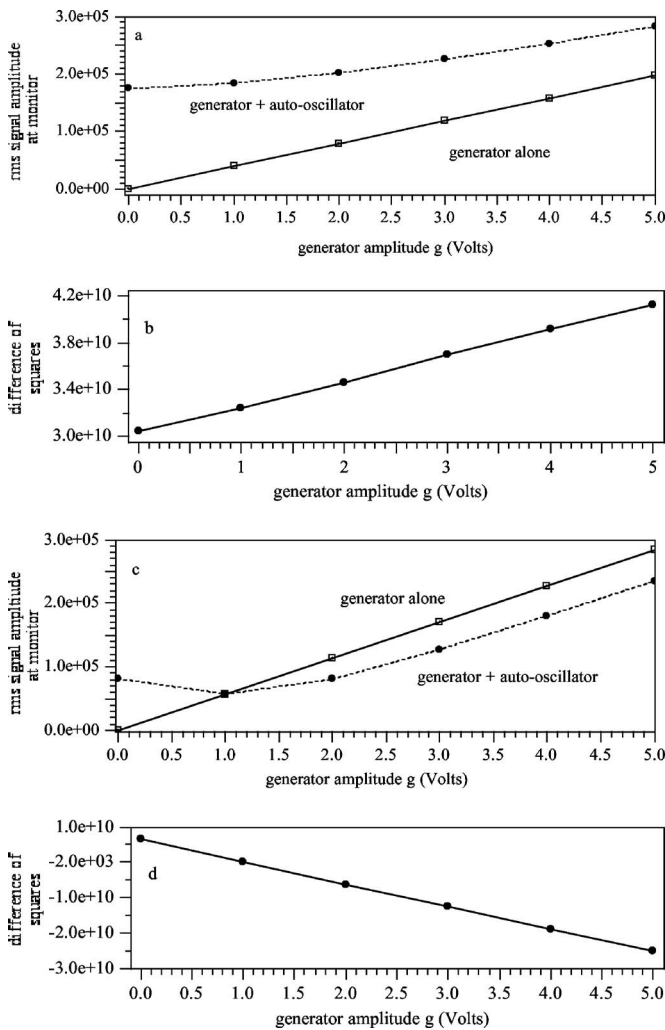


FIG. 8. Evidence of stimulated emission and absorption in the rod of Fig. 7. The rms signal amplitude at y is plotted for two (a, and c) transducer positions x against the amplitude of the signal from the generator which drives the transducer at z . Filled circles correspond to the case in which the auto oscillator is powered; empty squares to unpowered. The difference of their squares is also plotted (b,d), indicating a coherent interference, i.e. stimulated emission (a,b) and absorption (c,d).

Stimulated emission is apparent in the positive slope in Fig. 8(b); power output is greater than the sum of the power outputs from the nonlinear circuit and the generator when they operate alone. Figures 8(c) and 8(d) show the stimulated absorption observed with a different transducer position x .

IV. UASING

On replacing the generator-driven transducer at z with one or more additional van der Pol auto-oscillating transducer circuits, as in Fig. 9, we have what may be termed a “uaser” (Ultrasound amplification by stimulated emission of radiation²¹), an acoustic analog to a laser. Of special interest are circumstances under which all oscillators synchronize to their mutual radiation field. One imagines that, as in the Kuramoto model,^{23–27} one would describe a large number of such synchronizing auto-oscillators as experiencing a phase transition. This differs from the directed entrainment discussed in Sec. III in which an oscillator locks to a prescribed applied field. The system here is similar to that encountered

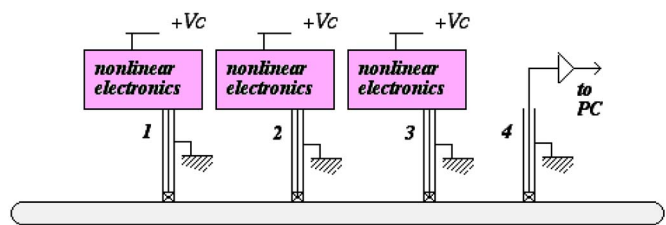


FIG. 9. (Color online) Two or three autonomous nonlinear piezoelectric oscillators are placed in contact with the aluminum rod of Fig. 7. The acoustic state is monitored with a passive detector.

when two or more Josephson junctions couple and synchronize through their microwave radiation field, a system noted elsewhere as analogous to a laser.²⁰ The data presented here are not meant to fully elucidate the properties of such systems. Rather these illustrations are presented to provoke imagination and suggest further studies. A thorough study awaits development of a theory to inform such experiments.

Figure 10 shows the behavior of the system of three auto-oscillators (labeled 1, 2 and 3) monitored by a receiver (labeled 4). For reference, we plot the three transfer functions h_{14} , h_{24} , and h_{34} between the auto-oscillators 1, 2 and 3, and the monitor at 4. Bold vertical lines indicate the rms amplitudes U_j received at 4 from each individual auto-oscillator when other oscillators are turned off, of each pair, U_{ij} when only one is off, and U_{123} when all three are on. It is apparent that any pair of oscillators synchronize to each other, and that all three oscillators synchronize to each other also. Frequency of synchronization is not always between the natural frequencies of the individual oscillators. This synchronization is similar to that described elsewhere (see Refs. 23–26, especially Ref. 27, Chapters 10 and 11). We may, however, also examine issues of ultrasonic power emission, stimulated emission and super radiance.

It behooves us to renormalize the amplitudes when comparing amplitudes emitted by an oscillator at two different frequencies, as the transfer function of the structure varies with frequency. An amplitude as measured by a receiver at position 4 due to emissions from oscillator j at frequency f_j ($j=1,2,3$) may be renormalized for comparison with the

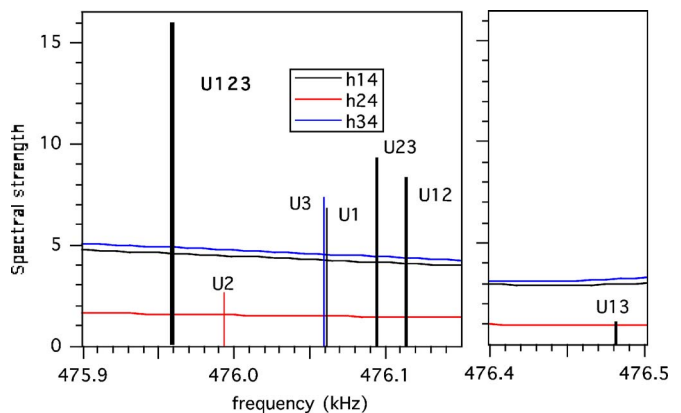


FIG. 10. (Color online) Three auto-oscillators are monitored by a single receiver at point 4. The rms signal strengths (rescaled to fit on the plot) are indicated by the vertical lines. Amplitudes of single auto-oscillators (thin vertical lines) are compared to those when auto-oscillators are taken in pairs (wider lines), and when all three auto-oscillators are powered (widest line).

same oscillator's contribution at another frequency f , by replacing the rms amplitude U_j with U_j times $h_{j4}(f)/h_{j4}(f_j)$. We compare the amplitude produced by the pair of oscillators 1 and 2 (labeled U12 in the figure) with their emissions U1 and U2 when isolated. The normalized amplitudes U1 and U2 of individual oscillators 1 and 2 (6.53 and 2.43 when referred to frequency $f=476.114$) add to approximate the amplitude of that pair together (U12=8.27). The normalized amplitudes of oscillators 2 and 3 (2.45 and 7.15 when referred to frequency 476.095) add to approximate the amplitude of that pair together (9.20). These pairs both add constructively. Oscillators 1 and 3 have normalized amplitudes (5.19 and 4.73 when referred to frequency 476.482) that roughly subtract to approximate the amplitude (1.05) of that pair together. Oscillators 1, 2 and 3 (normalized amplitudes 7.31, 2.66, and 7.88, respectively, when referred to frequency 475.959) add to approximate the amplitude of all three together (15.93). These comparisons show that acoustic energy is generated coherently, and at more than twice the rate that would be obtained in the absence of feedback. The system is super-radiant.³⁶

V. SUMMARY

It has been shown that nonlinear van-der Pol-like piezoelectric oscillators can be configured to exhibit the key behaviors required of sets of neighboring continuously pumped "atoms" in a classical analog for a laser. These include frequency locking, stimulated and spontaneous emission, stimulated absorption, and super-radiance. We conjecture that the principles illustrated here may find direct application in the construction of new kinds of acoustic generators, and indirect application in scale-model emulation of laser dynamics, in particular in research on random³¹ and chaotic³⁷ and photonic crystal lasers^{38,39} and arrays of Josephson junctions.^{22,30}

ACKNOWLEDGMENTS

This work was supported by the NSF, Grant No. CMS 05-28096. A.Y. acknowledges support from University of Missouri-Rolla.

- ¹K. Popp and P. Stelter, "Stick-slip vibrations chaos," *Philos. Trans. R. Soc. London* **332**, pp. 89–105 (1990); S. N. Patek, "Spiny lobsters stick and slip to make sound," *Nature (London)* **411**, 153–154 (2001).
- ²S. Adach, "Principles of sound production in wind instruments," *Acoust. Sci. & Tech.* **25**, 400–405 (2004); T. D. Mast and A. D. Pierce, "Describing-function theory for flow excitation of resonators," *J. Acoust. Soc. Am.* **97**, 163–172 (1995).
- ³P. van Dijk and H. P. Wit, "Amplitude and frequency fluctuations of spontaneous otoacoustic emissions," *J. Acoust. Soc. Am.* **88**, 1779–1793 (1990).
- ⁴G. W. Swift, "Thermoacoustic engines," *J. Acoust. Soc. Am.* **84**, 1145–1180 (1988).
- ⁵A. Larsen, "Ein akustischer Wechselstromerzeuger mit regulierbare Periodenzahl für schwache Ströme," *Elektrotechnische Zeitschrift*, ("An acoustic alternating-current generator for weak currents, with adjustable frequency") *Elektrotech. Z.* **32**, 284–285 (1911); A. Kjerbye Nielsen, "'Larsen-effekten' og den første elektriske tonegenerator baseret herpå," ("The 'Larsen effect' and the first electrical pure-tone generator based on this effect") *Teletek. (Engl. ed.)* **3**, 140–481 (1984); D. Barbaro, "Self starting acoustic oscillations in closed spaces," *Alta Freq.* **27**, 472–85 (1958).
- ⁶R. L. Weaver and O. I. Lobkis, "On the line width of the ultrasonic Larsen effect in a reverberant body," *J. Acoust. Soc. Am.* **120**, 102–109 (2006).

- ⁷A. Furduev, "Acoustic monitoring of the sea medium variability: Experimental testing of new methods," *Acoust. Phys.* **47**, 361–268 (2001).
- ⁸J. M. Kates, "Constrained adaptation for feedback cancellation in hearing aids," *J. Acoust. Soc. Am.* **106**, 1010–1019 (1999).
- ⁹S. M. Kim, S. J. Elliott, and M. J. Brennan, "Decentralized control for multichannel active vibration isolation," *IEEE Trans. Control Syst. Technol.* **9**, 93–100 (2001).
- ¹⁰P. S. Spoor and G. W. Swift, "The Huygens entrainment phenomenon and thermoacoustic engines," *J. Acoust. Soc. Am.* **108**, 588–599 (2000); P. S. Spoor and G. W. Swift, "Mode-locking of acoustic resonators and its application to vibration cancellation in acoustic heat engines," *J. Acoust. Soc. Am.* **106**, 1353–1362 (1999).
- ¹¹I. I. Blekhnman, *Synchronization in Science and Technology* (ASME Press, New York 1988); B. Bennett, M. F. Schatz, H. Rockwood, and K. Wiesenfeld, "Huygens' clocks," *Proc. R. Soc. London, Ser. A* **458**, 563–579 (2002).
- ¹²J. Pantaleone, "Synchronization of metronomes," *Am. J. Phys.* **70**, 992–1000 (2002).
- ¹³M. Borenstein and W. E. Lamb, "Classical laser," *Phys. Rev. A* **5**, 1298 (1972).
- ¹⁴M. Sargent III, M. O. Scully, and W. E. Lamb, Jr., *Laser Physics* (Addison-Wesley, Reading, MA, 1974).
- ¹⁵B. Fain and P. W. Milonni, "Classical stimulated-emission," *J. Opt. Soc. Am. B* **4**, 78–85 (1987).
- ¹⁶Y. A. Kobelev, L. A. Ostrovsky, and I. A. Soustova, "Nonlinear model of autophasing of classical oscillators," *Zh. Eksp. Teor. Fiz.* **99**, 470–480 (1991).
- ¹⁷V. V. Bredikhin, Yu. A. Kobelev, and N. I. Vasilinenko, "Autophasing of free volume oscillations of air cavities in water," *J. Acoust. Soc. Am.* **103**, 1775–1786 (1998).
- ¹⁸S. T. Zavtrak, "Acoustical laser with mechanical pumping," *J. Acoust. Soc. Am.* **99**, 730–733 (1996).
- ¹⁹I. V. Volkov, S. T. Zavtrak, and I. S. Kuten, "Theory of sound amplification by stimulated emission of radiation with consideration for coagulation," *Phys. Rev. E* **56**, 1097 (1997).
- ²⁰B. G. Danly and R. J. Temkin, "Generalized nonlinear harmonic gyrotron theory," *Phys. Fluids* **29**, 561–567 (1986).
- ²¹A. Yamilov, R. Weaver, and O. Lobkis, *UASER: Ultrasound Amplification by Stimulated Emission of Radiation*, *Photonics Spectra* (August 2006).
- ²²P. Barbara, A. B. Cawthorne, S. V. Shitov, and C. J. Lobb, "Stimulated emission and amplification in Josephson Junction arrays," *Phys. Rev. Lett.* **82**, 1963–1966 (1999).
- ²³A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear sciences*, (Cambridge University Press, Cambridge 2001). The terms *synchronization*, *locking* and *entrainment* are used in the literature in various ways. In the current paper *entrainment* is taken to describe the state of a limit-cycle oscillator when its frequency conforms precisely to that of an external forcing. We choose the term *locking* for the same case, but to emphasize the fixed phase relation. The term *synchronization* is used here to describe the autonomous entrainment of a set of auto-oscillators to their mutual radiation field in the absence of a prescribed forcing. A fourth term, *pulling*, describes change—but without full conformation—of the frequency of a nonlinear oscillator under the influence of a continuous harmonic force.
- ²⁴S. H. Strogatz, "From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators," *Physica D* **143**, 1–20 (2000).
- ²⁵H. Daido, "Quasi-entrainment and slow relaxation in a population of oscillators with random and frustrated interactions," *Phys. Rev. Lett.* **68**, 1073–1076 (1992).
- ²⁶P. C. Matthews, R. E. Mirollo, and S. H. Strogatz, "Dynamics of a large system of coupled nonlinear oscillators," *Physica D* **52**, 293–331 (1991).
- ²⁷P. S. Landa, *Regular and Chaotic Oscillations* (Springer New York, 2001).
- ²⁸R. Adler, "A study of locking phenomena in oscillators," *Proc. IEEE* **61**, 1380–1385 (1973).
- ²⁹Y. Kuramoto, *Chemical Oscillations and Waves* (Springer-Verlag, New York, 1984).
- ³⁰P. Hadley, M. R. Beasley, and K. Wiesenfeld, "Phase locking of Josephson-junction series arrays," *Phys. Rev. B* **38**, 8712–8719 (1988); K. Wiesenfeld, K. P. Colet, and S. H. Strogatz, "Synchronization transitions in a disordered Josephson series array," *Phys. Rev. Lett.* **76**, 404–407 (1996).
- ³¹D. S. Wiersma, M. P. VanAlbada, and A. Lagendijk, "Random laser," *Nature (London)* **373**, 203–204 (1995); C. Vanneste, and P. Sebbah, "Selective excitation of localized modes in active random media," *Phys. Rev.*

Lett. **87**, 183903 (2001).

³²G. S. Kino, *Acoustic Waves* (Prentice-Hall, Englewood Cliffs, NJ, 1987).

³³M. Lakshmanan and K. Murali, *Chaos in Nonlinear Oscillators. Controlling and Synchronization* (World Scientific, Singapore, 1996).

³⁴Ramon Vargas-Patron, personal communication; see also (last seen online 7/2007) <http://cidtel.inictel.gob.pe/cidtel/contenido/Publicaciones/rvargas/ORANRD.pdf>

³⁵A. L. Schawlow and C. H. Townes, "Infrared and optical masers," *Phys. Rev.* **112**, 1940 (1958).

³⁶R. L. Weaver, "Wave chaos in elastodynamics, pp. 141–186 in *Waves and Imaging through Complex Media*, edited by P. Sebbah, *Proceedings of the International Physics School on Waves and Imaging through Complex*

Media, Cargese, France (Kluwer, Dordrecht, 2001).

³⁷We use the term in the sense of R. H. Dicke, *Phys. Rev.* **43**, 102 (1954). "For want of a better term, a gas which is radiating strongly because of coherence will be called 'super-radiant'" to refer to emission that scales with the number N of emitters faster than the first power.

³⁸V. A. Podolskiy, E. E. Narimanov, W. Fang, and H. Cao, "Chaotic micro-lasers based on dynamical localization," *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10498–10500 (2004).

³⁹S. Strauf, K. Hennessy, M. T. Rakher, Y. S. Choi, A. Badolato, L. C. Andreani, E. L. Hu, P. M. Petroff, and D. Bouwmeester, "Self-tuned quantum dot gain in photonic crystal lasers," *Phys. Rev. Lett.* **96**, 127404 (2006).

Effects of coupled vibrations on the acoustical performance of underwater cylindrical shell transducers

Boris Aronov, David A. Brown,^{a)} and Corey L. Bachand

BTech Acoustics, LLC, Acoustics Research Laboratory, Advanced Technology and Manufacturing Center, and Department of Electrical and Computer Engineering, University of Massachusetts Dartmouth, 151 Martine Street, Fall River, Massachusetts 02723

(Received 26 November 2006; revised 26 March 2007; accepted 28 August 2007)

An experimental investigation of the effects of coupled vibrations on the acoustical performance of underwater transducers made from radially polarized, thin walled, air-backed, piezoelectric cylinders as a function of their height-to-diameter aspect ratio is presented. Characteristics of the frequency response, directivity patterns, and effectiveness of the transducers are considered in comparison with analogous characteristics for the transducers comprised of mechanically separated rings, otherwise having the same geometry. Recommendations are made on the application of cylindrical piezoelectric elements with different aspect ratios for underwater transducer applications. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2793602]

PACS number(s): 43.38.Ar, 43.38.Fx, 43.38.Pf, 43.30.Yj [AJZ]

Pages: 3419–3427

I. INTRODUCTION

Air-backed cylindrical piezoelectric transducers are commonly used as transmitters and receivers for broadband underwater applications. When designing underwater cylindrical transducers, guidance in choosing the dimensions of the comprising piezoelectric elements is useful in order to enhance the overall transducer performance. Experimental results were presented in a recent paper¹ on the dependence of the resonance frequencies, effective coupling coefficients, and vibration mode shapes of radially poled air-backed cylindrical elements on their height-to-mean diameter aspect ratio ($h/2a$). However, the previous experimental study¹ concentrates on the properties of unloaded mechanical cylinders and not the performance of loaded underwater electroacoustic transducers as a function of aspect ratio.

A theoretical approach for estimating the acoustical effects of the coupled mechanical vibrations in the transducer's mechanical system is achievable by solving the coupled electro-elastic-acoustic problem. An example of such an electroacoustic analysis, which takes into account the coupled vibrations and the radiation, of a tangentially polarized cylindrical piezoelectric element was reported.² The numerical data in that study were based on determining the radiation load from the assumption that the cylinder vibrates uniformly. Another paper³ solved the electroacoustic problem for a radial polarized cylindrical element using the free-flooded cylindrical transducer design and compared this with experimental results for a short ring and a relatively tall cylinder. These results cannot be applied directly to the case of the air-backed transducers under consideration in the current study. It is of note that the existence of capable numerical methods does not exclude the need for independent experimental verification, no matter how elaborate the mathematical model, due to the simplifying assumptions that inevitably must be made there within.

A literature search for publications on experimental studies related to the effect of coupled vibrations on the acoustical parameters of underwater air-backed cylindrical transducers resulted in one paper (Ref. 4) in which the transmit frequency response of air-backed transducers and their electrical power consumption as a function of piezoceramic element aspect ratio were reported. Unfortunately, that paper has a very limited distribution and availability. Further, many aspects of the electroacoustic study were not considered and no general conclusions or recommendations were presented.

Thus it is reasonable to proceed with an experimental investigation of the effects of coupled vibrations on the electroacoustic performance of air-backed, cylindrical transducers to gain insight into optimum performance and related tradeoffs. The required time and cost of such an investigation is reduced, given that only a limited range of aspect ratios is applicable for practical underwater transducers and the construction of air-backed, cylindrical shells with isolated ends is greatly simplified by holding the diameter of the samples fixed and using modular construction techniques.

The objectives of this paper are to illustrate the trends in the acoustical performance (radiation and reception of sound) of air-backed cylindrical transducers, resulting from the effects of coupled vibrations in their mechanical systems, and to estimate the merit of using radially poled cylindrical piezoelements of different aspect ratios for practical underwater applications.

The scope of the current paper includes a full scale experimental investigation of the parameters of transducers composed of thin hollow cylindrical or tubular piezoelectric elements having aspect ratios ranging from $h/2a=0.54$ to 2.8. Section II covers the reasons for choosing this range of aspect ratios for experimentation, the criteria for estimating transducers' properties, and the description of the samples used in this work. In Sec. III the experimental results of the transducers' parameters as a function of aspect ratio are presented, as well as results associated with peculiar characteristics in the frequency responses and corresponding directiv-

^{a)}Electronic mail: dbacoustics@cox.net

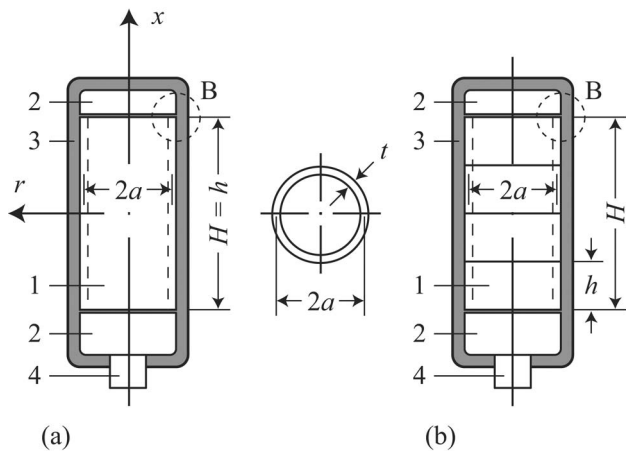


FIG. 1. Schematic representation of the air-backed cylindrical transducers for this work: (a) Made from the axially solid cylinders of different aspect ratio and (b) made from a number of the mechanically isolated short cylinders. In both cases the active piezoelectric elements (1) are mechanically isolated from the caps (2), and the entire structure is encapsulated in polyurethane (3). Item 4 presents an electrical connector. The coordinate system for the transducer is illustrated in (a) as the axis r extending radially and the axis x extending axially. The individual cylinder height is denoted as h , mean diameter as $2a$, thickness as t , and overall transducer active height as H . Location B indicates where mechanical losses from shear deformation are greatest.

ity patterns of the transducers. Results obtained are discussed and recommendations are made in Sec. IV for the application of transducer elements with different aspect ratios.

II. DESCRIPTION OF THE SAMPLES USED AND CRITERIA FOR THE COMPARISON OF TRANSDUCERS

The experimental investigations were made using air-backed cylindrical transducers comprised of radially poled cylinders, as illustrated in Fig. 1(a), having different aspect ratios. In all cases the distal ends of the cylinders are mechanically isolated from their base and cap. For comparison purposes, additional transducers of the corresponding overall height were built with the same transducer design methodology, presented in Fig. 1(b), from a number of mechanically separated and electrically connected (in parallel) cylinders with aspect ratio $h/2a=0.54$. The transducers are compared for the following reasons. First, the acoustic parameters of the finite height cylindrical transducers may change due to the change in overall height-to-wavelength ratio (wave size), H/λ , of the transducers, regardless of effects of coupled vibrations in the piezoelements on the axial velocity distribution. The comparison can help with characterizing the acoustical effects of transducers comprised of rings that are sufficiently short enough to be considered as one dimensional and to ensure that the velocity distribution over the transducer surface is uniform. Second, such comparison transducers made from short rings are in common use, and their characteristics can be conveniently used as a baseline for a comparison in terms of possible applications.

The electromechanical parameters and the mode shape of vibration of the cylinders with $h/2a=0.54$ are very close to those predicted under the assumption of one-dimensional

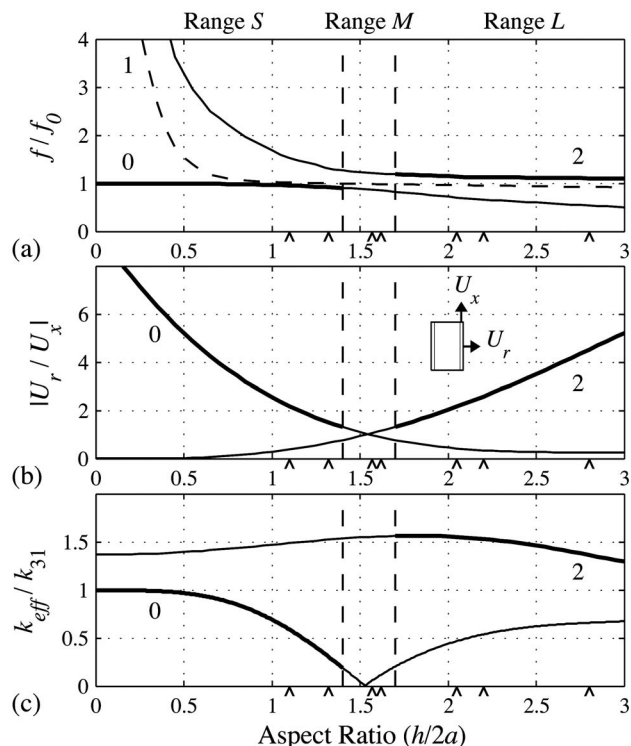


FIG. 2. Dependence of the electromechanical parameters of the radially poled cylindrical transducers on their aspect ratios (following Ref. 1): (a) The normalized resonance frequencies, (b) the radial-to-axial velocity ratios (U_r/U_x), and (c) the normalized effective coupling coefficients. Markers along the aspect ratio axes indicate the aspect ratio of axially solid transducers used in this work ($h/2a=1.1, 1.32, 1.57, 1.62, 2.05, 2.2, \text{ and } 2.8$). The aspect ratios are divided into three ranges: Range S is for small aspect ratios ($h/2a < 1.4$), range L is for large aspect ratios ($h/2a > 1.7$), and range M is for moderate aspect ratios ($1.4 \leq h/2a \leq 1.7$), where strong coupling between radial and axial modes takes place.

vibration for the short rings.¹ For brevity, the transducers made of single piezoelements with different aspect ratios will be identified as *axially solid* and the transducers with multiple ring elements will be identified as *axially segmented*. Particular values of the axially solid cylinder aspect ratios were chosen based on the data of dependence of the resonance frequencies, radial-to-axial velocities ratio, U_r/U_x , and the *effective coupling coefficients*, k_{eff} , for the piezoelectric tubular (ring) element aspect ratios, presented in Figs. 3, 5, and 6 of Ref. 1. For illustrative purposes these figures are reproduced in Figs. 2(a)–2(c), respectively. Although, a detailed explanation of the relations displayed in Fig. 2 are given in Ref. 1, to which this paper can be considered as a continuation, several brief summary comments might be useful for the current paper.

The resonance frequency branches are labeled as 0, 1, and 2 in the order of their appearance in the resonance frequencies spectrum at small aspect ratios. Label 0 for the lowest frequency branch originates from the common notation for the breathing mode (0 mode) of a very short ring with the resonance frequency $f_0=1/(2\pi a\sqrt{\rho s_{11}^E})$, where ρ is the density and s_{11}^E is the elastic constant of the piezoceramic material. Frequency branch 1 corresponds to the coupled axisymmetric flexural modes of the cylinder wall vibration as described in Ref. 1. While the resonance frequency of branch

1 (flexural mode) is relevant and shown by the dashed line in Fig. 2(a), the corresponding ratio of radial-to-axial velocity and coupling coefficient for branch 1 are omitted in Figs. 2(b) and 2(c), respectively. The coupling coefficient for the flexural branch is relatively small¹ as is the corresponding acoustic radiation.

In each of the modes of vibration related to the resonance frequency branches, vibrations take place in the radial, r , and the axial, x , directions. Ratios of magnitudes of vibrations, $|U_r/U_x|$, for both branches 0 and 2 are presented in Fig. 2(b) together with an icon showing at what points on the cylinder surface U_r and U_x are measured. The effective coupling coefficient, k_{eff} , for the modes change in correspondence with the distribution of deformations in vibration modes over the volume of the cylindrical piezoelement, which change as a function of the aspect ratio. The dependencies of k_{eff} for the modes related to the frequency branches 0 and 2 are presented in Fig. 2(c) as a function of aspect ratio. From Fig. 2 it follows that for the modes of vibration related to branch 0, the effective coupling coefficient drops to zero at the aspect ratio $h/2a = \pi/2 = 1.57$, i.e., at the point of the strongest coupling. This means that the corresponding mode of vibration cannot be generated through electromechanical transformation.

Note that the resonance frequency curves in Fig. 2(a) are normalized to the resonance frequency of a short ring, f_0 , the curves for the effective coupling coefficients in Fig. 2(c) are normalized to the coupling coefficient of the ceramic composition used, k_{31} , and the results do not depend on the thickness of the cylinder.

The whole range of aspect ratios tested can be divided conditionally in several ranges, shown in Fig. 2, by the following considerations. So far as the radiation of radially poled cylinders is concerned, there are certain ranges of aspect ratios that are more advantageous, for which U_r/U_x and k_{eff} are larger. There are two such ranges: range “S,” related to small aspect ratios ($h/2a < 1.4$) with corresponding resonance frequencies at the lower branch (0), and range “L,” related to larger aspect ratios ($h/2a > 1.7$) with resonance frequencies at the upper branch (2), which appear as bold lines in Fig. 2. In both ranges, interference between the radial components of vibrations in the modes belonging to the upper and lower branches and their effects on the frequency responses of the transducers may be expected. Range “M” around the point of the strongest coupling (around $h/2a = 1.57$) is of interest because a strong radiation can be expected at the resonance frequencies belonging to branch 2 ($k_{\text{eff},2}$ is large), and the contribution of the vibration related to branch 0 should be minimal ($k_{\text{eff},0}$ is very small). Based on these considerations several transducers were tested made from the PZT-4 cylindrical piezoelements having aspect ratios $h/2a = 1.1, 1.32, 1.57, 1.62, 2.05, 2.2,$ and 2.8 . (The corresponding points are shown by markers on the coordinate axes in Fig. 2).

All of the axially solid piezoelements were built by bonding several rings together, as done for the experiments described in Ref. 1, namely, from PZT-4 rings with the mean diameter $2a = 35$ mm, height $h = 19$ mm ($h/2a = 0.54$), and thickness $t = 3.2$ mm. (In the case that $h/2a = 1.32, 1.57,$ and

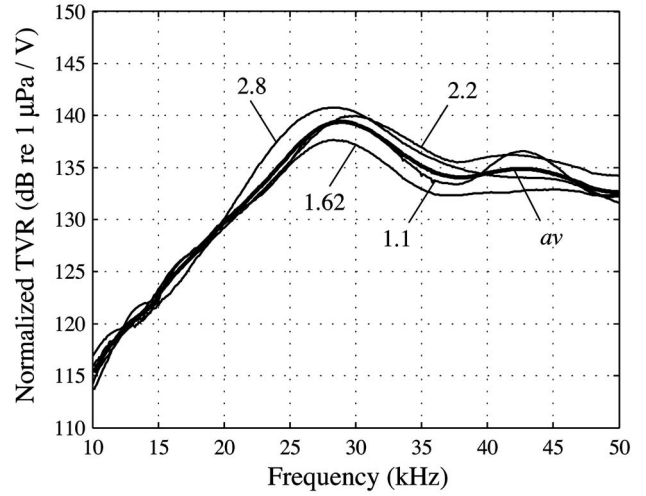


FIG. 3. The transmit frequency response (TVR) of each axially segmented transducer normalized to its overall aspect ratio $H/2a$, given by Eq. (5). Curves corresponding to the aspect ratios $H/2a = 1.1, 1.62, 2.2,$ and 2.8 are presented along with the curve labeled *av* representing the averaged normalized transmit frequency response for the axially segmented transducers.

2.05 the additional rings were obtained by cutting the cylinders.) As pointed out in Ref. 1, bonding of the rings with a rigid epoxy compound (Devcon 2-Ton Epoxy) ensured a uniformity of the elastic and electromechanical properties along the samples, having the aspect ratios in the range of interest. Consequent bonding of the cylindrical samples of the same geometry showed excellent repeatability of the resonance frequencies and effective coupling coefficients, which are parameters most vulnerable to deviation of the elastic properties of a mechanical system. This approach is preferred over experimentation of cylinders with very tall aspect ratios due to the manufacturing limitations on uniform composition.

In order to characterize the performance of each transducer, the transmit and receive frequency responses, directivity factors, and input impedance related parameters were measured. All the measurements were performed in the underwater calibration facility located at the School for Marine Science and Technology (SMAST) in New Bedford, MA. Standard reference calibration procedures were used with Brüel & Kjær Type 8104 reference hydrophones to achieve an accuracy of ± 1.5 dB.

The following has to be noted with regard to the transducer’s power-handling capacity. Although the corresponding axially segmented transducers of comparison have the same height, their directional factors in the vertical plane, and hence their directivities, may be different due to the nonuniform velocity distribution for the axially solid transducers. Therefore, the electroacoustic efficiency figure-of-merit, which takes into account the directivity, may not be accurate for comparing the power-handling capacity of transducers. It seems more appropriate for this purpose to use the figure-of-merit defined as

$$Ef = (I_{ax} \cdot 1\text{m}^2) / \dot{W}_{el}, \quad (1)$$

which will be called the *effectiveness* of a transducer. Here $I_{ax} 1\text{m}^2$ is the product of the intensity on axis of the transducer, referenced to 1m from its acoustic center, and an area

of 1 m^2 . \dot{W}_{el} is the total electrical power delivered to the transducer per 1 V applied. The factor 1 m^2 is used in order to make the effectiveness a nondimensional quantity. The effectiveness in Eq. (1) may be transformed after substituting the expressions $I_{ax} = P_{ax}^2 / \rho c$ for intensity and $\dot{W}_{el} = V^2 G / \cos \varphi$ for electric power into

$$Ef = P_{ax}^2 \cos \varphi / (\rho c V^2 G), \quad (2)$$

where P_{ax} is the sound pressure on the acoustic axis per 1 V applied referred to 1 m from the acoustic center of a transducer, G is the conductance, and $\cos \varphi$ is the power factor of the transducer input admittance.

An advantage of the effectiveness in comparison with the electroacoustical efficiency, η_{ea} , is that the effectiveness accounts for the “useful” acoustic power radiated in the direction of the acoustic axis only, whereas the efficiency includes the total acoustic power and hence radiation in unwanted directions. The relation between the effectiveness and efficiency is

$$\eta_{ea} = Ef(4\pi/D \cos \varphi), \quad (3)$$

where D is the directivity, which can be different for the transducers under comparison because of the different distribution of velocities over their radiating surfaces.

It is convenient to calculate effectiveness from experimental data and express it in decibels as

$$\begin{aligned} Ef \text{ (dB)} &= 10 \log Ef \\ &= \text{TVR} - 10 \log G_{mS} + 10 \log \cos \varphi - 152 \text{ dB}. \end{aligned} \quad (4)$$

When introducing this formula based on Eq. (2), it is assumed that $\rho c = 1.5 \times 10^6 \text{ kg s/m}^2$ and G_{mS} is taken in mS.

III. THE RESULTS OF EXPERIMENTAL INVESTIGATION

A. Axially segmented transducers

The transmit voltage response (TVR) of each transducer, normalized to its overall transducer active height-to-diameter aspect ratio ($H/2a$),

$$\text{Normalized TVR} = \text{TVR} - 20 \log(H/2a) \quad (5)$$

in dB re $1 \mu\text{Pa}/\text{V}$ at 1 m, is displayed in Fig. 3. The normalized TVR for all of the transducers tested, in which the diameter is held constant, can be represented by an averaged curve within an accuracy of $\pm 1.5 \text{ dB}$. The free-field voltage sensitivity (FFVS) response of each transducer is depicted in Fig. 4, where from it follows that the frequency responses

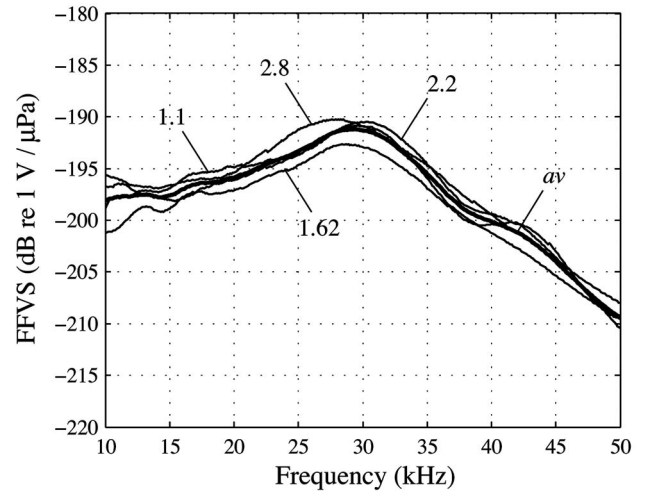


FIG. 4. The receive frequency responses (FFVS) of the axially segmented transducer for overall transducer height-to-diameter aspect ratios $H/2a = 1.1, 1.62, 2.2,$ and 2.8 . The curve labeled *av* represents the averaged receive frequency response for the axially segmented transducers.

can be represented by the averaged curve within an accuracy of $\pm 1.5 \text{ dB}$. FFVS is the open circuit sensitivity of the transducers, and therefore does not need to be normalized to the height as long as the rings are connected in parallel. When estimating the potential of a transducer as a receiver, one must keep in mind that the capacitance of the transducer is proportional to its height.

Parameters of the axially segmented transducers at their resonance frequency are presented in Table I. The mechanical quality factor, Q_m , in Table I was determined from the transmit frequency response as $Q_m = f_r / \Delta f_{-3 \text{ dB}}$, where $\Delta f_{-3 \text{ dB}}$ is the frequency band, over which the transmit response drops by less than 3 dB. The results show that for the axially segmented transducers Q_m practically does not depend on the number of rings. As $Q_m \approx \omega M_{eqv} / r_{ac}$, where M_{eqv} is the equivalent mass of a transducer and does not change per unit length for the axially segmented transducer, this implies that the radiation resistance per unit length also remains nearly constant. In other words, the combined effect of acoustical interaction between rings of a particular size is not significant, as verified by data in Figs. 3 and 4. Note that in the frequency range around the resonance frequency of the transducers, the individual ring height-to-wavelength ratio is $h/\lambda \approx 0.36$, and the overall transducer height-to-wavelength ratio is $H/\lambda > 0.7$.

TABLE I. Parameters of the axially segmented transducers at their resonance frequencies.

Ratio $H/2a$	No. of Rings	f (kHz)	Q_m	$\cos \varphi$	TVR (dB)	Ef_r (dB)
1.1	2	29.1	3.6	0.37	140	-16.0
1.62	3	29.0	3.5	0.36	142	-15.4
2.2	4	28.8	3.4	0.35	147	-12.0
2.8	5	28.9	3.4	0.36	150	-10.5

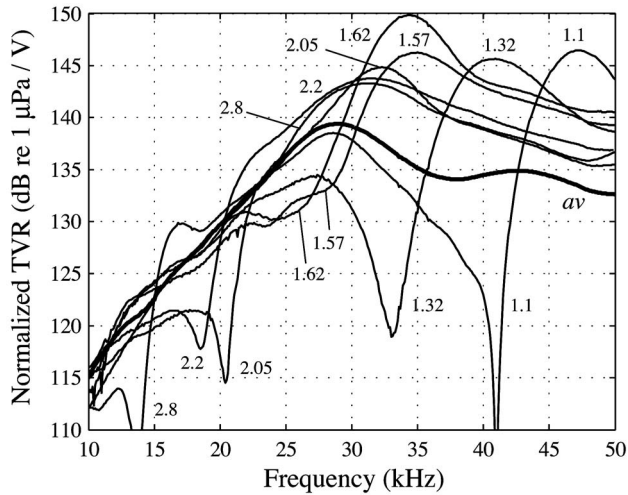


FIG. 5. The transmit frequency responses (TVR) of the axially solid transducers normalized to their aspect ratios. Curves corresponding to the aspect ratios $h/2a=1.1, 1.32, 1.57, 1.62, 2.05, 2.2,$ and 2.8 are presented, along with the curve labeled *av* representing the averaged normalized transmit frequency response for the axially segmented transducers.

B. Axially solid transducers with different aspect ratios

The frequency responses of the axially solid transducers with different aspect ratios are presented in Fig. 5 for the transmit mode (TVR) and in Fig. 6 for the receive mode (FFVS). Also in Figs. 5 and 6 the averaged frequency responses for the axially segmented transducers are shown for comparison. In order to make the comparison fair, TVRs for the axially solid transducers are normalized by their aspect ratio, as was already done for the axially segmented transducers. It makes sense to analyze the regions of the frequency response around the resonance frequencies of the transducers, which are typical for their operation as projectors or as reversible transducers, and the frequency bands

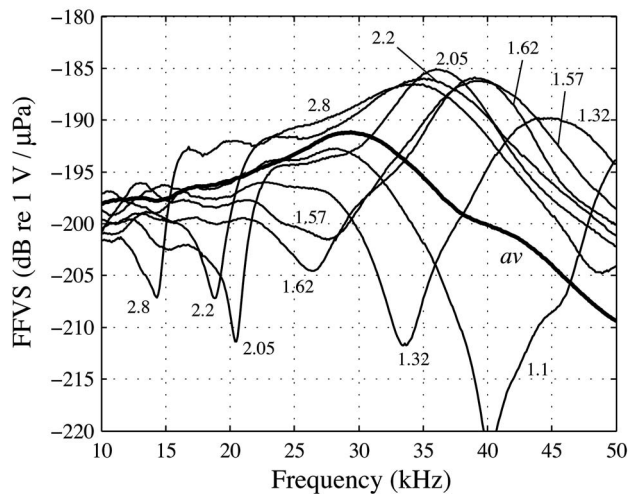


FIG. 6. The receive frequency response (FFVS) of the axially solid transducers for different aspect ratios. Curves corresponding to the aspect ratios $h/2a=1.1, 1.32, 1.57, 1.62, 2.05, 2.2,$ and 2.8 are presented, along with the curve labeled *av* representing averaged receive frequency response for the axially segmented transducers.

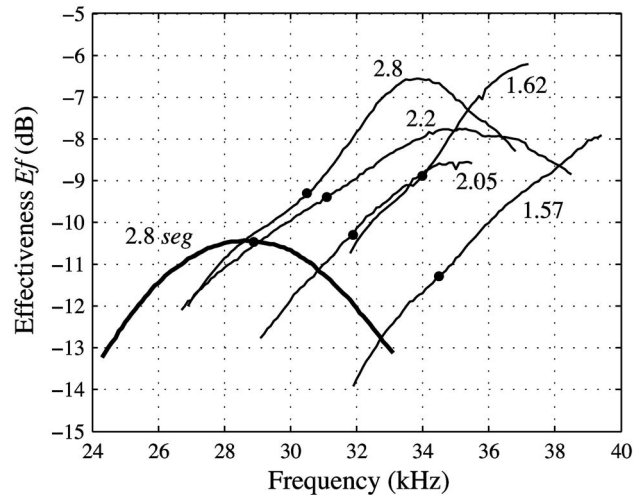


FIG. 7. The effectiveness (in decibels) of five axially solid transducers having different aspect ratios and one segmented transducer. The axially solid transducers have aspect ratios $h/2a=1.57, 1.62, 2.05, 2.2,$ and 2.8 . The curve labeled *2.8 seg* corresponds with the axially segmented five-ring transducer ($H/2a=2.8$). Markers (dots) on the curves indicate the resonance frequencies and the curves are band limited to the -3 dB frequency range.

below these regions, which are typical for the operation of transducers in the receive mode or separately as measurement projectors.

1. Operation near the resonance frequencies

The electroacoustical parameters of the axially solid transducers at their resonance frequencies are summarized in Table II. The effectiveness of the transducers over the $\Delta f_{-3 \text{ dB}}$ frequency band is shown in Fig. 7. The directional factors in the vertical plane for the axially solid transducers of different aspect ratios at their resonance frequencies are presented in Fig. 8, together with directional factors for the axially segmented transducers and uniformly vibrating line sources of the same length. Several conclusions can be drawn from the experimental results. The axially solid transducers in the small aspect ratio range ($h/2a=1.1$ and 1.32) are not optimal for operation at the lower resonance frequencies (corresponding to frequency branch 0) in comparison with the axially segmented transducers of the same overall height. The axially solid transducers can be used as dual resonance projectors with an adjustable separation between the resonances. However, at higher resonance frequencies (corresponding to branch 2), the axially solid transducers have larger Q_m and smaller effectiveness for two reasons. First, the acoustical loading is reduced due to the approximately $\cos(2\pi x/h)$ distribution of velocity as a function of height. Second, a significant component of vibration exists in the longitudinal direction, which does not produce radiation. (Remember that the caps are mechanically isolated from the ends of piezoelectric cylinders in the subject designs.)

The axially solid transducers with aspect ratios $h/2a = 1.57$ and 1.62 , close to the point of the strongest coupling, have a clear advantage over the axially segmented transducers of the same height in the sense that they provide greater sound pressure at higher effectiveness and power factor for operation on the upper resonance frequency branch. The rea-

TABLE II. Parameters of the axially solid transducers at their resonance frequencies.

Ratio $h/2a$	Branch	f (kHz)	Q_m	$\cos \varphi$	TVR (dB)	Ef_{ir} (dB)
1.1	0	28.0	3.7	0.25	139	-17.0
	2	46.0	8.9	1.00	147	-18.0
1.32	0	27.5	3.8	0.28	137	-21.0
	2	41.0	5.8	1.00	148	-14.0
1.57	2	34.6	4.6	0.94	150	-11.3
1.62	2	33.8	5.6	0.95	154	-8.9
2.05	2	31.7	4.9	0.85	151	-10.3
2.2	2	31.0	3.1	0.72	151	-9.4
2.8	2	30.6	3.0	0.77	152	-9.3

sions behind this are (1) the larger effective coupling coefficients [see Fig. 2(c) branch 0 for small aspect ratios and branch 2 for the aspect ratios around the point of the strongest coupling] and (2) the reduced acoustical loading due to nonuniform distribution of radial velocity as a function of height. (Note that the cylindrical transducers of this thickness-to-mean diameter ratio may be considered as overloaded under uniform distribution of vibration.)

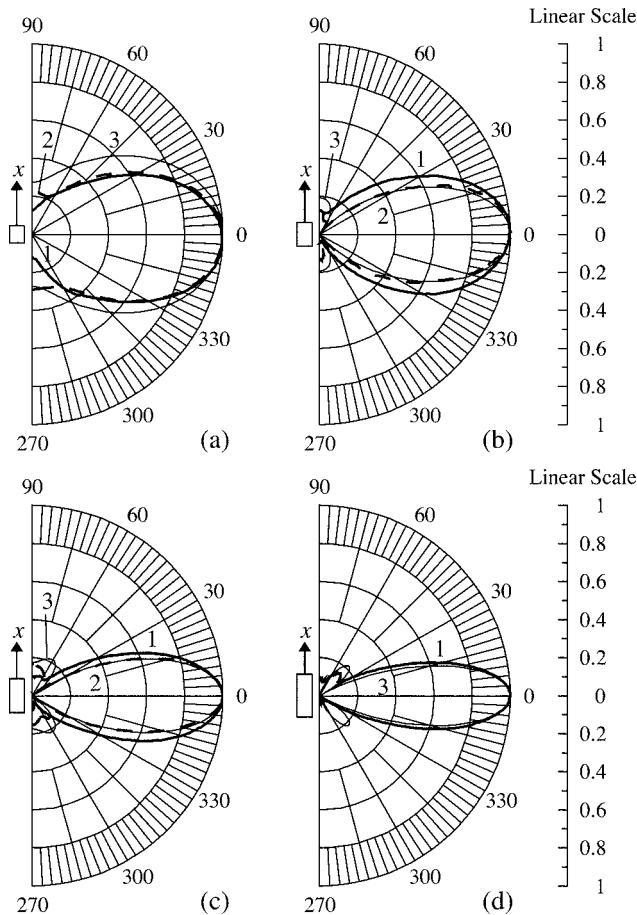


FIG. 8. The directional factors for (1—solid line) the axially solid transducers with different aspect ratios, (2—dashed line) axially segmented transducers of the same height, and (3—thin line) uniformly vibrating line sources having the same length: (a) $H/2a=1.1$, $f=30$ kHz; (b) $H/2a=1.57$, $f=35$ kHz; (c) $H/2a=2.2$, $f=32$ kHz, and (d) $H/2a=2.8$, $f=31$ kHz. Only one half of the directional factors is shown due to symmetry. The transducer icon near the origin of each plot indicates relative aspect ratio and orientation.

As the aspect ratio increases, the distribution of velocity gradually becomes more uniform, contribution from the longitudinal component of vibration drops, and the corresponding parameters of the transducers become independent of their aspect ratio. It is of note that the reduction of the effectiveness for transducers of small height, for both axially solid and axially segmented cases, has to be partly attributed to a relatively greater contribution of mechanical losses in the encapsulation of the parts of transducers located between the caps and the ends of vibrating cylinders (labeled as area B in Fig. 1), in which the shear deformations occur.

The directional factors for the axially solid transducers in the vicinity of their resonance frequencies are slightly broader and have smaller minor lobes than those for their axially segmented counterparts, which is especially noticeable in Fig. 8(b) at $H/2a=1.57$. This correlates with the nonuniform distribution of vibration that decreases at the ends. As the aspect ratio increases, and hence distribution of the radial velocity becomes more uniform, the directional factors become very close to the directional factors for a uniformly vibrating line source of the same length. A big difference between the directional factors calculated for the line source and those measured from axially solid and segmented transducers at aspect ratio $H/2a=1.1$ is due to diffraction effects on the real transducers of relatively small height-to-wavelength ratio (in this case $H/\lambda=0.7$).

The inherent peculiarity of the axially solid transducers is that their resonance frequencies change versus aspect ratio and remain higher than the resonance frequency of the axially segmented transducers of the same diameter. The axially solid transducers with aspect ratios $h/2a=1.57-1.62$ can be considered as the most applicable of solid transducers. The performance of transducers that are taller than these is not better, and their manufacturability becomes more difficult. The only reason to further increase the height of the solid transducers is to achieve a narrower beamwidth in the vertical plane for greater directivity.

It was found that the axially solid transducers may have an advantage over the axially segmented transducers in application for multimode operation.⁵ Operation of the multimode cylindrical baffled transducers is considered in Ref. 5. The transducers employ the two lower modes of extensional vibrations of the cylinder in the circumferential direction. These modes and their combination can be generated electromechanically, if the transducer electrodes are split into

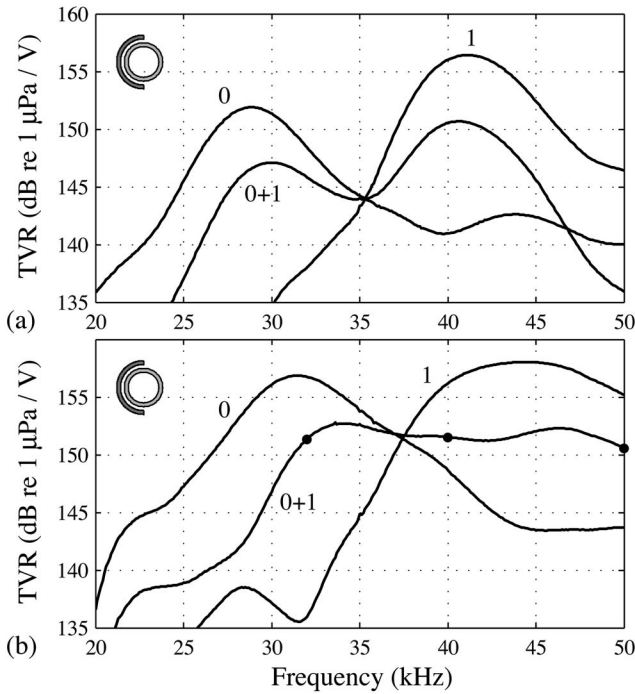


FIG. 9. The transmit frequency responses of (a) axially segmented and (b) axially solid baffled transducers with aspect ratio $H/2a=2.2$ in the multimode operation: (0) Breathing mode, (1) dipole mode, and (0+1) combined breathing and dipole mode. Markers (dots) on the curves indicate frequencies at which directional factors are presented in Fig. 10.

halves and then the halves are electrically connected in phase [breathing mode (mode 0) of vibration with resonance frequency f_0 generated], in antiphase [dipole mode (mode 1) of vibration with resonance frequency $f_1=f_0\sqrt{2}$ generated], or only one half of the electrodes is used (superposition of the two modes is generated).

In order to obtain unidirectional radiation in the horizontal plane for a frequency range of operation of both modes, the transducer is supplied with an acoustic baffle applied to one half of the cylinder surface. A detailed theoretical analysis of the transducer and its experimental verification in the case that the transducer is comprised of the same basic PZT-4 rings with aspect ratio $h/2a=0.54$, as were used throughout the current investigation, are presented in Ref. 5. In Fig. 9 comparison is made of the frequency responses of axially segmented transducer considered in Ref. 5 [shown in Fig. 9(a)] and the axially solid transducer of the same overall height $H/2a=2.2$ [shown in Fig. 9(b)]. The baffles had the configuration shown in Fig. 9. They were made from corprene with thickness 6 mm. The directional factors of the axially solid transducer in the horizontal plane for the frequency range of the multimode operation are presented in Fig. 10. They show that the directional factors are nearly frequency independent in a broad operational range.

2. Operation below the resonance frequencies

A notable feature of the frequency responses of axially solid transducers is the existence of notches at some frequencies in the range below their high frequency resonances (with exception of the transducers having the aspect ratio $h/2a$

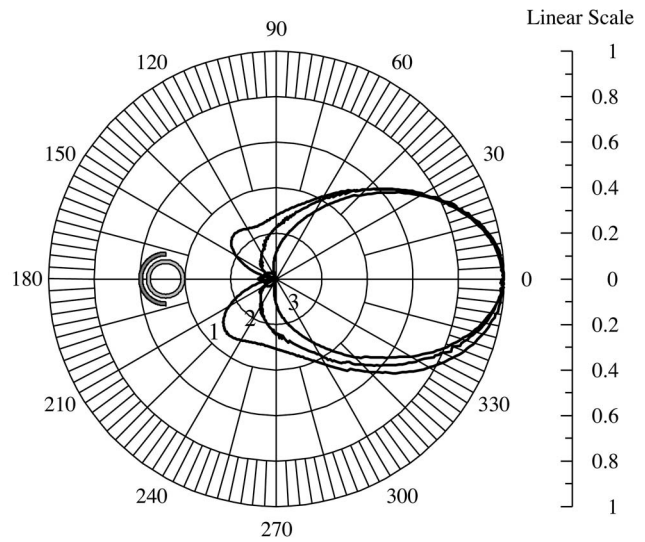


FIG. 10. The horizontal directional factors of a baffled axially solid transducer having $h/2a=2.2$ in the range of its combined (0+1) mode of operation at the frequencies: (1) $f=32$ kHz, (2) $f=40$ kHz, and (3) $f=50$ kHz. The corresponding points are marked on the frequency response of the transducer in Fig. 9(b). The directional factors are nearly frequency independent over a broad range of frequencies.

≈ 1.57). These irregularities may compromise the broadband operation of the transducers in the receive mode or when used as measurement projectors. In order to reveal the reasons for existence of the dips, an additional experimental investigation was undertaken. The directional factors were

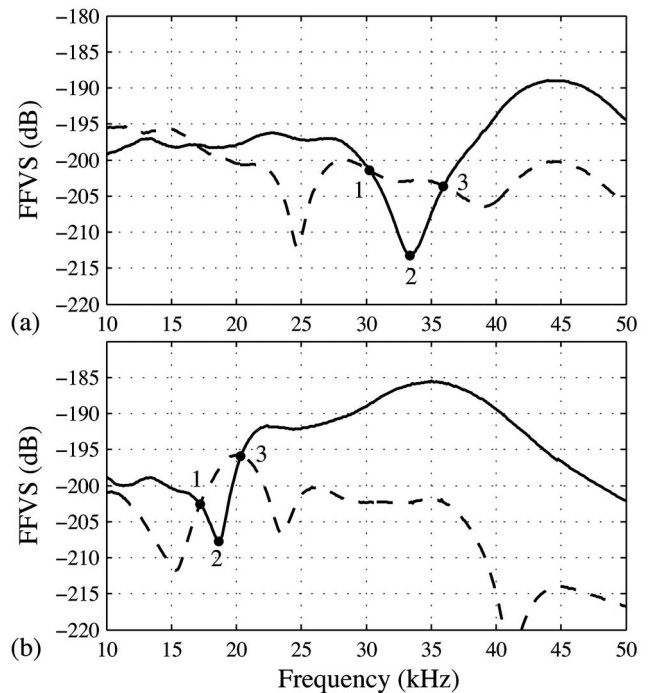


FIG. 11. Comparison of FFVS (dB re $1\text{ V}/\mu\text{Pa}$) measured in the radial (solid line) and axial (dashed line) directions for the transducers with different aspect ratios: (a) $h/2a=1.32$, (b) $h/2a=2.2$. Markers (dots) labeled 1 and 3 correspond to the frequencies at which the sensitivities in the radial and axial directions are equal, and point 2 corresponds to the frequency at which the sensitivity in the radial direction is minimal.

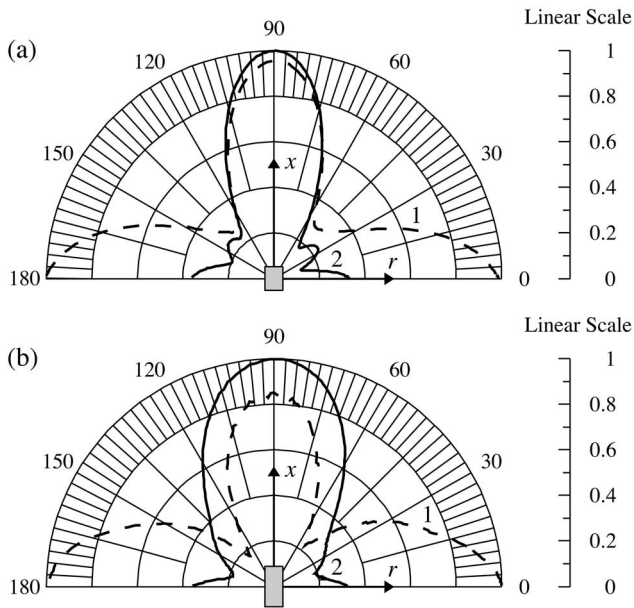


FIG. 12. The directional factors of (1—dashed line) the axially solid transducers at the frequency for which the sensitivities in the radial and axial directions are nearly equal and (2—solid line) at the frequencies of their minimal sensitivity for different aspect ratios: (a) $h/2a=1.32$ and (b) $h/2a=2.2$. The directional factors labeled 1 and 2 were measured at frequencies labeled as 1 and 2 in the FFVS responses in Fig. 11(a) and 11(b). The transducer icon near the origin of each plot indicates relative aspect ratio and orientation.

taken at the frequencies of the smallest receive and transmit levels. It occurred that the greatest response (sound pressures in the transmit mode) at these frequencies take place in the axial direction of the transducers. In order to quantify the frequency bands, over which the transmit response of a transducer in the axial direction exceeds that in the radial direction, the FFVS and TVR were measured in this direction and compared with those previously measured in the radial direction. The FFVSs for $h/2a=1.32$ and 2.2 are presented in Figs. 11(a) and 11(b). From the comparison of the FFVS measured in the radial and axial directions it clearly follows that the sensitivity on axis dominates in the frequency bands between points 1 and 3 in Fig. 11. Examples of the directional factors for transducers with aspect ratios typical for ranges S and L , taken (1) at the frequency where the radial and axial responses are nearly equal and (2) at the frequencies of the minimal response in the radial direction, are presented in Figs. 12(a) and 12(b). The directional factors, labeled as 1 and 2 in Fig. 12, correspond to points 1 and 2 in Fig. 11.

A qualitative explanation to the greater axial response follows. Radiation of a transducer at each frequency can be considered as the result of superposition of radiation by the modes of vibration belonging to the upper and lower frequency branches shown in Fig. 2(a). Each of these modes has the radial and axial components of vibration, U_r and U_x , respectively. It is shown in Fig. 2(b) that for small aspect ratios (range S) in the mode corresponding to branch 2 (denoted as mode 2 for brevity) the axial component dominates and in the mode corresponding to branch 0 (mode 0) the radial component dominates. It is of note that the radial com-

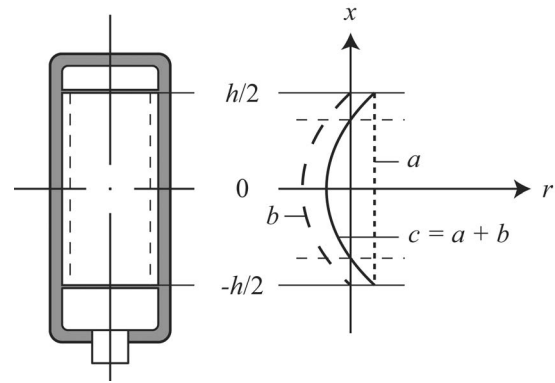


FIG. 13. Qualitative illustration of the superposition of modes of vibration, resulting in the velocity distribution that is favorable for radiation in the axial direction: (a) The radial component of the main mode belonging to branch 0 in range S and to branch 2 in range L (dotted line); (b) the radial component of the contributing mode belonging to branch 2 in range S and to branch 0 in range L (dashed line); and (c) the resulting distribution of velocity over the surface of the transducer (solid line).

ponent in mode 2 in this range has a nearly $\cos(2\pi x/h)$ distribution of magnitude. In the range L of aspect ratios the opposite is true: The radial component of vibration dominates in mode 2 and its distribution along the height becomes close to uniform as the aspect ratio increases, and in mode 0 the axial component dominates. The radial component in mode 0 has a nearly cosine distribution of magnitude of vibration along its height. In range M , the contribution of mode 0 should be weak due to the small effective coupling coefficient, whereas for mode 2, the effective coupling coefficient has a maximum value.

In a broad frequency range, the radiation is predominantly generated by the radial component of mode 0 in range S and by the radial component of mode 2 in the range L of aspect ratios. A contribution of the radial component of mode 2 in the range S , being otherwise small, may become significant in the vicinity of frequencies close to the resonance frequency corresponding to the upper branch for a given aspect ratio. Similarly, in range L , a contribution of the radial component of mode 0 may become significant in the vicinity of frequencies close to the resonance frequency of the lower branch for a given aspect ratio. In both cases the contributing components of vibration have a cosine-like distribution along the height of the transducer. Their magnitude and phase change significantly and the sign of the phase may also change in a frequency band around the resonance frequency of the contributing mode.

The superposition of the main and contributing modes of vibration may result in a distribution of velocity on the surface of a transducer shown qualitatively in Fig. 13. At some frequencies the average velocity over the transducer surface may become very small and the main radiation may then be in the axial direction due to a favorable phase distribution on the radiating surface for this direction.

IV. DISCUSSION OF RESULTS

As the experimental investigation reported in this paper was performed on transducers built from PZT-4 piezoceramic rings of a particular geometry (mean diameter of $2a$

=35 mm and thickness of 3.2 mm), a question may arise as to the extent to which the results obtained may be considered as general. Note that the curves in Fig. 2 do not depend on the cylinder thickness. The resonance frequency branches in Fig. 2(a) are normalized to the resonance frequency of a short ring, f_0 , and the effective coupling coefficients in Fig. 2(c) are normalized to the coupling coefficients of the ceramic material, k_{31} . Otherwise they depend only on the aspect ratio $h/2a$. Thus, qualitatively the results related to the electromechanical transducer behavior do not depend on the particular material properties and mean diameter. Quantitatively, the results can be scaled for changes in the material parameters.

The acoustical field related parameters (radiation impedances, diffraction coefficients, directivity patterns) of the cylindrical transducer for a given outer diameter depend on the wave size H/λ . The resonance frequency can be expressed as $f_r = f_0 \varphi(H/2a)$, where $\varphi(H/2a)$ presents a corresponding frequency branch in Fig. 2(a) and $f_0 = c_c/2\pi a$. The wavelength is given by

$$\lambda = \frac{c_w}{f_0} \frac{1}{\varphi(H/2a)} = \frac{c_w}{c_c} \frac{2\pi a}{\varphi(H/2a)}, \quad (6)$$

therefore the wave size can be expressed as

$$\frac{H}{\lambda} = \frac{H}{2a} \varphi(H/2a) \frac{c_c}{\pi c_w}. \quad (7)$$

In this relation, $c_c = 1/\sqrt{\rho s_{11}^E}$ and $c_w = 1500$ m/s is the sound speed in water. From Eq. (7) it follows that the wave size of a cylindrical transducer depends only on aspect ratio, with an accuracy up to the deviation of sound speed for a particular piezoceramic composition.

Although the thickness of the cylinder does not influence the resonance frequencies, or the general trends in the transducer performance, the thickness will affect the quality factor, Q_m , and accordingly the frequency response. However, the frequency response of the axially solid transducer will be similarly affected by the change in Q_m as is the axially segmented transducer, the behavior of which is well known.

In summary, a conclusion can be made that results of the experimental investigation of effects of coupled vibrations on the performance of cylindrical transducers composed from the rings of a particular geometry made of a particular ceramic composition are qualitatively fully representative, in terms of the trends in changes of the transducer parameters as a function of aspect ratio. Quantitatively, the relation between the frequency responses of axially solid transducers and axially segmented transducers made from the rings of the same geometry and of the same ceramic composition is presented.

Following this conclusion, general suggestions are made for the aspect ratios of cylindrical radially poled piezoelements used for underwater transducers based on the presented results of the experimental investigation.

For transducers used in arrays with electronically steerable beams, it is desirable to have the height-to-wavelength ratio of the piezoelements $h/\lambda < 0.5$. Given that the mean diameter of a transducer is related to its resonance frequency, and for a short ring $2a = c_c/\pi f_r$, where c_c is the sound speed in a ceramic composition used, it follows from $h/\lambda < 0.5$ that the height-to-diameter ratio is $h/2a < 0.5\pi c_w/c_c$. For PZT-4, the aspect ratio should be $h/2a < 0.7$ assuming that $c_w = 1500$ m/s. A higher effective coupling coefficient can be obtained by using piezoelements of smaller height (aspect ratio), but this may result in an increase in the acoustical interaction between elements.

Piezoelements with larger aspect ratios $h/2a > 1.5$ can be used effectively as single element projectors. In a frequency band around resonance they have an advantage in terms of electroacoustical parameters over transducers of the same height comprised of rings having small aspect ratios. However, these transducers also have disadvantages. First, they have a larger diameter (ranging from a 20% to 6% increase as the aspect ratio increases) for the same resonance frequency. Second, their directional factors in the vertical plane are related to their aspect ratio and therefore to their resonance frequency, whereas for the axially segmented transducers, increasing the height by adding rings does not affect the resonance frequency.

Piezoelements with large aspect ratios are not recommended for use as very broadband receivers (hydrophones) that operate predominantly below the resonant frequency because of the deep notches that occur in the frequency response arising from the effects of coupled vibrations. For this purpose, the best results can be achieved by employing mechanically separated piezoelements with small aspect ratios ($h/2a$ below approximately 0.6) that may be connected in parallel or series depending on the desired capacitance.

ACKNOWLEDGMENTS

This work was supported by BTech Acoustics, LLC and assisted in part by related work funded by ONR and the SBIR program. The transducers were designed and built by BTech Acoustics, LLC and testing was conducted at the Underwater Test Facility of the School for Marine Science and Technology, University of Massachusetts.

¹B. S. Aronov, D. A. Brown, and S. Regmi, "Experimental investigation of coupled vibrations in piezoelectric cylindrical shells," *J. Acoust. Soc. Am.* **120**, 1374–1380 (2006).

²H.-Z. Wang, "On the tangentially and radially polarized piezoceramic thin cylindrical tube transducers," *J. Acoust. Soc. Am.* **79**, 164–171 (1986).

³P. H. Rogers, "Mathematical model for a free-flooded piezoelectric cylindrical transducer," *J. Acoust. Soc. Am.* **80**, 13–18 (1986).

⁴V. I. Pozern, "Dependence of cylindrical projector parameters on height," *Voprosy Radioelectronics, Hydroacoustics Series XVII* (1961), Vol. **7** (in Russian).

⁵T. Oishi, B. S. Aronov, and D. A. Brown, "Broadband multimode baffled cylindrical shell transducers," *J. Acoust. Soc. Am.* **121**(6), 3465–3471 (2007).

Development of a micromachined piezoelectric microphone for aeroacoustics applications

Stephen Horowitz

*Department of Mechanical and Aerospace Engineering, Interdisciplinary Microsystems Group,
University of Florida, Gainesville, Florida 32611-6250*

Toshikazu Nishida

*Department of Electrical and Computer Engineering, Interdisciplinary Microsystems Group,
University of Florida, Gainesville, Florida 32611-6250*

Louis Cattafesta and Mark Sheplak^{a)}

*Department of Mechanical and Aerospace Engineering, Interdisciplinary Microsystems Group,
University of Florida, Gainesville, Florida 32611-1250*

(Received 5 December 2006; revised 10 July 2007; accepted 24 August 2007)

This paper describes the design, fabrication, and characterization of a bulk-micromachined piezoelectric microphone for aeroacoustic applications. Microphone design was accomplished through a combination of piezoelectric composite plate theory and lumped element modeling. The device consists of a 1.80-mm-diam, 3- μm -thick, silicon diaphragm with a 267-nm-thick ring of piezoelectric material placed near the boundary of the diaphragm to maximize sensitivity. The microphone was fabricated by combining a sol-gel lead zirconate-titanate deposition process on a silicon-on-insulator wafer with deep-reactive ion etching for the diaphragm release. Experimental characterization indicates a sensitivity of 1.66 $\mu\text{V}/\text{Pa}$, dynamic range greater than six orders of magnitude (35.7–169 dB, re 20 μPa), a capacitance of 10.8 nF, and a resonant frequency of 59.0 kHz. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2785040]

PACS number(s): 43.38.Fx, 43.38.Kb, 43.40.Dx, 43.38.Ar [AJZ]

Pages: 3428–3436

I. INTRODUCTION

As part of an ongoing effort to reduce the impact of aircraft noise on passengers, airports, and surrounding communities, the Federal Aviation Administration restricts the level of noise that can be emitted from aircraft.¹ These restrictions provide ample motivation for aircraft manufacturers to reduce emitted noise. A key component of noise reduction is the identification and localization of the noise sources. This requires measurement of the radiated noise using acoustic transducers with appropriate performance characteristics. The goal of the research presented in this paper is to develop a microphone for aeroacoustic measurement applications.²

There are two basic approaches for the measurement of aircraft noise: full-scale and reduced-scale testing. Full-scale testing employs fly-overs of full size aircraft, whereas reduced-scale testing uses 1/*n*th scale mock-ups in wind tunnels. In either case, noise source localization is performed using arrays of microphones, often in quantities on the $O(100)$ with a unit cost on the $O(1\text{k}\$)$.³ Furthermore, for 1/4 scale testing and certification, the microphones must possess a maximum sound pressure level (SPL) in excess of 160 dB, bandwidths of 50 kHz or greater, and a noise floor of $<50 \text{ dB}/\sqrt{\text{Hz}}$. Micromachined transducers offer the potential to provide an economically viable solution to these requirements.³

Microelectromechanical systems (MEMS)-based microphones have been developed utilizing piezoresistive, optical,

capacitive, and piezoelectric transduction techniques.^{4–11}

Most of these designs, however, have focused on audio applications that typically require a maximum SPL of $<120 \text{ dB}$, A-weighted noise floors of 25–35 dBA, and bandwidths of $<20 \text{ kHz}$. Microphones designed for aeroacoustic applications must instead meet more stringent requirements, such as described earlier for 1/4 scale testing.

Each transduction technique has unique advantages and disadvantages. Piezoelectric transduction offers the potential for a durable, high sensitivity, low noise device that requires no external power to operate. Capacitive sensors are capable of high sensitivity and dynamic range; however, their high impedance necessitates the use of an impedance buffer close to the sensor. Piezoresistive techniques generally have lower sensitivity and require temperature compensation for accurate measurement. Unlike capacitive techniques, they have a relatively low impedance. However, in contrast to piezoelectric transduction, both capacitive and piezoresistive transduction require some form of external power for operation. For aeroacoustic applications, a microphone with no external power requirement has a key advantage for widespread deployment. The main drawback to piezoelectrics is material integration and compatibility with standard micromachining. A number of authors have developed piezoelectric microphones mainly for audio applications and have fabricated them using a wide variety of piezoelectric materials.^{4–8,10,12}

Thin film piezoelectric materials include zinc oxide (ZnO), aluminum nitride (AlN), aromatic polyurea, polyvinylidene fluoride (PVDF), and lead zirconate titanate (PZT). Of the three most commonly used materials, PZT, AlN, and ZnO,

^{a)}Electronic mail: sheplak@ufl.edu

TABLE I. Typical material properties of select piezoelectric materials.

	d_{31} (pC/N)	d_{33} (pC/N)	$\epsilon_{33,r}$	E (GPa)	ρ (kg/m ³)
PZT ^{a,b}	-130	290	1300	96	7.7
AlN ^c	n/a	3.4	10.5	330	3.26
ZnO ^d	-4.7	12	12.7	210	5.6

^aReference 13.

^bValues are for bulk PZT.

^cReference 27.

^dReferences 20, 37, and 38.

PZT offers the largest piezoelectric modulus.¹³ They each offer different mechanical and electrical properties, as well as transduction capabilities as shown in Table I. However, these properties are also highly dependent upon the deposition method. In comparison to PZT, AlN and ZnO have relatively small piezoelectric coefficients and coupling factors. Additionally, they have significantly lower dielectric constants than PZT, leading to a two order of magnitude lower output capacitance. This lower output capacitance necessitates an integrated, on-chip amplifier to avoid effects due to parasitics that would otherwise result in substantial performance degradation.^{5,7,12} The use of PZT avoids the need for an integrated amplifier and simplifies the system design. In Table I, d_{31} is the piezoelectric coefficient relating electrical displacement in the “3” direction (z axis) to a mechanical stress in the “1” direction (radial) and is the relevant coefficient for bending-mode transducers, whereas d_{33} relates the electrical displacement in the “3” direction to a mechanical stress in that same direction. The coefficient, d_{33} , is thus the relevant coefficient for compression-mode transducers. Additionally, $\epsilon_{33,r}$, E , and ρ are the relative dielectric constant, elastic modulus, and mass density of the materials, respectively.

In addition, some of the materials are compatible with complementary metal oxide semiconductor (CMOS) technology while others are not, which has led to many differing approaches to fabricate piezoelectric microphones. The ease of integrating the piezoelectric material into a fabrication sequence is an important consideration in the development of a microphone. PZT, for instance, contains lead, barring it from all CMOS facilities, whereas AlN is fully CMOS compatible. ZnO and PVDF, while not fully CMOS compatible, are integrated more easily with conventional processing.^{6,14} Despite the various microphones that utilize piezoelectric transduction, to date, no piezoelectric microphones have been developed specifically for aeroacoustic applications.

The paper addresses this need by presenting the design, fabrication, packaging, and characterization of a piezoelectric microphone for aeroacoustic applications, extending upon previous conference papers that reported preliminary results.^{15,16} The details of the sensor design including lumped element modeling are presented in Sec. II. Then the fabrication and packaging are presented in Sec. III. Finally, the experimental setup and results are presented in Sec. IV.

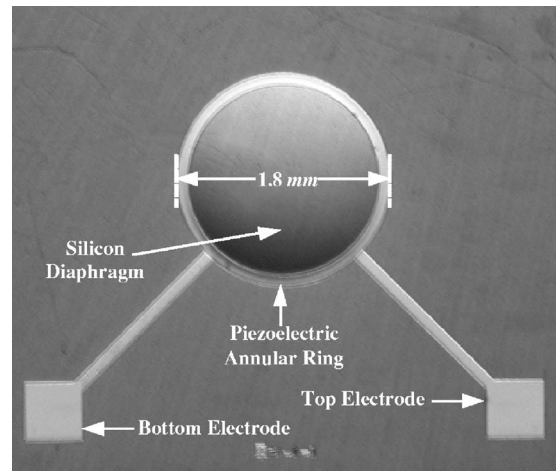


FIG. 1. An optical photograph of the piezoelectric microphone.

II. SENSOR DESIGN

A. Overview

This microphone was designed by combining a fundamental piezoelectric composite plate model^{17,18} with a lumped-element, electroacoustic model.^{18,19} Lumped-element modeling (LEM) was used to estimate the theoretical sensitivity and bandwidth of the device. Finite element modeling was then used to verify the LEM predictions prior to fabrication.

In micromachined piezoelectric devices, thin ferroelectric films are deposited on substrates comprised of a composite of silicon, metals, and oxides. Large in-plane residual stresses often result because of the fabrication-induced stresses, different thermal expansion coefficients between the films, and packaging effects.^{20,21} These in-plane stresses can dominate the mechanics of these devices and must be accounted for during design.²² For this microphone, the electromechanical design employed small-deflection plate theory for the multilayer piezoelectric diaphragm subjected to in-plane stresses, transverse pressure, and transverse electric field.¹⁸

The microphone structure consists of a 1.80-mm-diam, 3- μm -thick, silicon diaphragm with a thin ring of PZT placed near the boundary of the diaphragm (Figs. 1 and 2). The 1.85-mm-o.d., 95- μm -wide, piezoelectric annular ring was placed near the diaphragm edge to maximize sensitivity due to higher stress concentrations in this region. The placement of the ring near the boundary also eliminates the complexity of running electrodes across the diaphragm. The

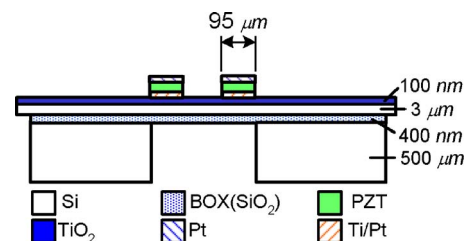


FIG. 2. (Color online) Cross-sectional schematic showing silicon diaphragm with circular piezoelectric composite ring (not to scale).

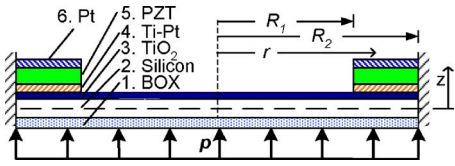


FIG. 3. (Color online) Cross-sectional schematic of the piezoelectric composite circular plate model.

267-nm-thick PZT was positioned between a 220 nm Ti/Pt layer and a 180 nm Pt layer that serve as bottom and top electrodes, respectively. The entire ring structure is separated from the silicon diaphragm by a 100-nm-thick TiO₂ layer that serves as a diffusion barrier for the PZT during processing. Inclusion of this diffusion barrier helps prevent cracking of the PZT that would otherwise arise from loss of lead via diffusion into the silicon during later processing steps.²³

The transversely isotropic composite plate model^{17,18} accounts for material properties of each of the thin-film layers, as shown in Fig. 3. These include the mechanical properties of the layers such as Young's modulus, E , density, ρ , and Poisson's ratio, ν , and electrical properties including the relative permittivity constant, ϵ_r , and piezoelectric coefficient, d_{31} , of the PZT. The model also accounts for the in-plane compressive stress, σ_0 , of each layer, which alters the overall deflection and subsequently affects the dynamic behavior. Furthermore, the geometry of the structure is accounted for via the inner PZT radius, R_1 , outer PZT radius, R_2 , and the various film thicknesses. The material properties used for the electromechanical model are given in Table II.

B. Lumped element model

A lumped element model was then developed to facilitate design of the complete microphone system, which involves electrical, mechanical, and acoustical energy domains. In order to use lumped element modeling, an underlying assumption must be met. The wavelength of the signal of interest must be significantly larger than any relevant length scales, so that little variation occurs in the signal energy distribution over space. As a result, the distributed energy is lumped at a chosen spatial location. This assumption is easily met for MEMS devices in many applications due to the typically small dimensions relative to the signal wavelengths of interest.

In LEM, the coupling between the various energy domains is realized via equivalent two-port models of the physical system.^{19,24,25} An equivalent circuit model is constructed by lumping the distributed energy storage and dissipation into ideal, generalized one-port circuit elements. In an

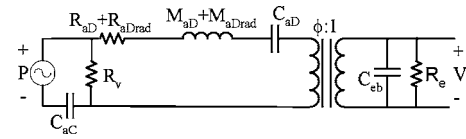


FIG. 4. Equivalent circuit representation of a lumped-element model of a piezoelectric microphone.

electroacoustic system, differential pressure and voltage are effort variables, while volumetric flow rate and current are flow variables.¹⁹ The lumped elements are represented in an equivalent electrical circuit by equating acoustic and mechanical elements to their electrical equivalents. In the electrical domain, a resistor represents dissipation of energy, while inductors and capacitors represent storage of kinetic and potential energy, respectively. Similarly, in the mechanical domain, kinetic and potential energy are stored in mass and compliance (inverse of stiffness), respectively. The same technique of lumping elements and equating them to their electrical equivalent is also extended to the acoustical energy domain, thereby permitting a single equivalent circuit that represents the dynamics of the entire system. In this paper, an impedance analogy is employed, in which elements that share a common effort are connected in parallel, while those sharing a common flow are connected in series.¹⁹

Figure 4 shows a schematic of an equivalent circuit representation of the lumped element model where R_{aDrad} , M_{aDrad} , C_{aD} , and M_{aD} are the radiation resistance and radiation mass of the diaphragm, and diaphragm compliance and mass, respectively, and are defined in Table III. Also, ϕ , C_{sb} , and R_e are the electroacoustic transduction coefficient, blocked electrical capacitance, and dielectric loss resistance, respectively. Finally, C_{aC} is the cavity compliance, R_{aD} is the diaphragm resistance representing structural losses, and R_v is the vent resistance.

The structure of the equivalent circuit is explained as follows. An acoustic pressure, P , drives the motion of the diaphragm, which in turn generates a voltage, V , across the piezoelectric thin film represented by C_{sb} . This represents a conversion from the acoustical to the electrical domain and is accounted for via a transformer possessing a turns ratio¹⁷

$$\phi = \frac{-d_A \left[\frac{\text{Pa}}{\text{V}} \right]}{C_{aD}}, \quad (1)$$

where d_A is the effective acoustic piezoelectric coefficient and C_{aD} is the diaphragm acoustic compliance when an electrical short circuit is imposed across the electrodes. The time-harmonic, two-port network equations describing the behav-

TABLE II. Material properties used for design.

	d_{31} (pC/N)	ν	ϵ_r	E (GPa)	ρ (kg/m ³)	σ_0 (MPa)
Silicon	—	0.3	—	169 ^a	2.5	0
Pt/Ti	—	0.38	—	170	21.44	—
TiO ₂	—	0.28	—	283	2.15	1809
PZT ^b	-50	0.3	500	30	7.5	176.0

^aReference 39.

^bValues are for thin films.

TABLE III. Lumped element modeling parameters.

Lumped element	Definition
d_A	$d_A = \frac{\Delta \hat{V}}{V} \Big _{p \rightarrow 0} = \int_0^{R_2} w(r) _{p \rightarrow 0} 2\pi r dr / V; \left[\frac{\text{m}^3}{\text{V}} \right]$
C_{aD}	$C_{aD} = \int_0^{R_2} 2\pi r w(r) dr / P \Big _{v=0} \left[\frac{\text{m}^3}{\text{Pa}} \right]$
M_{aD}	$M_{aD} = \frac{2\pi}{(\Delta \bar{V})^2} \int_0^{R_2} \rho_A (w(r))^2 r dr \Big _{v=0} \left[\frac{\text{kg}}{\text{m}^4} \right]$
	where ρ_A is the areal density of the piezoelectric composite plate given by
	$\rho_A = \int_{z_1}^{z_2} \rho_l dz; \left[\frac{\text{kg}}{\text{m}^2} \right]$
	and ρ_l is the density of the corresponding layer.
$M_{aD\text{rad}}^a$	$M_{aD\text{rad}} \approx \frac{8kR_2\rho_c}{3\pi\omega A_{\text{eff}}}; \left[\frac{\text{kg}}{\text{cm}^4} \right]$
	where A_{eff} is an effective diaphragm area to maintain continuity of volume velocity.
$R_{aD\text{rad}}^a$	$R_{aD\text{rad}} \approx \frac{(kR_2)^2\rho_c}{2A_{\text{eff}}}; \left[\frac{\text{kg}}{\text{m}^4\text{s}} \right]$
C_{eB}	$C_{eB} = C_{eF}(1 - \kappa^2) = \frac{\epsilon_r\epsilon_0\pi R_1^2}{h_p}(1 - \kappa^2); [F]$
	where ϵ_r is the relative dielectric constant of the piezoelectric material, ϵ_0 is the permittivity of free space, and $\kappa^2 = d_a^2/C_{eF}C_{aD}$ is the electroacoustic coupling factor.
C_{aC}	$C_{aC} = \frac{\hat{V}}{\rho c^2}; \left[\frac{\text{m}^3}{\text{Pa}} \right]$

^aReference 26.

ior of the linear reciprocal piezoelectric transducer are written in the acoustic domain as

$$\begin{bmatrix} I \\ Q \end{bmatrix} = \begin{bmatrix} j\omega C_{eF} & j\omega d_A \\ j\omega d_A & j\omega C_{aD} \end{bmatrix} \begin{bmatrix} V \\ P \end{bmatrix}, \quad (2)$$

where

$$Q = \int \frac{dw(r,t)}{dt} \cdot dS \quad (3)$$

is the volume velocity of the plate, dS is the normal projection of an infinitesimal element of area, $w(r,t)$ is the transverse deflection, and I is the electrical current.¹⁸

The radiation impedance is found by modeling the structure as a piston in an infinite baffle. For $kR_2 \ll 1$, the radiation resistance and mass are approximated as given in Table III.²⁶ The effects of structural damping are included by putting a resistor R_{aD} in series with $R_{aD\text{rad}}$. Additionally, the compliance of the cavity, C_{aC} , formed between the back of the diaphragm and the microphone package affects the dynamic response and is therefore included in the equivalent circuit. Finally, the resistance, R_v , represents viscous losses incurred in a vent channel that typically connects a microphone cavity to ambient pressure. In the actual device described in this paper, a vent channel was not specifically designed into the structure; however packaging-induced leakage between the cavity and atmospheric pressure provides the same effect.

The general expression for the sensitivity of the piezoelectric microphone is written as

$$\frac{V}{P} = \frac{\frac{R_v R_e \phi^2}{R_e s C_{eB} + 1}}{\frac{R_v}{s C_{aC}} + \left(R_v + \frac{1}{s C_{aC}} \right) \left(\frac{R_e \phi^2}{R_e s C_{eB} + 1} + s M_{aD} + \frac{1}{s C_{aD}} \right)}, \quad (4)$$

where $s = j\omega$. This expression includes the effects of cavity compliance and vent resistance in addition to the piezoelectric plate parameters. An important point to note from the equivalent circuit and the resulting sensitivity expression is the effect that the cavity compliance, C_{aC} , has on the overall sensitivity. If C_{aC} is much larger than the diaphragm compliance, C_{aD} , then the sensitivity is not noticeably affected. However, if the cavity compliance is the same order of magnitude or smaller than C_{aD} , then it has a noticeable stiffening effect because C_{aC} and C_{aD} are in series. In order to avoid this generally undesirable effect, the cavity must be made large enough to have a much greater compliance than the diaphragm.

In order to determine most of the lumped element parameters, the distributed deflection, $w(r)$, must be determined. To obtain $w(r)$, an analytical model was developed for the piezoelectric composite circular plate by Wang *et al.*¹⁸ Analytical modeling was accomplished by dividing the problem of Fig. 3 into two portions, an inner circular plate surrounded by an annular composite ring with matching boundary conditions at the interface. The boundary conditions consist of equal moments and forces at the interface as well as equal slope and radial displacement. After solving for the deflection in each region, the deflection equation for each region was then combined into a single equation.¹⁸ The resulting equation also predicts any initial static deflection arising from stress in the composite structure as well as the onset of buckling.

Using the lumped element model, a theoretical frequency response is computed and is shown in Fig. 5. The predicted resonant frequency for the designed microphone geometry, using the material properties as given in Table II, was computed to be approximately 63.8 kHz. The theoretical response assumed a perfectly sealed cavity below the diaphragm (i.e., $R_v \rightarrow \infty$). Any leakage from this cavity in the actual device leads to a low frequency roll-off in the re-

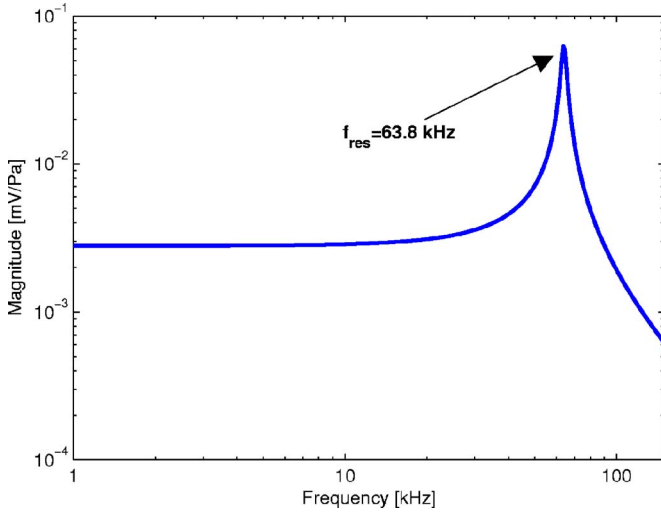


FIG. 5. (Color online) Magnitude of the theoretical frequency response based on the lumped element model for the piezoelectric microphone.

sponse. Furthermore, the additional low frequency roll-off effect of the dielectric loss resistance, R_e , is not visible in this plot as the cutoff frequency is around 1 Hz.

A noise model was also developed in order to understand and predict the noise floor of the device. Returning to the equivalent circuit of Fig. 4, a noise equivalent circuit is developed by adding a voltage or current noise source in series or parallel to each resistor as shown in Fig. 6. The choice of voltage or current noise source is based strictly on convenience in solving for the output voltage.

In order to compute the total output noise voltage power spectral density (PSD), S_{vOUT_e} , the individual contributions to the PSD are first found by considering the output voltage PSD due to a single source, while removing the others from the circuit (i.e., short-circuiting voltage sources, and open-circuiting current sources). The total output noise voltage PSD is then the summation of the individual PSDs, $S_{vOUT_e} = S_{vRv_e} + S_{vRe_e} + S_{vRD_e}$, where S_{vRv_e} is the output voltage noise PSD due to the microphone vent resistance R_v , S_{vRe_e} is the output voltage noise PSD due to the dielectric loss resistance, R_e , and S_{vRD_e} is the output voltage noise PSD due to the diaphragm and radiation resistances, $R_{aD} + R_{aDrad}$. The current noise PSD in the diaphragm, S_{iRv_D} , due to S_{iRv} is

$$S_{iRv_D} = \left| \frac{(Z_D + \phi^2 Z_e) // Z_{ac}}{Z_D + \phi^2 Z_e} \right|^2 S_{iRv}, \quad (5)$$

where

$$Z_e = \frac{1}{sC_{eb}} // R_e, \quad (6)$$

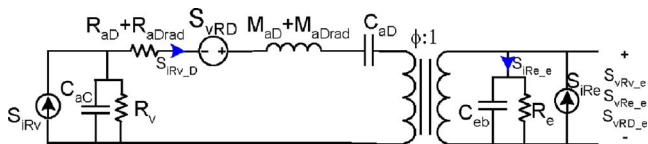


FIG. 6. (Color online) Equivalent circuit noise model for the piezoelectric microphone.

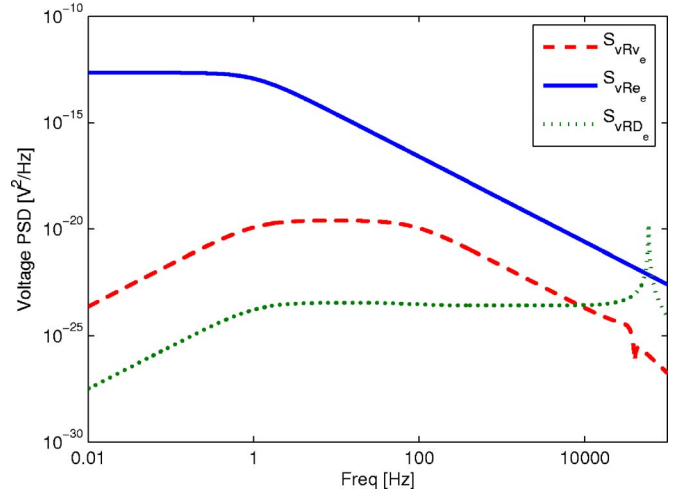


FIG. 7. (Color online) Output noise power spectral density resulting from the thermal noise of three resistive elements and shaped by the dynamics of the sensor.

$$Z_{ac} = \frac{1}{sC_{ac}} // R_v, \quad (7)$$

$$Z_D = R_{aD} + R_{aDrad} + s(M_{aD} + M_{aDrad}) + \frac{1}{sC_{aD}}, \quad (8)$$

and $S_{iRv} = 4k_B T / R_v$ where k_B is Boltzmann's constant and T is the ambient temperature (K). The output voltage PSD due to the vent resistance is then $S_{vRv_e} = \phi^2 S_{iRv_D} Z_e$. Similarly the output voltage PSD due to the dielectric loss resistance is

$$S_{vRe_e} = \left| \frac{Z_e // \frac{(Z_D + Z_{ac})}{\phi^2}}{Z_e} \right|^2 S_{iRe_e} Z_e^2, \quad (9)$$

where $S_{iRe} = 4k_B T / R_e$. Finally the contribution from the diaphragm losses is computed as

$$S_{vRD_e} = \left| \frac{\phi^2 Z_e}{\phi^2 Z_e + Z_D + Z_{ac}} \right|^2 \frac{S_{vRD}}{\phi^2}, \quad (10)$$

where $S_{vRD} = 4k_B T R_D$.

The resulting output noise voltage PSD is shown in Fig. 7. Note that over much of the frequency span, the noise is dominated by the contribution from the dielectric loss resistance, R_e , while the noise contribution from the diaphragm resistance is only significant near the diaphragm resonance. Meanwhile, the contribution from the vent resistance R_v is negligible over the frequency span.

III. FABRICATION AND PACKAGING

The devices were fabricated using a sol-gel PZT deposition process on a 4 in. silicon-on-insulator wafer combined with deep reactive ion etching (DRIE) for the release of the silicon diaphragm. Three facilities were used for the fabrication: Sandia National Laboratories in Albuquerque, NM for the pre-PZT and PZT deposition steps, the Army Research Laboratory in Adelphi, MD for the PZT patterning, and the University of Florida in Gainesville, FL for the DRIE release. The device fabrication is illustrated in Fig. 8, high-

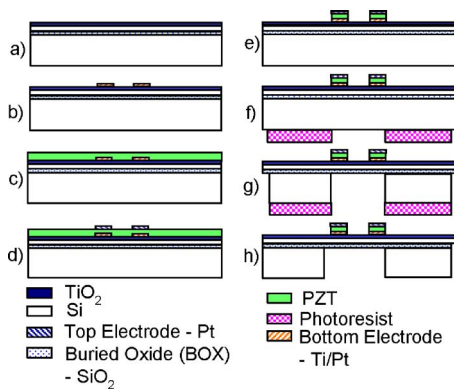


FIG. 8. (Color online) Concise process flow for the piezoelectric microphone. (a) Deposit 100 nm of Ti and oxidize to form TiO_2 . (b) Deposit and liftoff Ti/Pt (40/180 nm) to form bottom electrode. (c) Spin PZT 52/48 and pyrolyze 4 layers for 267 nm total. (d) Deposit and liftoff Pt (180 nm). (e) Wet etch PZT in 3:1:1 of $(\text{NH}_4)\text{HF}_2/\text{HCl}/\text{DI}$ water. (f) Spin and pattern thick PR on back. (g) DRIE to BOX. (h) Ash resist.

lighting the major steps in the process flow. The first step was to deposit a 100-nm-thick layer of Ti, which is then oxidized in a tube furnace to create a TiO_2 layer that serves as a diffusion barrier²³ as shown in Fig. 8(a). A liftoff process was then used to pattern the bottom electrodes with 40 nm of evaporated Ti followed by 180 nm of evaporated Pt as shown in Fig. 8(b). The Ti layer served as an adhesion layer in this step.

The next step was the deposition of a previously mixed 52/48 sol-gel solution of PZT as shown in Fig. 8(c). The solution was spin-cast at 2500 rpm for 30 s, then pyrolyzed at 350 °C in air for 2 min on a hot plate. Repeating the spin and pyrolysis step six times yielded a total PZT thickness of approximately 267 nm. The wafers were then furnace annealed at 650 °C for 30 min to achieve a perovskite phase of PZT for the piezoelectric properties. Following the PZT deposition, the top electrodes were deposited and patterned in a liftoff process similar to the bottom electrodes as shown in Fig. 8(d). The top electrodes also served as a hard etch mask for the PZT etch. The PZT was etched using a 3:1:1 solution of ammonium bifluoride $[(\text{NH}_4)\text{HF}_2]$, hydrochloric acid (HCl), and de-ionized water as shown in Fig. 8(e). The etchant leaves a residue that then must be removed with a dilute nitric acid and hydrogen peroxide solution. The next step was to deposit a thick photoresist (PR) on the back side, and pattern the back side release etch as in Fig. 8(f). Then the silicon wafer is back side etched using a DRIE process as shown in Fig. 8(g), stopping on the buried oxide (BOX) layer. Finally, the resist was ashed to yield the finished structure.

The completed chip was then flush mounted in a Lucite package shown in Fig. 9, using a two-part silver epoxy (Epotek™ H20E) to secure it. The electrical connections to the package leads were accomplished via short lengths of bare wire that were bonded using silver epoxy on the chip bond pads as well as to the copper pads on the package. The completed packaged microphone is shown in the photograph of Fig. 10.

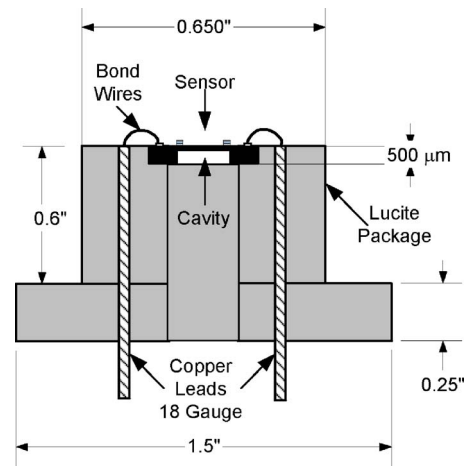


FIG. 9. (Color online) Lucite package with leads and flush mounted chip.

IV. EXPERIMENTAL SETUP AND RESULTS

After completing the fabrication and packaging, several experiments were used to characterize the electrical, acoustical, and transduction properties. From these experiments, the bandwidth, sensitivity, linearity, and noise floor were obtained.

Prior to full characterization of the device, the ferroelectric properties were measured by attaching the two electrodes to a Radiant Technologies Precision LC, Precision Materials Analyzer. This ferroelectric tester applies a bipolar triangular wave form, while it measures the resulting charge, from which it calculates the polarization of the material as a function of applied voltage. The bipolar wave form excitation to the ferroelectric material leads to the familiar hysteresis loop. For these measurements, the voltage was swept first between -0.5 and 0.5 V. Subsequent loops were swept to higher voltages in steps of 0.25 V up to a maximum of 5 V, corresponding to a maximum applied electric field of 187.95 kV/cm. For each loop, 101 points were measured and the total loop time was set to 10 ms. The resulting measured polarization reached a maximum of $16.9 \mu\text{C}/\text{cm}^2$ with a remanent polarization of $6.26 \mu\text{C}/\text{cm}^2$ for the 5 V sweep. It was not possible to extend the measured hysteresis curve into the fully saturated region as the maximum voltage that could be placed across the piezoelectric was approximately 5 V. Beyond this voltage, unexpected effects occurred in the material that produced a short circuit between the two electrodes. The nature of this short circuit is believed to be related to an overhanging top electrode resulting from problems arising during the fabrication. This was substantiated by

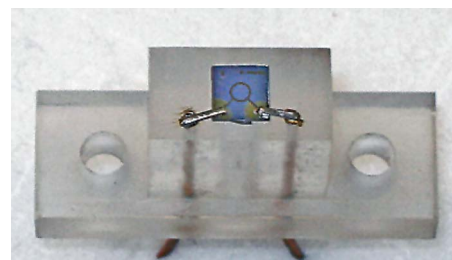


FIG. 10. (Color online) Optical photograph of packaged microphone.

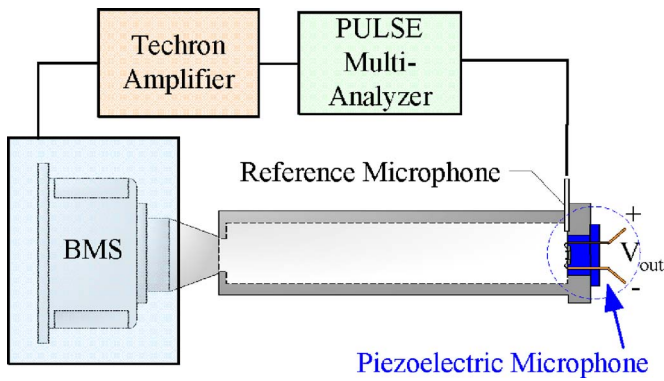


FIG. 11. (Color online) Experimental setup in a plane-wave tube.

resistance measurements performed on the failed devices. The measurements showed a resistance that was proportional to the perimeter of the ring, indicating it was due to an edge effect rather than a surface effect, which would have otherwise suggested dielectric breakdown in the piezoelectric material.

The blocked electrical properties were then measured before release of the device using a vector impedance meter (Agilent Model HP4294A) and are shown in Table IV. The value obtained for ϵ_r of 629 was found to be lower than the typically reported values for thin film PZT in the range of 900–1100,^{27–29} leading to a lower capacitance. However, the measured loss tangent was found to be 0.024, which matches well with previously published values ranging from 0.02 to 0.05 for thin film PZT.^{30–33}

Following the electrical characterization, the packaged device was mounted at the end of a 2.54 cm square, normal incidence plane wave tube (Fig. 11) to measure the frequency response and device linearity against a reference microphone [Brüel & Kjær (B&K) 1/8 in. Type 4138].³⁴ The excitation signal was provided by a B&K PULSE Multi-Analyzer System to a 50.8 mm coaxial compression driver (BMS Model 4590P) via a power amplifier (Techron Model 7540). The B&K PULSE was also used for data acquisition.

The frequency response was measured using periodic random noise at around 75 dB, re 20 μPa and 500 spectral averages over a bandwidth from 0 to 6.4 kHz with a 1 Hz bin width. A representative frequency response function of the sensor is shown in Fig. 12 in terms of magnitude and phase. The response is flat over the testable frequency range up to 6.7 kHz, except for a low frequency roll-off below approximately 100 Hz. This is most likely due to the vent resistance; however, the magnitude of this contribution is unknown, as the value for the vent resistance due to packaging leaks has not been measured. Representative linearity results, shown in Fig. 13, were taken at 1 kHz. For each measurement, 100 averages were taken of the rms output voltage from the measured power spectrum using a 1 Hz bin width with a rectangular window centered at 1 kHz to avoid

TABLE IV. Measured blocked electrical properties.

C_{eb} (nF)	R_e (M Ω)	ϵ_r	$\tan \delta$
10.8	14	629	0.001

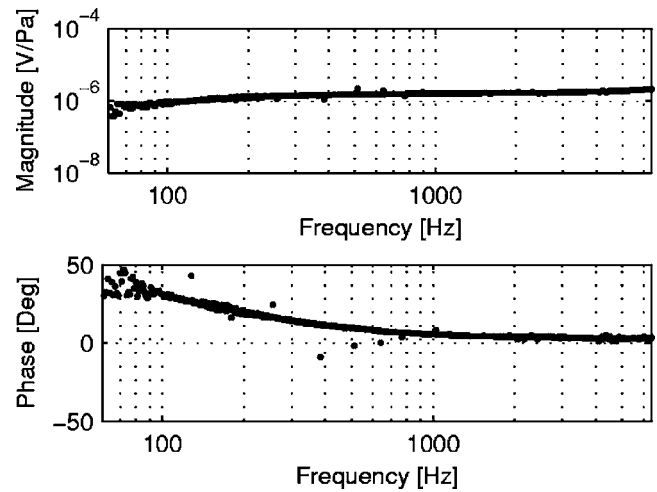


FIG. 12. Measured sensor frequency response.

spectral leakage. The device is linear ($R^2=0.9997$) up to (at least) the maximum testable sound pressure level of 169 dB. The maximum testable level is limited by the output capacity of the compression driver. The measured sensitivity is 1.66 $\mu\text{V}/\text{Pa}$.

To obtain noise floor measurements, the sensor was mounted in a triple Faraday cage along with an SRS 560 low-noise preamplifier, through which the sensor output was routed, to triple shield the sensor from potential electromagnetic interference.³⁵ For the measurement, the preamplifier was set to a gain setting of 1000, with a high pass filter cutoff of 0.03 Hz and a low pass filter cutoff of 1 MHz. The output of the preamplifier was then routed, via feed-through adapters, out of the Faraday cage to an SRS 785 dynamic signal analyzer. The raw input referred pressure noise spectrum of the sensor is shown in Fig. 14 to 4 kHz, along with the noise spectrum due to the experimental setup alone and the extracted microphone pressure noise floor determined from the difference between the sensor and setup noise PSD. Overlaid with these data is the theoretical computation for the noise floor based upon the model of Fig. 6. The noise spectrum of the experimental setup alone was obtained by shorting the inputs to the preamplifier and recording the resulting output voltage noise PSD. The measured noise spec-

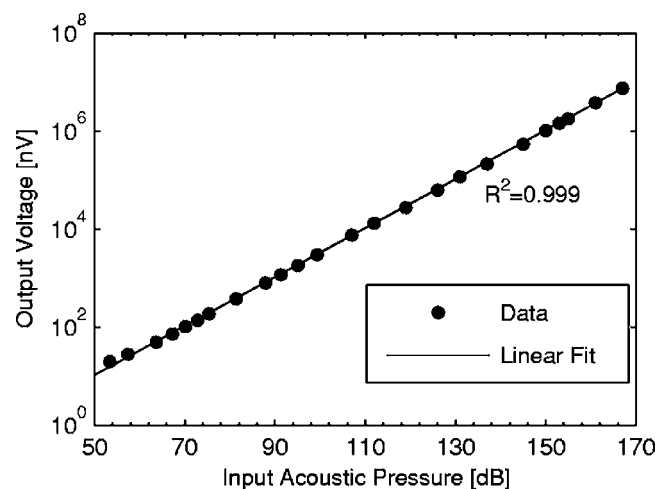


FIG. 13. Measured sensor linearity at 1 kHz.

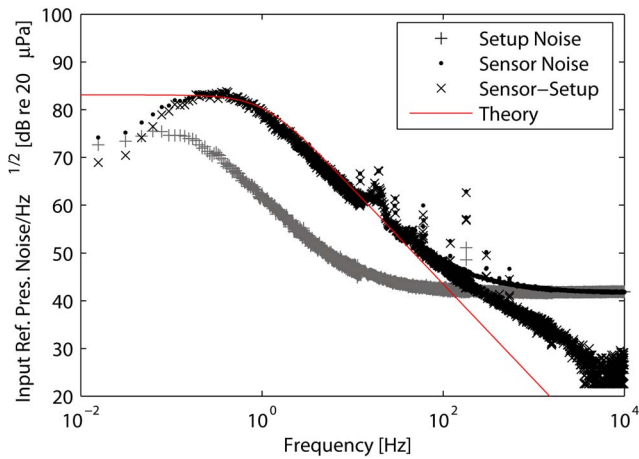


FIG. 14. (Color online) Measured noise floor spectra including sensor noise, setup noise, the differential noise, and the theoretical noise based on the piezoelectric microphone noise model.

trum agrees well with the theoretical model. For this device, there is a low pass, flat-band region with a cutoff frequency of 1.1 Hz. Here the noise is dominated by the dielectric loss resistance, R_e , of the sensor. Then, from 1.1 Hz up to approximately 225 Hz a $1/f$ rolloff is evident as the broadband noise is shaped by the spectral characteristics of the sensor. Above 225 Hz, the noise spectrum flattens out again as the sensor noise approaches the setup noise. For a 1 Hz bin width centered at 1 kHz, the output voltage with no acoustic signal applied is 2.02 nV due to just the sensor, which corresponds to an equivalent acoustic pressure of 2.69 mPa or 35.7 dB. It should be noted though that the system noise floor may be greater due to the contributions of output referred noise from additional electronics such as a preamplifier or A/D converter.

Because the device is a reciprocal transducer, the actuator frequency response is used to gain additional information about the sensor response, such as the bandwidth.¹⁹ In particular, the acoustically actuated frequency response could only be measured up to 6.7 kHz due to testing limitations of the impedance tube. The bandwidth is estimated via use of the electrically actuated frequency response as the piezoelectric microphone. In order to determine the actuator frequency response, scanning laser vibrometry was performed. The packaged device was mounted under an Olympus BX60 microscope, with a $5\times$ objective lens. The device was excited using a periodic random signal from a Polytec OFV 3001S laser scanning vibrometer. Velocity data were obtained over the surface of the composite diaphragm using the vibrometer, with a Polytec OFV-074 microscope adapter. Diaphragm displacement was obtained via integration of the measured velocity frequency response. The actuator frequency response, recorded at the diaphragm center, is shown in Fig. 15 in terms of magnitude and phase. The measured resonant frequency is 59.0 kHz and provides an estimate of the usable bandwidth as a microphone. The device characteristics are summarized in Table V.

Despite the improved performance, theoretical predictions do not match the experimental results. Specifically, the predicted sensitivity of $2.8 \mu\text{V}$ was roughly 70% larger than

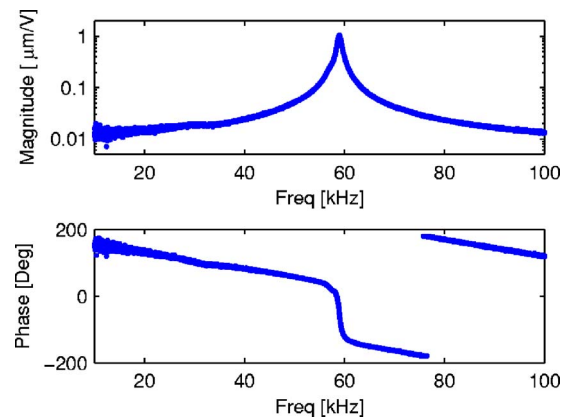


FIG. 15. (Color online) Measured actuator frequency response.

measured, while the predicted resonant frequency of 63.8 kHz was 8% larger than measured. These discrepancies are due to a combination of factors. First, there tends to be large uncertainties in the thin-film material properties. More important, the devices were statically deflected on the order of the diaphragm thickness due to large in-plane compressive stresses from the TiO_2 . These large static deflections render the small-deflection assumption in the composite plate theory invalid. More specifically, the model was developed assuming a small residual stress as compared to the piezoelectrically excited stress. The assumption allowed for a decoupling of transverse and radial displacements. In practice, however, our devices exhibited an unexpectedly significant amount of residual stress, thereby violating this assumption. An extension of the composite plate model that allows for the transverse and radial displacements to remain coupled is therefore necessary to more accurately predict the performance of this device.

V. CONCLUSIONS

A MEMS-based piezoelectric microphone was developed for aeroacoustic measurements and consists of a circular composite annular ring of PZT for electromechanical

TABLE V. Summary of experimental results and comparison to previous work.

Researcher	Material	Sensitivity (mV/Pa)	f_{res} (kHz)	Noise floor (dB)
Present Work	PZT	0.00166	59.0	35.7 ^a , 95.3 ^b
Kim <i>et al.</i> ^c	ZnO	1	16	50 ^b
Ried <i>et al.</i> ^d	ZnO	0.92	18	57 ^b
Lee <i>et al.</i> ^e	ZnO	30	0.89	NR
Royer <i>et al.</i> ^f	ZnO	0.25	10	66.02 ^g
Schellin <i>et al.</i> ^h	polyurea	0.004–0.030	NR	NR
Ko <i>et al.</i> ⁱ	ZnO	0.51 ^j	7.3	NR

^aAt 1 kHz with 1 Hz bin.

^bA-weighted.

^cReference 6.

^dReference 7.

^eReference 4.

^fReference 5.

^gMethod not reported.

^hReference 8.

ⁱReference 10.

^jAt Resonance.

transduction. An electroacoustic equivalent circuit model was used to design the device. Experimental characterization yielded a sensitivity of $1.66 \mu\text{V}/\text{Pa}$, with a linear dynamic range from (35.7–169 dB, re $20 \mu\text{Pa}$). The resonant frequency was estimated to be 59.0 kHz by measuring the actuator frequency response. This is the first reported piezoelectric microphone developed specifically for aeroacoustic applications. Based on the experimental results reported earlier, this microphone meets and exceeds the general specifications required for 1/4 scale wind tunnel testing and certification.² Furthermore, the self-powered nature of the transduction facilitates the construction of microphone arrays for airframe noise source localization, by eliminating the need to distribute power throughout the array.

Future work includes the extension of the composite plate model to large deflections. These models will then be validated against the current design and then subsequently used to synthesize an optimal design.³⁶ A new package design is necessary to provide a vent channel to the microphone for a controlled low frequency roll-off. Further characterization of the microphone is needed in order to determine variability from device to device, and to investigate reliability issues such as fatigue and stability of the piezoelectric modulus.

ACKNOWLEDGMENTS

Financial support for this research was provided by Sandia National Laboratories and monitored by Kent Pfeifer. The authors are also grateful for significant fabrication assistance provided by Stephanie Jones of Sandia National Laboratories.

- ¹“Aeronautics and space, noise standards: Aircraft type and airworthiness certification,” Technical Report Part 36, Title 14 US Code of Federal Regulations, 2004.
- ²*Aeroacoustic Measurements*, edited by T. Mueller (Springer, Berlin, 2002), p. 313.
- ³D. P. Arnold, T. Nishida, L. N. Cattafesta, and M. Sheplak, “A directional acoustic array using silicon micromachined piezoresistive microphones,” *J. Acoust. Soc. Am.* **113**, 289–298 (2003).
- ⁴S. Lee, R. Ried, and R. White, “Piezoelectric cantilever microphone and microspeaker,” *J. Microelectromech. Syst.* **5**, 238–242 (1996).
- ⁵M. Royer, J. Holmen, M. Wurm, and O. Aadland, “ZnO on Si integrated acoustic sensor,” *Sens. Actuators, A* **4**, 357–362 (1983).
- ⁶E. Kim, J. Kim, and R. Muller, “Improved IC-compatible piezoelectric microphone and CMOS process,” in *IEEE International Conference on Solid-State Sensors and Actuators*, San Francisco, 1991, pp. 270–273.
- ⁷R. Ried, E. Kim, D. Hong, and R. Muller, “Piezoelectric microphone with on-chip CMOS circuits,” *J. Microelectromech. Syst.* **2**, 111–120 (1993).
- ⁸R. Schellin, G. Hess, W. Kuehnel, G. Sessler, and E. Fukada, “Silicon subminiature microphones with organic piezoelectric layers: Fabrication and acoustical behaviour,” *IEEE Trans. Electr. Insul.* **27**, 867–871 (1992).
- ⁹P. Scheeper, A. van der Donk, W. Olthuis, and P. Bergveld, “A review of silicon microphones,” *Sens. Actuators, A* **44**, 1–11 (1994).
- ¹⁰S. Ko, Y. Kim, S. Lee, S. Choi, and S. Kim, “Micromachined piezoelectric membrane acoustic device,” *Sens. Actuators, A* **103**, 130–134 (2003).
- ¹¹W. Lee, N. A. Hall, Z. Zhou, and F. L. Degertekin, “Fabrication and characterization of a micromachined acoustic sensor with integrated optical readout,” *IEEE J. Sel. Top. Quantum Electron.* **10**, 643–651 (2004).
- ¹²E. Kim, R. Muller, and P. Gray, “Integrated microphone with CMOS circuits on a single chip,” in *IEEE International Electron Devices Meeting*, Washington, DC, 1989, pp. 880–883.
- ¹³*Piezoelectric Materials in Devices*, edited by N. Setter (EPFL Swiss Federal Institute of Technology, Lausanne, 2002).

- ¹⁴M. J. Vellekoop and C. C. G. Visser, “An integrated SAW voltage sensor,” in *Proc.-IEEE Ultrason. Symp.* **1**, 575–578 (1988).
- ¹⁵S. Horowitz, T. Nishida, L. Cattafesta, and M. Sheplak, “Design and characterization of a micromachined piezoelectric microphone,” in *11th AIAA/CEAS Aeroacoustics Conference*, Monterey, CA, 2005.
- ¹⁶S. Horowitz, T. Nishida, L. Cattafesta, and M. Sheplak, “A micromachined piezoelectric microphone for aeroacoustics applications,” in *Solid-State Sensor, Actuator and Microsystem Workshop*, Hilton Head Island, SC, 2006.
- ¹⁷S. Prasad, B. V. Sankar, L. N. Cattafesta, S. Horowitz, Q. Gallas, and M. Sheplak, “Two-port electroacoustic model of an axisymmetric piezoelectric composite plate,” *AIAA J.* **4**, 2311–2318 (2006).
- ¹⁸G. Wang, B. Sankar, L. Cattafesta, and M. Sheplak, “Analysis of a composite piezoelectric circular plate with initial stresses for MEMS,” in *International Mechanical Engineering Congress and Exposition*, ASME, New Orleans, 2002.
- ¹⁹S. Senturia, *Microsystems Design* (Kluwer Academic, Boston, 2001), Chap. 5.
- ²⁰D. L. DeVoe and A. P. Pisano, “Surface micromachined piezoelectric accelerometers (PiXLs),” *J. Microelectromech. Syst.* **10**, 180–186 (2001).
- ²¹E. Zakar, M. Dubey, R. Polcawich, B. Piekarski, R. Piekarz, J. Conrad, and R. Widuta, “Study of PZT film stress in multilayer structures for MEMS devices,” in *Mater. Res. Soc. Symp. Proc.* **605**, 287–292 (2000).
- ²²M. Sheplak and J. Dugundji, “Large deflections of clamped circular plates and transitions to membrane behavior,” *J. Appl. Mech.* **65**, 107–115 (1998).
- ²³B. Tuttle *et al.*, “Ferroelectric thin film microstructure development and related property enhancement,” *Ferroelectrics* **151**, 11–20 (1994).
- ²⁴M. Rossi, *Acoustics and Electroacoustics* (Artech House, Norwood, MA, 1988).
- ²⁵J. Merhaut, *Theory of Electroacoustics* (McGraw-Hill, New York, 1981).
- ²⁶D. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000), pp. 153–156.
- ²⁷M. A. Dubois and P. Muralt, “Measurement of the effective transverse piezoelectric coefficient e_{31} of AlN and Pb(Zr,Ti)O₃ thin films,” *Sens. Actuators, A* **77**, 106–112 (1999).
- ²⁸Y. Yee, H. J. Nam, S. H. Lee, J. U. Bu, and J. W. Lee, “PZT actuated micromirror for fine-tracking mechanism of high-density optical data storage,” *Sens. Actuators, A* **89**, 166–173 (2001).
- ²⁹D. F. Bahr, J. C. Merlino, P. Banerjee, C. M. Yip, and A. Bandyopadhyay, “Reliability and properties of PZT thin films for MEMS applications,” in *Mater. Res. Soc. Symp. Proc.* **546**, 153–158 (1999).
- ³⁰C. Lee, T. Itoh, R. Maeda, and T. Suga, “Characterization of micromachined piezoelectric pzt force sensors for dynamic scanning force microscopy,” *Rev. Sci. Instrum.* **68**, 2091–2100 (1997).
- ³¹B. Lee, K. Lee, and B. Ju, “Compositional dependence of the properties of ferroelectric $Pb(zrTi)O_3$ thin film capacitors deposited on single-layered Pt electrode barriers,” *Ceram. Int.* **30**, 1543–1546 (2004).
- ³²T. Kanda, T. Morita, M. Kurosawa, and T. Higuchi, “Flat type touch probe sensor using pzt thin film vibrator,” *Sens. Actuators, A* **83**, 67–75 (2000).
- ³³K. Kunz, P. Enoksson, and G. Stemme, “Highly sensitive triaxial silicon accelerometer with integrated pzt thin film detectors,” *Sens. Actuators, A* **92**, 156–160 (2001).
- ³⁴T. Schultz, L. Cattafesta, and M. Sheplak, “Modal decomposition method for acoustic impedance testing in square ducts,” *J. Acoust. Soc. Am.* **120**, 3750–3758 (2006).
- ³⁵R. Dieme, G. Bosman, M. Sheplak, and T. Nishida, “Source of excess noise in silicon piezoresistive microphones,” *J. Acoust. Soc. Am.* **119**, 2710–2720 (2006).
- ³⁶M. Papila, R. Haftka, T. Nishida, and M. Sheplak, “Piezoresistive microphone design pareto optimization: Tradeoff between sensitivity and noise floor,” *J. Microelectromech. Syst.* **15**, 1632–1643 (2006).
- ³⁷D. L. DeVoe, “Piezoelectric thin film micromechanical beam resonators,” *Sens. Actuators, A* **88**, 263–272 (2001).
- ³⁸C. H. Han and E. S. Kim, “Micromachined piezoelectric ultrasonic transducers based on parylene diaphragm in silicon substrate,” in *IEEE International Ultrasonic Symposium*, San Juan, Puerto Rico, 2000, pp. 919–923.
- ³⁹N. Ono, K. Kitamura, K. Nakajima, and Y. Shimanuki, “Measurement of young’s modulus silicon single crystal at high temperature and its dependency on boron concentration using the flexural vibration method,” *Jpn. J. Appl. Phys., Part 1* **39**, 368–371 (2000).

Reconstruction of Rayleigh–Lamb dispersion spectrum based on noise obtained from an air-jet forcing

Eric Larose,^{a)} Philippe Roux, and Michel Campillo

Lab. de Géophysique Interne et Tectonophysique, Université J. Fourier and CNRS, BP53, 38041 Grenoble, France.

(Received 7 June 2007; revised 29 August 2007; accepted 26 September 2007)

The time-domain cross correlation of incoherent and random noise recorded by a series of passive sensors contains the impulse response of the medium between these sensors. By using noise generated by a can of compressed air sprayed on the surface of a plexiglass plate, we are able to reconstruct not only the time of flight but the whole wave forms between the sensors. From the reconstruction of the direct A_0 and S_0 waves, we derive the dispersion curves of the flexural waves, thus estimating the mechanical properties of the material without a conventional electromechanical source. The dense array of receivers employed here allow a precise frequency-wavenumber study of flexural waves, along with a thorough evaluation of the rate of convergence of the correlation with respect to the record length, the frequency, and the distance between the receivers. The reconstruction of the actual amplitude and attenuation of the impulse response is also addressed in this paper. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2799913]

PACS number(s): 43.40.Dx, 43.35.Cg, 43.50.Yw, 43.40.Ph [JJM]

Pages: 3437–3444

I. INTRODUCTION

Elastic waves at kilohertz and megahertz frequencies are widely used to evaluate the mechanical properties of structures, material, and tissues. In plates and shells, the elastic wave equation admits specific propagation modes, denoted Lamb waves (Viktorov, 1967; Royer and Dieulesaint, 2000), which are related to the traction free condition on both sides of the medium. Depending on the purpose of the experiment, different measurement configurations have been proposed. From the dispersion of Lamb waves, obtained from pitch-catch measures repeated for different ranges, one has access to the velocities of bulk waves and to the thickness of the plate (see for instance Gao *et al.*, 2003). Other pulse-echo (or impact-echo) techniques have been proposed to assess the mechanical properties of the plate, based on the simple and multiple reflection of bulk waves within the plate (Krautkrämer and Krautkrämer, 1990) or on the resonance of high order Lamb modes (Clorennec *et al.*, 2007). Some studies also concern the dynamic evaluation of fatigue and/or crack growth [see for instance, Ihn and Chang (2004) and Ing and Fink (1996)]. In this last application, several impulse responses of the medium are acquired at different dates and are eventually compared to each others to monitor the medium.

All these techniques require the use of controlled sources and receivers. In the following, they are referred to as “active” experiments. Another idea has undergone a large development after the seminal experiments of Weaver and Lobkis (2001) [see for instance the review of Wapenaar and Fokkema (2006), Weaver and Lobkis (2006), or Larose (2006)]. By cross correlating the incoherent noise recorded by two passive sensors, Weaver and Lobkis demonstrated that one could reconstruct the impulse response of the medium as if a source was placed at one sensor. This noise

correlation technique [also referred to as “passive imaging” or “seismic interferometry” [Schuster *et al.* (2004)]] requires the use of synchronized sensors, and has the advantage of eliminating any controlled source.

In Sec. II of our paper, we compare the dispersion curves of flexural waves (Lamb waves in the low frequency regime) in a plexiglass plate obtained by an active (pitch-catch) experiment to the ones obtained in a passive experiment. During the passive acquisition, we deployed synchronized receivers only, and used a random noise source: a 1 mm thick high pressure air jet produced by a can of compressed air (McBride and Hutchison, 1976). The experiment conducted in a plexiglass plate show at the same time: dispersive (A_0) and nondispersive (S_0) waves, reverberations, and absorption. As the earth’s crust has similar properties for seismic waves, we believe that a plexiglass plate at kHz frequencies is a good candidate to build small scale seismic analogous experiments. The question of reconstructing not only the phase, but also the amplitude of the wave, is addressed at the end of part II. This is of interest in seismology, where the role of absorption and attenuation in the correlation has not yet been subject to experimental investigation. In Sec. III, we analyze the rate of convergence of the correlations to the real impulse response. The role of the record length, the central frequency, and the distance between the receivers is investigated.

II. EXPERIMENTAL SETUP AND DISPERSION CURVES

To study actively and passively the dispersion of flexural waves, we built a laboratory experiment using a 1.5 m × 1.5 m large, 0.6 cm thick plexiglass plate. The plate is laid on an open steel frame that supports the edges of the plate but leaves free the upper and lower sides. The traction-free

^{a)}Electronic mail: eric.larose@ujf-grenoble.fr

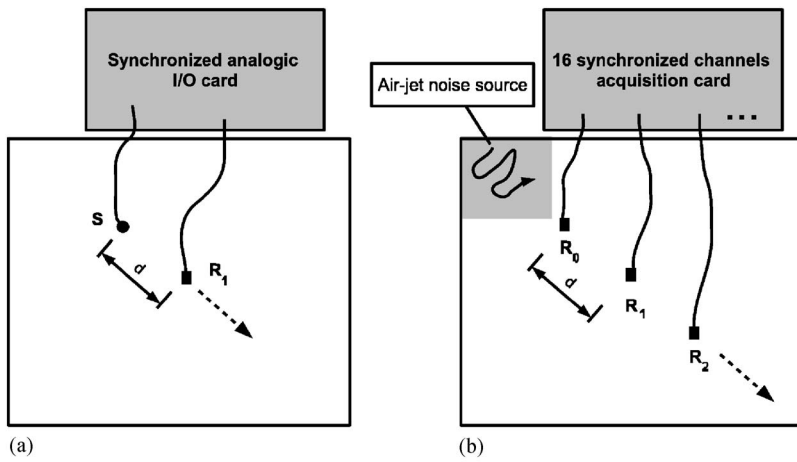


FIG. 1. Experimental setup. (a) In the active experiment (pitch-catch configuration), a broadband piezoelectric source S emits a chirp that is sensed by a vertical accelerometer R placed at a distance d . (b) In the passive experiment, a turbulent jet produced by a can of compressed air generates a white random noise recorded simultaneously at all sensors R_i . The jet is randomly sprayed over the area in gray.

condition is therefore achieved on both horizontal sides. The resulting dispersion relation that connects the pulsation ω and the wave-vector k reads:

$$\frac{\omega^4}{c_s^4} = 4k^2 q^2 \left(1 - \frac{p \tan(ph/2 + \gamma)}{q \tan(qh/2 + \gamma)} \right), \quad (1)$$

where $p^2 = \omega^2/c_p^2 - k^2$ and $q^2 = \omega^2/c_s^2 - k^2$; c_s (respectively, c_p) is the shear (respectively, compressional) velocity, and h is the thickness of the plate. The parameter γ equals 0 for symmetric (S) modes, and $\pi/2$ for antisymmetric (A) modes. In the low frequency regime, only two modes are solutions: they are labeled A_0 and S_0 .

For both acquisitions we used broad-band miniature (3 mm radius) accelerometers (Ref. No. 4518 from Bruel&Kjaer). They show a flat response in the 20 Hz–70 kHz frequency range. We fixed our accelerometers on our plate using a hot chemical glue (phenyl-salicylic acid) that solidifies with cooling (below 43°C) (see Fig. 1).

A. Active experiment

In the active experiment, a source S (a piezoelectric polymer) is in a corner of the plate, approximately 30 cm from each side and is emitting a 3 s chirp $s(t)$ in a linear range of frequencies f from 1 kHz to 60 kHz. A receiver R is initially placed at $d=1$ cm away from the source toward the center of the plate on a straight graduated line. After each acquisition, we moved the receiver a centimeter away from the source down the diagonal; we repeated this operation 100 times to cover 100 cm of the plate. The dynamic impulse response $h_d(t)$ of the plate is reconstructed for each distance d by correlating the record $r_d(t)$ by the source chirp:

$$h_d(t) = r_d(t) \times s(t). \quad (2)$$

As an example, impulse responses obtained for three different distances $d=10, 40, 80$ cm are displayed in Fig. 2. The dispersive A_0 mode is dominating the record (a), but the nondispersive S_0 wave (b) and reverberations from the edges of the plate (c) are also visible. The transit time in the plate is of the order of a few milliseconds. The absorption time of the plate that strongly depends on the frequency is of the order of a few hundreds of milliseconds. The signal-to-noise in the experiment do not allow to record more than a few

tens of millisecond. Records are therefore dominated by ballistic waves, but waves reverberated from the plate boundaries are also visible.

From the set of 100 impulse responses $h_d(t)$, a spatiotemporal Fourier (f - k) transform was applied. The resulting dispersion curves are displayed in Fig. 3, showing the dispersive A_0 mode and the weaker nondispersive S_0 mode. Theoretical dispersion curves are numerically obtained after Eq. (1) (dots), they perfectly fit the experimental data for $c_p=3130$ m/s and $c_s=1310$ m/s. The dispersion curves are widely used to evaluate the mechanical properties of the plate, including thickness, presence of flaws, etc. Nevertheless, the source–receiver configuration is sometimes at a disadvantage. As mentioned in the introduction, instead of using the conventional source–receiver configuration, it is possible to take advantage of random elastic noise to reconstruct the impulse response $h_d(t)$ between two sensors. This is performed in the following section.

B. Passive experiment

In the passive experiment, we removed the source and placed 16 receivers separated by 7 cm from each other on the

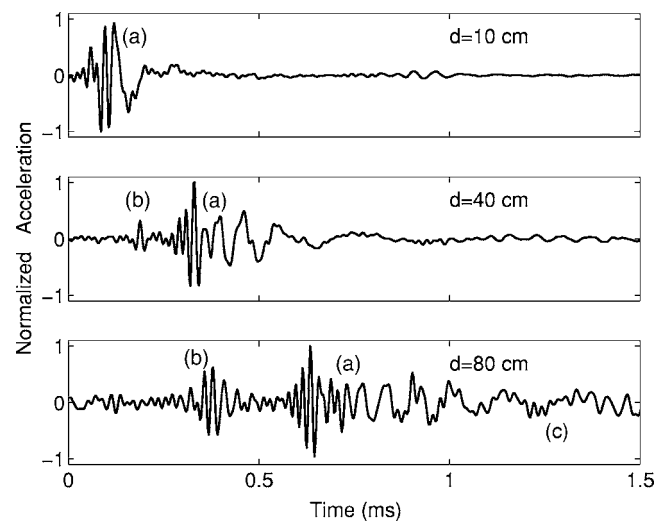


FIG. 2. Source–receiver impulse response $h_d(t)$ for different distances d . (a) Dispersive anti-symmetric A_0 mode propagating at velocities approximately ranging from 150 to 900 m/s. (b) Nondispersive S_0 mode propagating at ≈ 2400 m/s. (c) Reflexions from surrounding edges.

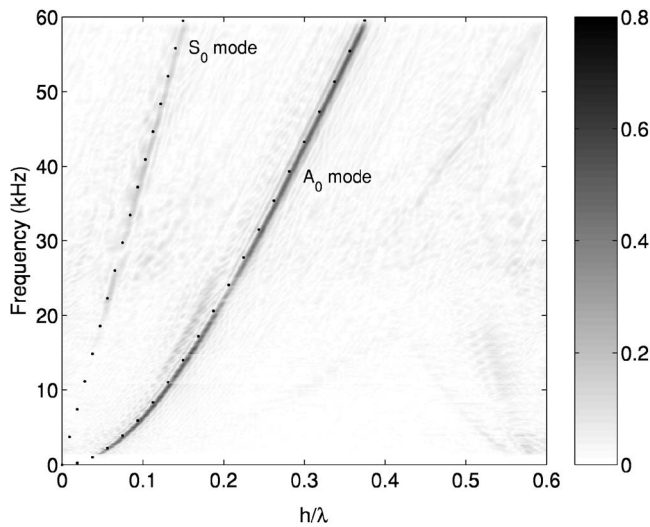


FIG. 3. f - k transform of the source-receiver impulse responses. X-axis: dimensionless wave number (h is the slab thickness). Dots are theoretical solutions of Eq. (1).

graduated line. At the former position of the source, we placed the accelerometer R_0 that was kept fixed all through the experiment. As proposed by Sabra *et al.* (2007), we used the noise generated by a turbulent flow. As a noise source, we employed a dry air blower. Note that, contrary to Sabra's work, our source was easy-to-handle. During the experiment, it was randomly moved to cover a 30 cm \times 30 cm large area located between the corner of the plate and the former active source. Note that the precise knowledge of the noise source is not necessary in the passive experiment. We also deployed an array of 16 receivers that allow for precise frequency-wave number analysis, and worked at higher frequencies, meaning a much thinner spatial resolution. We focused our attention on the A_0 and S_0 direct arrivals. As we want to reconstruct direct waves, we chose to spray only at one end of the array of receivers [end-fire lobes described by Roux *et al.* (2004) or coherent zones described by Larose (2006)]. Nevertheless, spraying elsewhere gives the same waveforms but requires much more data, thus much longer acquisitions (see Sec. III).

The noise in the plate was created by spraying continuously for approximately $T=10$ s. The 16 receivers recorded synchronously this 10 s noise sequence, each record is labeled after its distance d from R_0 . Then receivers R_1 - R_{15} were translated 1 cm down the graduated line and the 10 s acquisition performed again. This operation was repeated seven times to cover the 100 cm of the diagonal with a pitch of 1 cm. We ended with a set of 106 $r_{d=0.105}(t)$ records. The time domain cross-correlation between the receiver R_0 and the other receivers is processed afterward:

$$C_i(\tau) = \int_0^T r_0(t)r_i(t + \tau)dt. \quad (3)$$

As mentioned in the introduction, this cross correlation $C_d(\tau)$ is very similar to the impulse response $h_d(t)$ obtained in the pitch-catch experiment. Strictly speaking, the impulse response equals the time derivative of the cross correlation convolved by the source spectrum (which is almost flat

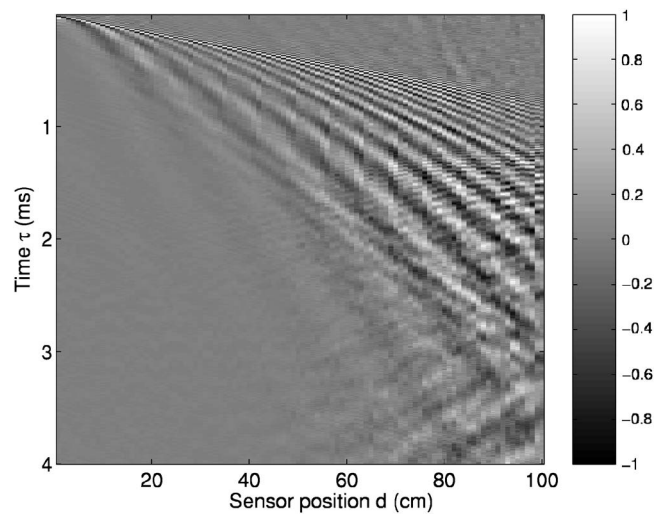


FIG. 4. Passively reconstructed impulse responses (linear color bar, normalized amplitude).

here). Nevertheless, the time-derivative operation was not performed here as it does not change the spatiotemporal Fourier transform of the data. The series of 100 correlations is displayed in Fig. 4 as a time-distance plot. The dispersive A_0 mode is clearly visible, including reverberations at the edges. The symmetric S_0 mode is very weak and almost invisible in Fig. 4. In the passive experiment, the size of the noise source is ≈ 1 mm large, which is quasi-punctual compared to the A_0 and S_0 wavelength. In that case, and taking into account the large difference in phase velocities, the energy ratio is clearly in favor of the A_0 mode, the S_0 mode is much weaker (and additionally less converged). In the active experiment the source was 6 mm large, thus exciting the S_0 mode more strongly and the S_0 mode was more visible in Fig. 3.

Like in the active experiment, we computed the spatiotemporal Fourier transform of the set of 100 traces $C_d(\tau)$. The resulting dispersion curves are plotted in Fig. 5. The A_0

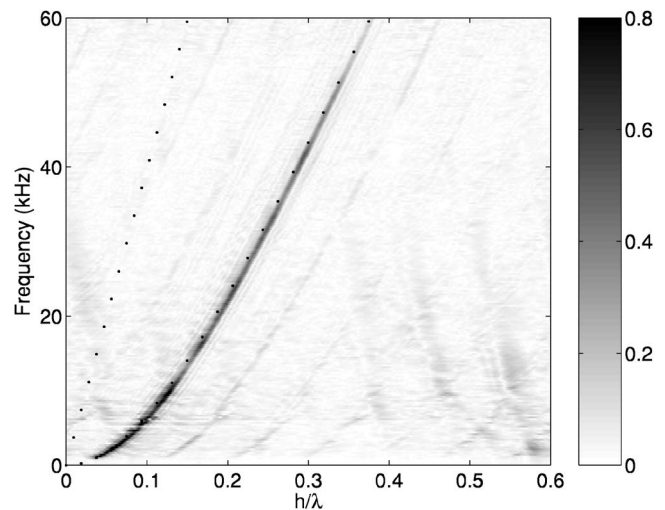


FIG. 5. Spatiotemporal (f - k) Fourier transform applied to the passive data (linear color bar, normalized amplitude). The dispersion curve of the A_0 mode is perfectly matching the active experiment. S_0 mode is very weak, probably feebly excited by the point-like noise source. The dots show the theoretical dispersion curves as obtained in the active experiment.

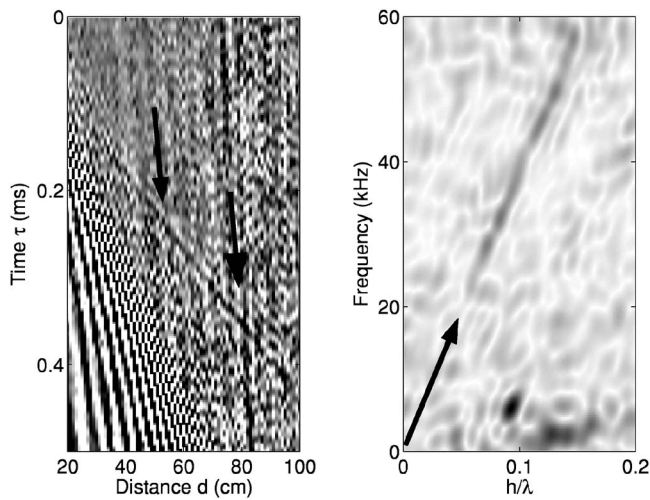


FIG. 6. (Left) Zoom into the early part of the passive data (saturated color bar). The nondispersive S_0 wave is pointed by the arrows. (Right) f - k transform applied to the passive data after muting the A_0 mode arrivals (set to zero). The S_0 non-dispersive mode is now visible.

mode is clearly reconstructed, the agreement between the active, passive, and theoretical dispersion curves is perfect. Ghosts are also visible in Fig. 5, and are due to (1) small errors in positioning the sensors and (2) the presence of reflections from the edges. As to Fig. 4, the S_0 mode is not

visible. We therefore zoomed into the early part of the correlations [Fig. 6 (left)] and muted all the data but the weak nondispersive arrival. We then computed again the f - k transform of the data, as plotted in Fig. 6 (right). The straight line that emerges from the noise is exactly the S_0 dispersion curve, thus demonstrating that the weak arrival in Fig. 6 (left) is indeed the S_0 mode. It is important to emphasize that the weakness of the S_0 mode is not a limitation of our correlation technique, but is connected to the noise generation.

C. Reconstruction of the amplitudes

Since now, the reconstruction of the phase (the arrival time of the wave) of the Green function (GF) by correlation of noise has been widely studied. Feeble attention was paid to the information carried by the amplitude of the reconstructed GF. We can report the first experiments of [Weaver and Lobkis \(2001\)](#) who noted that both the phase and amplitude of the signals were passively reconstructed, and also [Larose et al. \(2006\)](#) who used this amplitude information to study weak localization without a source. In Sec. II B, we have seen that the phase information in the passive experiment perfectly matches the active data, what about the amplitude of the wave?

To see if the amplitude decay is recovered using correlations, we plot in Fig. 7 the amplitude decay obtained with

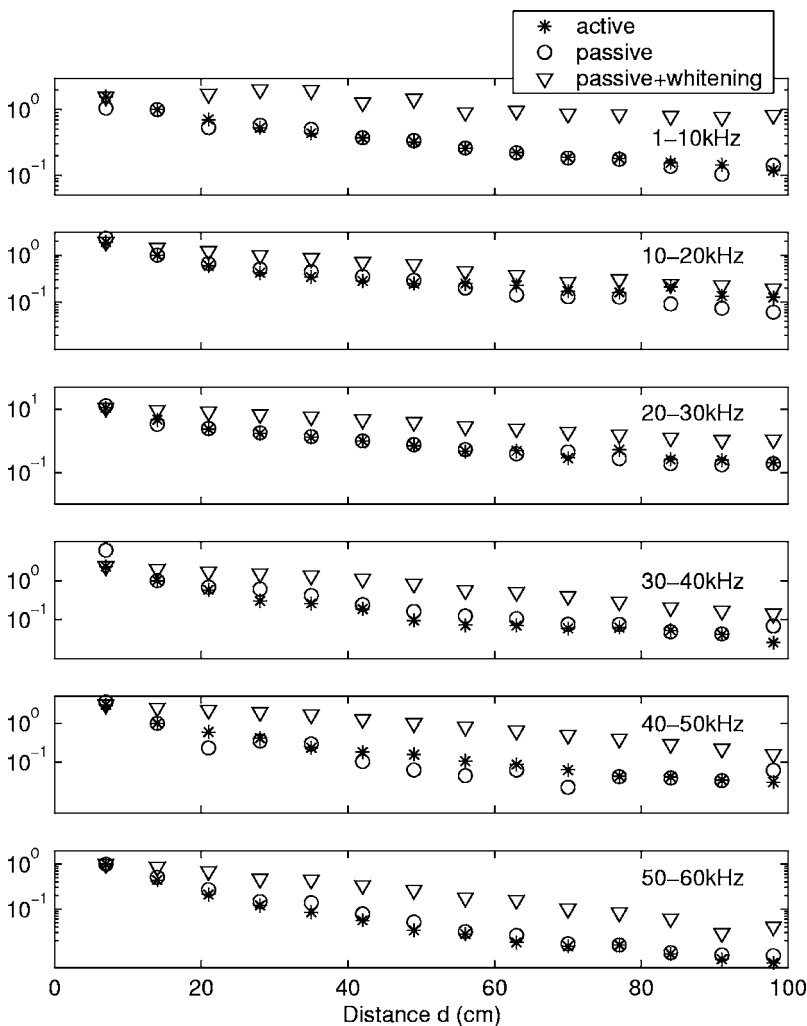


FIG. 7. Amplitude decay obtained in an active experiment (stars) and by correlation (circles) without additional processing, and with whitening (triangles). Attenuation includes geometrical spreading and absorption.

an active experiment (stars). In a homogeneous and lossless two-dimensional (2D) plate, one would expect a decrease of the amplitude as $\propto 1/d^{0.5}$. Because of dispersion and the bandwidth used here (10 kHz), the exponent is slightly greater ($\propto 1/d^{0.6}$). To be more rigorous, absorption adds another exponential decay. From the active experiments, we found that the absorption time was about 260 ms at 60 kHz and 2000 ms at 1 kHz [similar to Safaeinili *et al.* (1996)].

In Fig. 7, we also plot the maximum of the amplitude of the correlations versus the distance d between the sensors (circles). Active and passive amplitudes are comparable: the geometrical spreading of the wave is well recovered. Nevertheless, small discrepancies are visible and are due the imperfect or irregular coupling between the sensors and the plate. Estimating the absorption in the plate from the passive data is therefore more speculative with this experimental setup and will be subject to further investigations.

Nevertheless, in several applications (Roux *et al.*, 2004; Shapiro *et al.*, 2005; Sabra *et al.*, 2005a), an operation is performed prior to the correlation in order to balance the contribution of all frequencies. The whitening operation, performed here in the 1–60 kHz frequency range, reads:

$$\tilde{r}_d(t) = IFFT\left(\frac{FFT(r_d(t))}{|FFT(r_d(t))|}\right), \quad (4)$$

then the correlation $\tilde{r}_0(t) \times \tilde{r}_d(t)$ is evaluated. We compared the active amplitude with the amplitude of the correlations using whitened data (triangles in Fig. 7). This time, though a decay in the correlation is visible, we do not recover the decay of the actual GF. Note that 1-bit processing [correlate the sign of the records, see for instance Larose *et al.* (2004)] give similar disappointing results. As whitening (or 1-bit) does not preserve the amplitude of the records, we stress that correlations will unlikely give the actual attenuation. This is of practical interest for applications like seismology. To summarize, in order to recover the amplitude by correlating the incoherent noise, we suggest that each record be filtered in a narrow frequency band, and then correlated without additional processing. The amplitude decay would therefore be estimated for each frequency, and attenuation derived. On the other hand, the whitening (or 1-bit) processing is valuable when the only phase information is targeted (tomography for instance).

III. RATE OF CONVERGENCE OF CORRELATIONS

Several authors (Campillo and Paul, 2003; Derode *et al.*, 2003) have experimentally noticed that the correlations $C_d(\tau)$ do not only contain the impulse response (the “signal”) between the receivers but also show residual fluctuations that blur the weaker part of the signal. As soon as residual fluctuations are negligible compared to the reference impulse response, we consider that the correlations have converged. The convergence of correlations has been theoretically studied by Weaver and Lobkis (2005) and Sabra *et al.* (2005b). We propose here to check the validity of their predictions with our experimental data. Of course, theoretical works apply to a perfectly diffuse wavefield. Here, the noise is generated over a delimited area. Nevertheless, as developed in

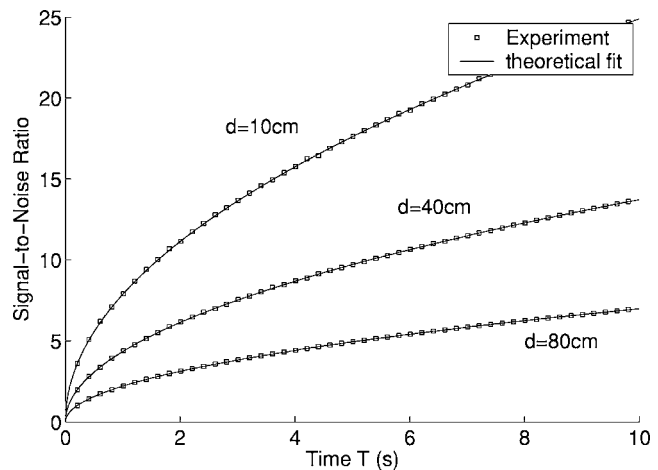


FIG. 8. Signal-to-noise ratio of the correlation versus the duration T of the records. The theoretical fit for each distance d is of the form $\alpha\sqrt{T}$ with α the fit parameter.

Sec. III D, reverberations compensate the uneven distribution of source, and we believe that previous theories should apply to our experiment.

A. Convergence with time T

To evaluate the degree of convergence of the correlations toward the real impulse response between the sensors, we split each noise record into $N=1000$ subrecords. Each sub-record lasts $\delta t=10$ ms. Correlations are processed for each sub-record and each couple of receivers R_0-R_d . We then compute the amplitude of the residual fluctuations:

$$\sigma_d(N, \tau) = \sqrt{\frac{\langle C_d(\tau)^2 \rangle - \langle C_d(\tau) \rangle^2}{N-1}}, \quad (5)$$

where $\langle \cdot \rangle$ represents an average over N subrecords. The fluctuations were found to vary weakly with time τ , we therefore average $\sigma_d(T=N\delta t, \tau)$ over τ to get a more robust estimation of the fluctuations $\sigma_d(T)$. To evaluate the “signal-to-noise ratio” (SNR) of the correlation, the maximum of each correlation is divided by the residual fluctuations σ_d . Experimental results are plotted in Fig. 8 for three distances d and for the whole 1–60 kHz frequency range. As noticed theoretically and experimentally by several authors (Weaver and Lobkis, 2005; Sabra *et al.*, 2005b), the SNR is found to increase like:

$$\text{SNR} = \alpha\sqrt{T}, \quad (6)$$

where α is the fit coefficient. This means that the longer the records, the better the reconstruction. The SNR increases as the square root of the amount of “information grain” contained in the records. This amount corresponds to the quantity $T\Delta f$ (where Δf is the frequency bandwidth) and represents the number of uncorrelated pieces of information transported by the waves (Derode *et al.*, 1999; Larose *et al.*, 2004). In our experiment, a few seconds was enough to get the direct arrival of the A_0 mode.

Another point visible in Fig. 8 is the decrease of the coefficient α with the distance d between the pair of receivers. This means that the passive reconstruction of the impulse

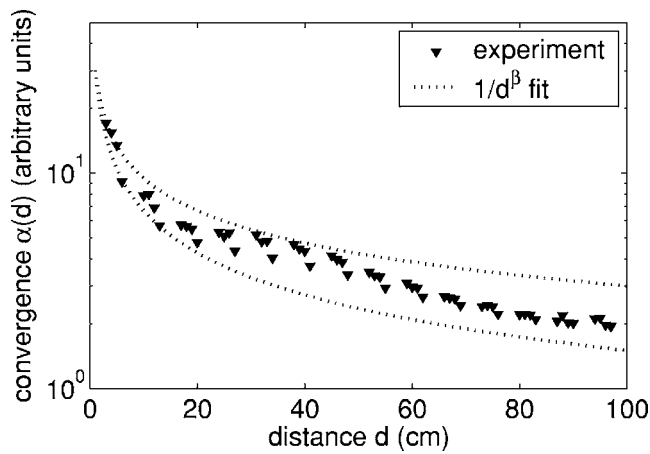


FIG. 9. Fit coefficient α of the SNR is plotted versus the distance d between the two correlated receivers.

response is harder when the receivers are far apart. This was noticed in seismology for instance (Shapiro and Campillo, 2004; Paul *et al.*, 2005). A more precise estimation of the convergence with the distance d is developed in the following subsection.

B. Convergence with distance d

To study the dependence of the convergence rate on the distance d between the passive sensors, each coefficient α was evaluated for each SNR curves (as in Fig. 8) in the 1–60 kHz frequency range. The resulting $\alpha(d)$ is plotted in Fig. 9 and is delimited by two theoretical curves of the form:

$$\alpha \propto 1/d^\beta, \quad (7)$$

with $0.5 < \beta < 0.65$. At first sight, our result is similar to previous theoretical works that predicted a SNR evolution as $\sqrt{1/d}$ in a 2D space. Nevertheless, the theory was developed for a nondispersive medium, where the wave spreading factor is $\sqrt{1/d}$. In our dispersive plate, the exponent is slightly greater (≈ 0.6 , see Sec. II A). The SNR decay seems to be driven by the decay of the actual Green function. Note that the role of the absorption cannot be evaluated in our experiment, and will be subject to further investigations.

C. Convergence with frequency f

It is now commonly acknowledged that high frequencies are harder to reconstruct by cross correlation than low frequencies. We now propose to quantitatively evaluate the role of the central frequency f of the record in the rate of convergence of the correlations. To that end, we performed several whitening operations in different frequency bands (every 5 kHz, ranging from 1 to 60 kHz). Fluctuations were then evaluated for each frequency band and each subrecord. The experimental SNR was fitted by $\alpha(f)\sqrt{T}$, with α the fit parameter plotted in Fig. 10. The prediction that $\text{SNR} \propto 1/\sqrt{f}$ is well recovered.

D. Convergence with the noise sources position

Several theoretical approaches have been proposed to model the reconstruction of the GF between passive sensors.

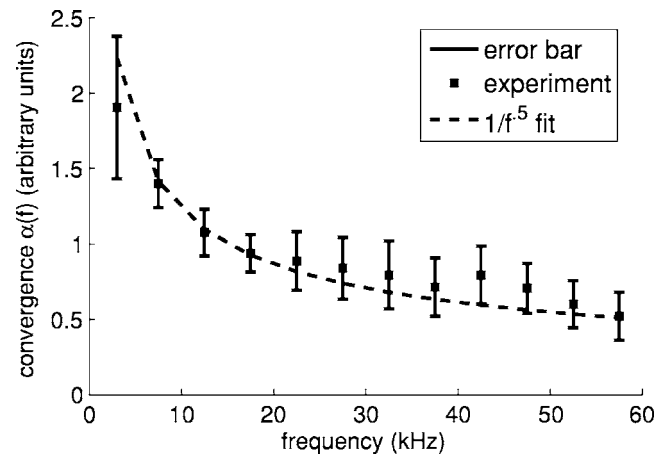


FIG. 10. Fit coefficient α of the SNR is plotted versus the central frequency f of the records.

Some invoke a perfect distribution of sources surrounding the medium to image (Derode *et al.*, 2003; Wapenaar, 2004). Others refer to a perfect diffuse wavefield, obtained either with scatterers or with thermal noise (Weaver and Lobkis, 2001). Our experiment does not correspond to the first case, since the noise is generated in limited area. Does it correspond to the diffuse case? Because ballistic waves strongly dominate the records, this is not guaranteed and should be checked thoroughly!

To answer this point, we chose to spray over an area in the side of the array of receivers (see the dark gray area in Fig. 11). If only ballistic waves were present, the direct waves along the receiver line would not be reconstructed. This was not the case here: after an integration over a 35 s long record, we could reconstruct the same wave forms as in the active experiment. The only difference is the rate of convergence of the correlations: it was found to be much slower (see SNR in Fig. 11 in log–log scale). This proves the elastic wavefield to be (at least slightly) diffuse in the plate employed in the experiment. To be more precise, the ballistic

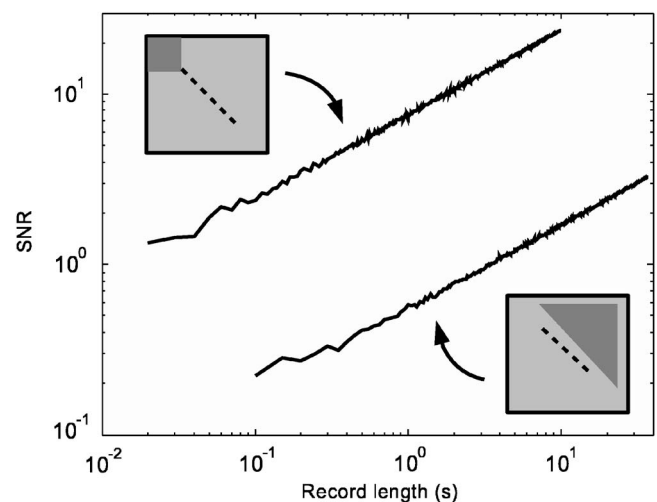


FIG. 11. Rate of convergence of the correlation (SNR) with the record time T . The air-jet is randomly sprayed over the dark gray areas, dotted line is the array of receivers. (Upper left plot) Sources located on the side of the array of receivers. (Lower right plot) Sources located in others places.

and the diffusive regime coexist in our plate. Right after a noise source occurs, the propagating wave packet is in the ballistic regime. It turns into the diffusive regime after a dozen of reverberation. The first regime dominates the record; the second is weaker but not completely negligible. This last experiment is also a clear evidence that even with imperfectly diffusive wavefield, correlations of incoherent noise yield the Green function. This is of importance in applications like seismology where the medium is not always perfectly diffusive and the noise sources are not perfectly distributed. Figure 11 also demonstrates that when the reconstruction of the direct waves is expected, spraying over an area located in the direction of the array requires much less data than any other configuration.

IV. CONCLUSION AND DISCUSSION

In this article, we presented two experiments using Lamb waves detected by a series of accelerometers fixed on a plate. In the first experiment, a conventional source–receiver configuration was employed to construct the dispersion curves of the A_0 and S_0 flexural waves. In the second experiment, we only used receivers: we recorded the noise generated by a can of compressed air whose turbulent stream generates random excitations. Contrarily to an active experiment, the advantage here was that we neither needed to know the position of the noise source, nor to employ any electronics for the emission. By correlating the noise records, we reconstructed the impulse response between the sensors, and recovered the same dispersion curves as in the active experiment (except for the low frequency S_0 mode that was not excited by the thin turbulent jet).

The amplitude decay of the actual A_0 mode with distance was successfully retrieved when correlating the raw records, but was not retrieved using either whitened or 1-bit records. This is of importance for applications where passive reconstruction of the attenuation of the media is envisioned. To summarize, whitening (or 1 bit) the data is helpful to reconstruct the phase of the GF (for imaging applications or f – k analysis), but should not be employed to reconstruct the amplitude.

Using our dense array of receivers, we also carefully studied the fluctuations of the correlations, which are connected to the rate of convergence of the correlations to the real impulse response. The SNR (or correlation-to-fluctuation ratio) obtained in our experiments is in agreement with previous studies (Weaver and Lobkis, 2005; Sabra *et al.*, 2005b), both qualitatively and quantitatively. We showed that the reconstruction of the impulse response is better when: (1) we use long record (more data); (2) we employ close receivers; and (3) we work at low frequency. The resulting SNR curves were best fitted by

$$\text{SNR} = B \sqrt{\frac{T\Delta fc}{d^{1.1}f}}. \quad (8)$$

The parameter B , as defined by Weaver and Lobkis (2005) was found to be $[0.25 \pm 0.02]$ (we take $c = 350$ m/s as a mean value for A_0 mode), which is much greater than seismic experiments (Shapiro and Campillo, 2004). This is in part due

by the very uneven distribution of sources in our experiment, resulting in an anisotropic wave flux that accelerates the convergence of the correlation to the direct A_0 wave. This parameter B was found to be ten times smaller, whereas spraying on one side of the array of receiver. Concerning the dependence with distance d , the decay matches the amplitude decay of the actual Green function. We emphasize that it is not a trivial term when dispersive waves are considered. A proper prediction for the SNR should account for the dispersion and the absorption.

Practically speaking, in the 1–60 kHz frequency range, we obtained very good reconstruction of the impulse responses ($\text{SNR} \geq 5$) for distances up to 100 cm with less than 10 s of noise. This estimation is fast enough to be repeated continuously and provides a route for the passive monitoring of structures and materials. At seismic frequencies, correlations of ambient noise are already used to monitor active volcanoes (Sabra *et al.*, 2006; Sens-Schönfelder and Wegler, 2006; Brenguier *et al.*, private communication). As suggested by Sabra *et al.* (2007) in a recent paper, we foresee similar applications in the field of on board passive structural health monitoring, in noisy environments such as aircrafts or ground vehicles, and also on civil engineering structures as bridges and buildings.

ACKNOWLEDGMENTS

The authors are thankful to D. Anache-Menier, P. Gouedard, Stefan Hiemer, L. Margerin, Adam Naylor, B. Van Tiggelen, and R. L. Weaver for fruitful discussions and experimental help. Jean-Paul Masson is strongly acknowledged for technical help in designing the experiment. This work was funded by a French ANR "chaire d'excellence 2005" grant.

- Campillo, M., and Paul, A. (2003). "Long range correlations in the diffuse seismic coda," *Science* **299**, 547–549.
- Clorenec, D., Prada, C., and Royer, D. (2007). "Local and noncontact measurements of bulk acoustic wave velocities in thin isotropic plates and shells using zero group velocity lamb modes," *J. Appl. Phys.* **101**, 034908.
- Derode, A., Larose, E., Tanter, M., de Rosny, J., Tourin, A., Campillo, M., and Fink, M. (2003). "Recovering the Green's function from field-field correlations in an open scattering medium," *J. Acoust. Soc. Am.* **113**, 2973–2976.
- Derode, A., Tourin, A., and Fink, M. (1999). "Ultrasonic pulse compression with one-bit time reversal through multiple scattering," *J. Appl. Phys.* **85**, 6343–6352.
- Gao, W., Glorieux, C., and Thoen, J. (2003). "Laser ultrasonic study of lamb waves: Determination of the thickness and velocities of a thin plate," *Int. J. Eng. Sci.* **41**, 219–228.
- Ihn, J.-B., and Chang, F.-K. (2004). "Detection and monitoring of hidden fatigue crack growth using a built-in piezoelectric sensor/actuator network. I. Diagnostics," *Smart Mater. Struct.* **13**, 609–620.
- Ing, R. K., and Fink, M. (1996). "Time recompression of dispersive Lamb waves using a time-reversal mirror—Application to flaw detection in thin plates," *Proc.-IEEE Ultrason. Symp.* **11**, 659–663.
- Krautkrämer, J., and Krautkrämer, K. (1990). *Ultrasonic Testing of Materials* (Springer, New York).
- Larose, E. (2006). "Mesocopies of ultrasound and seismic waves: Application to passive imaging," *Ann. Phys. (Paris)* **31(3)**, 1–126.
- Larose, E., Derode, A., Campillo, M., and Fink, M. (2004). "Imaging from one-bit correlations of wideband diffuse wave fields," *J. Appl. Phys.* **95**, 8393–8399.
- Larose, E., Margerin, L., Derode, A., van Tiggelen, B., Campillo, M., Shapiro, N., Paul, A., Stehly, L., and Tanter, M. (2006). "Correlation of random wave fields: An interdisciplinary review," *Geophysics* **71**, S111–S121.

- McBride, S., and Hutchison, T. (1976). "Helium gas jet spectral calibration of acoustic emission transducers and systems," *Can. J. Phys.* **54**, 1824–1830.
- Paul, A., Campillo, M., Margerin, L., Larose, E., and Derode, A. (2005). "Empirical synthesis of time-asymmetrical Green functions from the correlation of coda waves," *J. Geophys. Res.* **110**, B08302.
- Roux, P., Kuperman, W., and the NPAL Group (2004). "Extracting coherent wave fronts from acoustic ambient noise in the ocean," *J. Acoust. Soc. Am.* **116**, 1995–2003.
- Royer, D., and Dieulesaint, E. (2000). *Elastic Waves in Solids* (Springer, New York).
- Sabra, K. G., Gerstoft, P., Roux, P., Kuperman, W. A., and Fehler, M. C. (2005a). "Surface wave tomography from microseisms in Southern California," *Geophys. Res. Lett.* **32**, L14311.
- Sabra, K. G., Roux, P., Gerstoft, P., Kuperman, W. A., and Fehler, M. C. (2006). "Extracting coherent coda arrivals from cross-correlations of long period seismic waves during the Mount St. Helens 2004 eruption," *Geophys. Res. Lett.* **33**, L06313.
- Sabra, K. G., Roux, P., and Kuperman, W. A. (2005b). "Emergence rate of the time-domain Green's function from the ambient noise cross-correlation function," *J. Acoust. Soc. Am.* **118**, 3524–3531.
- Sabra, K. G., Winkel, E. S., Bourgoyne, D. A., Elbing, B. R., Ceccio, S. L., Perlin, M., and Dowling, D. R. (2007). "Using cross correlations of turbulent flow-induced ambient vibrations to estimate the structural impulse response application to structural health monitoring," *J. Acoust. Soc. Am.* **121**, 1987–1995.
- Safaenili, A., Lobkis, O. I., and Chimenti, D. E. (1996). "Air-coupled ultrasonic estimation of viscoelastic stiffnesses in plates," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **43**, 1171–1180.
- Schuster, G. T., Yu, J., Sheng, J., and Rickett, J. (2004). "Interferometric/daylight seismic imaging," *Geophys. J. Int.* **157**, 838–852.
- Sens-Schönfelder, C., and Wegler, U. C. (2006). "Passive image interferometry and seasonal variations of seismic velocities at Merapi volcano, Indonesia," *Geophys. Res. Lett.* **33**, L21302.
- Shapiro, N. M., and Campillo, M. (2004). "Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise," *Geophys. Res. Lett.* **31**, 614.
- Shapiro, N. M., Campillo, M., Stehly, L., and Ritzwoller, M. H. (2005). "High resolution surface wave tomography from ambient seismic noise," *Science* **307**, 1615–1618.
- Viktorov, I. A. (1967). *Rayleigh and Lamb waves: Physical Theory and Applications* (Plenum, New York).
- Wapenaar, K. (2004). "Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation," *Phys. Rev. Lett.* **93**, 254301.
- Wapenaar, K., and Fokkema, J. (2006). "Greens function representations for seismic interferometry," *Geophysics* **71**, SI33–SI46.
- Weaver, R. L., and Lobkis, O. I. (2001). "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," *Phys. Rev. Lett.* **87**, 134301.
- Weaver, R. L., and Lobkis, O. I. (2005). "Fluctuations in diffuse field-field correlations and the emergence of the Green's function in open systems," *J. Acoust. Soc. Am.* **117**, 3432–3439.
- Weaver, R. L., and Lobkis, O. I. (2006). "Diffuse fields in ultrasonics and seismology," *Geophysics* **71**, S15–S19.

Response variance prediction for uncertain vibro-acoustic systems using a hybrid deterministic-statistical method

R. S. Langley^{a)}

Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

V. Cotoni

ESI US R&D Inc., 12555 High Bluff Drive, Suite 250, San Diego, California 92130, USA

(Received 19 December 2006; revised 4 September 2007; accepted 20 September 2007)

Imperfections during the manufacturing process can cause significant variations in the noise and vibration levels exhibited by nominally identical structures. Any response calculations employed during the design process should ideally take account of these uncertainties and predict the expected range in performance. Recently a hybrid method has been developed to predict the ensemble average response of a built-up system by combining a deterministic model of parts of the system with a statistical model of other components [Shorter, P. J., and Langley, R. S. (2005) *J. Sound. Vib.*, **288**, 669–700]. In this paper the method is extended to predict the ensemble variance of the response. Expressions are derived for the variance of the vibrational energies in the statistical components, and for the variance of the cross spectrum of the response of the deterministic components, which augment the mean values of these quantities predicted by the original theory. The method employs a nonparametric model of uncertainty, in the sense that the statistical components are taken to carry diffuse wave fields, and this obviates the requirement for a detailed description of the system uncertainties. The method is validated by application to a range of coupled plate structures, and good agreement with detailed Monte Carlo simulations is found.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799499]

PACS number(s): 43.40.Qi, 43.20.Ks, 43.40.Dx, 43.20.Bi [DF]

Pages: 3445–3463

I. INTRODUCTION

The prediction of the dynamic response of a structure that has uncertain or random properties is of great interest across a range of industries. For example, manufacturing imperfections arising in aerospace and automotive production lines can lead to significant performance differences between nominally identical vehicles, as exemplified by the data presented by [Kompella and Bernhard \(1993\)](#). Any analysis of the system response in this type of situation must allow for the presence of uncertainty, and example target prediction quantities include the response of a “typical” system, the mean response, the higher order statistics of the response, and nonstatistical measures of the response uncertainty, such as bounds or fuzzy distributions. The prediction of such quantities can be extremely difficult, and when the concern is with high frequency vibration and noise then the modeling problem is made even more challenging by the fact that the short wavelength of the system deformation can require the use of very many degrees of freedom. The present work is concerned with the development of an efficient method for predicting the mean and variance of the frequency response functions of a complex uncertain system across a broad frequency range.

There are two basic approaches to modeling an uncertain system, the first of which employs a *parametric* model

of the underlying uncertainties. In this case the system is modeled by using a standard deterministic technique, such as the finite element method, and the physical properties of the system (for example, the material properties) are then considered to be uncertain. The aim of the approach is to propagate a description of the physical uncertainties through the dynamic model, and thus obtain a description of the response uncertainty. The uncertainty description may take the form of a statistical distribution, a fuzzy distribution, or a bounded interval; examples of these approaches have been reported respectively by [Schuëller \(2006\)](#), [Moens and Vandepitte \(2005\)](#), and [Elishakoff \(1995\)](#). The propagation of uncertainty through the model may be done analytically or by using Monte Carlo simulations, but in either case the computational difficulty is significant for complex systems with many degrees of freedom. It can also be extremely difficult to obtain the data required to describe the uncertainty of the physical parameters, or even to identify which physical parameters should be considered to be uncertain. It might be imagined that such modeling problems can only increase if high frequency excitation is of concern, and this is certainly true of the parametric approach. However, if the system is sufficiently complex and random then there is the possibility of employing the second, more efficient, analysis approach: *nonparametric* modeling of the system uncertainties, as described in what follows.

The nonparametric approach can be explained by considering the example of a single plate. If the plate has a small degree of randomness, then the statistics of the natural fre-

^{a)}Author to whom correspondence should be addressed. Electronic mail: rsl21@eng.cam.ac.uk

quencies and mode shapes will be dependent on the detailed distribution of the random parameters. However, if the plate is sufficiently random, then the statistics of the natural frequencies and mode shapes are known to saturate to universal distributions—this can be viewed as a kind of central limit theorem, wherein the statistics of the “output quantities,” the modal properties, become independent of the statistics of the “input properties,” the system parameters. Somewhat paradoxically, it then becomes possible to predict the statistics of the response of a highly random system without having any detailed knowledge of the nature of the underlying physical uncertainties. This type of analysis has been used to predict the mean and variance of the response of a single plate by Langley and Brown (2004) and of general built up structures by Langley and Cotoni (2004). In the latter case the analysis was employed within the framework of statistical energy analysis (SEA, see, for example, Lyon and DeJong, 1995), thus allowing that long established method to be extended to variance prediction.

The nonparametric analysis of Langley and Cotoni (2004) assumes that a complex system can be broken down into a set of subsystems, and that the natural frequencies and mode shapes of each subsystem follow the statistics of the Gaussian orthogonal ensemble (GOE, see, for example, Mehta, 1991). There is significant evidence for the occurrence of the GOE in structural systems that are sufficiently random (for example, Weaver, 1989; Lobkis *et al.*, 2000; Langley and Brown, 2004), and this condition is most likely to be met at high frequencies, where the short wavelength vibrations are highly sensitive to uncertainties. At lower frequencies, these assumptions will not be valid, and a number of alternative nonparametric methods have been proposed in the literature. For example, Soize (2000, 2005) has developed a family of positive definite random matrices that can be used to represent the system mass, damping, and stiffness matrices of a complete system; the statistical distributions of these matrices are centered on reference values obtained from a complete finite element model of the system. This approach allows efficient Monte Carlo simulations of the response to be performed. Alternatively, Mace and Shorter (2001) have developed a finite element substructuring technique in which substructure natural frequencies can be randomized without reformulating the system model, again allowing efficient Monte Carlo simulations to be performed. The aim of the present work is to develop an efficient nonparametric approach that avoids the need to perform Monte Carlo simulations, yet which is applicable to a wider class of structure than the method of Langley and Cotoni (2004). As explained below, the approach represents an extension of a recently developed method for predicting the ensemble mean response of a complex system.

As an alternative to the previously mentioned approaches, another class of nonparametric model is based on dividing the structure into two parts: the first part consists of those components which are highly random and suitable for nonparametric modeling, while the second part consists of modally sparse components which display essentially deterministic behavior. The fuzzy structure theory of Soize (1993) and the hybrid FE-SEA theory of Shorter and Langley

(2005a) fall into this category. In the latter method the finite element (FE) method and SEA are combined in a single model of the system. Very few degrees of freedom are employed compared to a full FE model of the system, and the statistical nature of the SEA modeling implies that the computed results automatically represent the ensemble average response. The coupling between the FE and SEA components of the model is effected by a diffuse field reciprocity relation (Shorter and Langley, 2005b), which relates the cross spectrum of the forces exerted by the SEA subsystems on the FE model to the subsystem energies. This formulation allows the method to be applied to a wide range of structures (see, for example, Cotoni, Shorter and Lowenstein, 2006; Cotoni, Gardner *et al.*, 2006; Charpentier *et al.*, 2006) and a general computer code has now been developed which implements the technique (ESI Group, 2006). In what follows this approach is extended to predict the ensemble variance of the response, and since the method is based on non-parametric techniques this is done without the requirement for a detailed description of the system uncertainty. One extreme application of the hybrid FE-SEA method is to produce a fully SEA model of the structure, and in this case it is shown that the present theory reduces to that of Langley and Cotoni (2004), although the derivation is quite different. Whereas Langley and Cotoni (2004) extensively employed GOE statistics, the present method is derived mainly on the basis of the diffuse field reciprocity relation. The connection between the GOE and the reciprocity relation has been discussed by Langley (2006).

In what follows the hybrid FE-SEA equations governing the ensemble average response of an uncertain system are summarized in Secs. II A and II B. Before considering the ensemble variance of the response, it is necessary to consider whether the general form of the hybrid equations is applicable to a single member of the random ensemble. It is demonstrated in Sec. II C that this is in fact the case: the hybrid equations can be applied to each member of the ensemble, although the coefficients in the equations differ from those in the “mean” equation and vary randomly across the ensemble. This result is then used in Sec. III to express the ensemble variance of the response in terms of the covariances of the coefficients that appear in the hybrid equations. Expressions for the covariances of the hybrid coefficients are then obtained in Sec. IV, thus completing the analysis of the response variance. Example applications of the theory are presented in Sec. V, where the present predictions are compared with those of detailed Monte Carlo simulations of full system models. The main conclusions of the work are summarized in Sec. VI.

II. HYBRID EQUATIONS

A. Terminology and definitions

The hybrid approach to the analysis of complex vibro-acoustic systems has been described by Shorter and Langley (2005a). The system is divided into those components whose behavior can be predicted deterministically, and those components which are subject to a high degree of uncertainty arising from manufacturing and/or material variability. Typi-

cally the sensitivity of the latter components to system uncertainties will arise from the fact that they display a short wavelength of vibration over the frequency range of interest, meaning that small imperfections in the system are commensurate with the length scale of the structural deformation. Conversely the deterministic components will deform with a longer wavelength, and will thus be more robust to system variability. A typical example is an aircraft fuselage (Cotoni, Gardner *et al.*, 2006), where the frames are relatively stiff and would be classed as deterministic, while the thin skin panels would form uncertain or “statistical” components. The main feature of the hybrid method is that the two types of component are modeled in very different ways: the deterministic components are modeled using a standard technique such as the finite element (FE) method, while the statistical components are described in terms of their vibrational energy, as is done in statistical energy analysis (SEA). Thus, taken together, the complete set of deterministic components can be described by a set of displacement (or equivalent) degrees of freedom \mathbf{q} , while the statistical components are described by a set of energies \mathbf{E} . To conform with existing terminology, (for example, Soize, 1993, and Lyon and DeJong, 1995) the set of deterministic components is referred to here as the “master system,” while each of the statistical components is referred to as a “subsystem.” Thus the system is described by a collection of subsystems which are coupled together via the master system; for completeness, it can be noted that if two subsystems are directly connected, without any obvious deterministic system in the junction between them, then nonetheless the junction is assigned a set of freedoms which are included in \mathbf{q} , which ensures that all the subsystems are coupled solely through the master system.

The equations of motion of the master system at a specified vibration frequency ω can be written in the form

$$\mathbf{D}_d \mathbf{q} = \mathbf{f} + \sum_k \mathbf{f}_k, \quad (1)$$

where \mathbf{D}_d is the dynamic stiffness matrix of the master system as given by the finite element method (or some other deterministic approach), \mathbf{f} is the vector of external forces applied directly to the master system, and \mathbf{f}_k is the vector of forces arising from subsystem k . It can be noted that all force and response quantities that appear in Eq. (1) represent complex amplitudes, and the time dependency $\exp(i\omega t)$ is implicitly assumed, so that, for example, the time history of the master system response is given by $\text{Re}\{\mathbf{q} \exp(i\omega t)\}$. A central aspect of the hybrid method (Shorter and Langley, 2005a) is that the force vector \mathbf{f}_k is written as the sum of two components, so that

$$\mathbf{f}_k = \mathbf{f}_{\text{rev}}^{(k)} - \mathbf{D}_{\text{dir}}^{(k)} \mathbf{q}. \quad (2)$$

Here $\mathbf{D}_{\text{dir}}^{(k)}$ is termed the “direct field dynamic stiffness matrix” for subsystem k , and $\mathbf{f}_{\text{rev}}^{(k)}$ is termed the “reverberant field force vector.” Equation (2) arises from a consideration of the physics of the response of subsystem k when viewed in terms of propagating waves. The motion of the boundary (described in terms of the master freedoms \mathbf{q}) will generate waves which travel into the subsystem and are then reflected multiple times. The term $\mathbf{D}_{\text{dir}}^{(k)}$ represents the matrix that

would arise in the absence of reflections, i.e., if all of the waves generated at the boundary were absorbed by the subsystem before reflection, so that there is only a “direct field” of waves driven by the boundary. This matrix is deterministic, and it can be calculated in a number of ways, as described by Shorter and Langley (2005a) and Langley (2006). As an example, the derivation of the matrix for a line junction on the edge of a plate has been detailed by Cotoni *et al.* (2007). The vector $\mathbf{f}_{\text{rev}}^{(k)}$ represents the “reverberant” forces arising from all of the reflected waves, together with any waves that are generated by loads applied directly to the subsystem. The randomness of the subsystem is accounted for by considering the waves which produce $\mathbf{f}_{\text{rev}}^{(k)}$ to comprise a random diffuse field; this is a form of “maximum entropy” assumption, which requires the system properties to be sufficiently random to scatter the waves in all directions with equal probability. The relationship of this assumption to common assumptions regarding the ensemble statistics of the subsystem mode shapes and natural frequencies will be discussed in Sec. IV B. It was shown by Shorter and Langley (2005b) that the cross-spectral matrix of the reverberant forces arising from the diffuse field can be written as

$$\mathbf{S}_{ff}^{(k),\text{rev}} = \text{E}[\mathbf{f}_{\text{rev}}^{(k)} \mathbf{f}_{\text{rev}}^{(k)*T}] = \left(\frac{4E_k}{\omega \pi n_k} \right) \text{Im}\{\mathbf{D}_{\text{dir}}^{(k)}\}, \quad (3)$$

where $\text{E}[\]$ indicates the ensemble average, E_k is the ensemble average vibrational energy of the subsystem, and n_k is the modal density, i.e., the average number of natural frequencies in a unit frequency band (see, for example, Cremer *et al.* 1990).

Equations (1) and (2) can be re-expressed in the form

$$\mathbf{D}_{\text{tot}} \mathbf{q} = \mathbf{f} + \sum_k \mathbf{f}_{\text{rev}}^{(k)}, \quad (4)$$

$$\mathbf{D}_{\text{tot}} = \mathbf{D}_d + \sum_k \mathbf{D}_{\text{dir}}^{(k)}. \quad (5)$$

These equations, together with Eq. (3), can be used to derive the complete set of equations which govern the response of the master system and the ensemble average vibrational energies of the subsystems; the resulting expressions are summarized in the following subsection.

B. Hybrid equations for the ensemble average response

By considering the conservation of vibrational energy flow for each subsystem j , Shorter and Langley (2005a) have shown that Eqs. (3)–(5) lead to the following set of coupled equations for the ensemble averaged subsystem energies

$$\omega(\eta_j + \eta_{d,j})E_j + \sum_k \omega \eta_{jk} n_j (E_j/n_j - E_k/n_k) = P_j + P_{\text{in},j}^{\text{ext}}. \quad (6)$$

Here η_j is the damping loss factor for subsystem j and P_j is the power input arising from external loads applied directly to the subsystem. The other terms in the equation are given by

$$\omega \eta_{d,j} = \left(\frac{2}{\pi n_j} \right) \sum_{r,s} \text{Im}\{D_{d,rs}\} (\mathbf{D}_{\text{tot}}^{-1} \text{Im}\{\mathbf{D}_{\text{dir}}^{(j)}\} \mathbf{D}_{\text{tot}}^{-1*T})_{rs}, \quad (7)$$

$$\omega \eta_{jk} n_j = (2/\pi) \sum_{r,s} \text{Im}\{D_{\text{dir},rs}^{(j)}\} (\mathbf{D}_{\text{tot}}^{-1} \text{Im}\{\mathbf{D}_{\text{dir}}^{(k)}\} \mathbf{D}_{\text{tot}}^{-1*T})_{rs}, \quad (8)$$

$$P_{\text{in},j}^{\text{ext}} = (\omega/2) \sum_{r,s} \text{Im}\{D_{\text{dir},rs}^{(j)}\} (\mathbf{D}_{\text{tot}}^{-1} \mathbf{S}_{ff} \mathbf{D}_{\text{tot}}^{-1*T})_{rs}, \quad (9)$$

where S_{ff} is the cross-spectral matrix of the loading \mathbf{f} applied directly to the master system. The set of equations represented by Eq. (6) has exactly the form of the statistical energy analysis (SEA) equations (Lyon and DeJong, 1995), with the term η_{jk} being known as a coupling loss factor; it can be shown from Eq. (8) that the reciprocity condition $\eta_{jk} n_j = \eta_{kj} n_k$ applies. The term $\eta_{d,j}$ in Eq. (6) represents an additional subsystem loss factor arising from energy dissipation in the master system, while $P_{\text{in},j}^{\text{ext}}$ is the power input to the subsystem arising from the loads on the master system. In addition to Eq. (6), the hybrid approach yields the following equation for the response of the master system:

$$\mathbf{S}_{qq} = \mathbf{E}[\mathbf{q}\mathbf{q}^{*T}] = \mathbf{D}_{\text{tot}}^{-1} \left[\mathbf{S}_{ff} + \sum_k \left(\frac{4E_k}{\omega \pi n_k} \right) \text{Im}\{\mathbf{D}_{\text{dir}}^{(k)}\} \right] \mathbf{D}_{\text{tot}}^{-1*T}. \quad (10)$$

The first term in this equation is the response caused by the directly applied forces \mathbf{f} , while the second term is the response caused by the reverberant subsystem forces $\mathbf{f}_{\text{rev}}^{(k)}$, which are assumed to be uncorrelated between different subsystems.

Equations (6) and (10) allow the system response to be computed: first Eq. (6) is solved to yield the subsystem energies E_k and then Eq. (10) is used to compute the cross-spectral matrix of the master system response. However, it should be noted that the results obtained represent the *ensemble average* (i.e., the average taken over a population of random systems) of both the subsystem energies and the master system spectral response. To consider the *ensemble variance* of these quantities, it is first necessary to derive a set of equations which govern the response of a *single member* of the ensemble, rather than the ensemble average response. It is not clear whether the form of Eqs. (6)–(10) is maintained for this case, and this issue is considered in the following subsection.

C. Hybrid equations for an ensemble member

To investigate the more general applicability of the hybrid equations, it is helpful to consider initially a completely deterministic model of the system. The degrees of freedom and the applied loads for this model can be written, respectively, as \mathbf{q}_{full} and \mathbf{f}_{full} say, where

$$\mathbf{q}_{\text{full}}^T = (\mathbf{q}^T \mathbf{q}_1^T \dots \mathbf{q}_N^T), \quad (11)$$

$$\mathbf{f}_{\text{full}}^T = (\mathbf{f}^T \mathbf{f}_1^T \dots \mathbf{f}_N^T). \quad (12)$$

Here \mathbf{q} and \mathbf{f} relate to the master system, as in the previous sections, and \mathbf{q}_k and \mathbf{f}_k relate to subsystem k , with N denoting the total number of subsystems. Suppose now that the cross-

spectral matrix of the applied forces can be written in the restricted form

$$\mathbf{S}_{\text{full}} = \langle \mathbf{f}_{\text{full}} \mathbf{f}_{\text{full}}^{*T} \rangle = \mathbf{S}_{ff} + \sum_{j=1}^N c_j \mathbf{S}_j, \quad (13)$$

where \mathbf{S}_{ff} is the cross-spectral matrix of the forces on the master system (padded with zeros to match the dimension of \mathbf{S}_{full}), and each matrix \mathbf{S}_j is fixed (i.e., nonvariable in what follows) but is multiplied by a variable amplitude c_j . It can be noted that the average $\langle \rangle$ in Eq. (13) is not taken over the ensemble of random structures, but rather over the ensemble of the applied forces, which is independent of the structural ensemble. Equation (13) limits the following analysis to a particular set of matrices \mathbf{S}_j , and this point will be discussed towards the end of the present section; at this stage it can be noted that there is no reason for \mathbf{S}_j to be identified solely with the loading on subsystem j , i.e., the index j that appears in Eq. (13) does not range over the subsystems, but over the N independent components of the loading on the subsystems. It is, however, assumed that each \mathbf{S}_j excludes the forces on the master system, so that these forces are taken to be uncorrelated from those on the subsystems.

The fact that the system is assumed to be linear implies that the vibrational energy in each subsystem will be a linear function of the cross-spectral matrix of the applied loading. It then follows from Eq. (13) that

$$\mathbf{E} = \mathbf{g}_L(\mathbf{S}_{ff}) + \mathbf{A}\mathbf{c}, \quad (14)$$

where the vectors \mathbf{E} and \mathbf{c} contain, respectively, the subsystem energies (*not* assumed to be ensemble averaged at this stage) and the coefficients c_j ; the vector \mathbf{g}_L is a linear function of the matrix \mathbf{S}_{ff} , and \mathbf{A} is an $N \times N$ matrix. The terms \mathbf{g}_L and \mathbf{A} will be determined by the equations of motion of the system and the matrices \mathbf{S}_j that appear in Eq. (13). The power input to the subsystems by the loads applied directly to the subsystems can similarly be written in the form

$$\mathbf{P} = \mathbf{B}\mathbf{c}, \quad (15)$$

where the vector \mathbf{P} contains the input powers P_j , and the $N \times N$ matrix \mathbf{B} is again determined by the equations of motion of the system and the matrices \mathbf{S}_j . It can be noted that by employing Eq. (13) it has been assumed that the subsystem energies and power inputs are averaged over the ensemble of applied forces. By eliminating \mathbf{c} from Eqs. (14) and (15) it follows that

$$\mathbf{B}\mathbf{A}^{-1}\mathbf{E} = \mathbf{P} + \mathbf{B}\mathbf{A}^{-1}\mathbf{g}_L(\mathbf{S}_{ff}). \quad (16)$$

This equation has exactly the same structure as Eq. (6) in the sense that: (i) the left hand side is a linear function of the subsystem energies, (ii) the first term on the right hand side is the power input from subsystem loading, and (iii) the second term on the right hand side is a linear function of the cross spectrum of the master system loading. The key difference between the two results is that none of the terms in Eq. (16) are averaged over the ensemble of random systems. This implies that Eq. (6) is actually applicable to each member of the ensemble, although in this case Eqs. (7)–(9) will no longer give the details of the coefficients in the equation—

the coefficients will be random across the ensemble of structures, and in particular there is no reason for the reciprocity condition $\eta_{jk}n_j = \eta_{kj}n_k$ to apply in general. It can further be shown that the cross-spectral matrix of the master system response, \mathbf{S}_{qq} , will be a linear function of \mathbf{S}_{ff} and the subsystem energies \mathbf{E} , as in Eq. (10), although the coefficients of these terms will not be the same as those that appear in Eq. (10). However, this result is not of the same importance as that regarding Eq. (6), and the master system response will be considered in more detail in Sec. III C.

To summarize the main conclusion of the above analysis, it has been shown that if the loading has the form of Eq. (13) then an equation in the form of Eq. (6) is applicable to *each member* of the random ensemble. Equation (13) represents, in addition to the loads on the master system, a set of N fixed loading spectra (represented by \mathbf{S}_j) with variable amplitudes c_j . Because the coefficients that appear in Eq. (6) will depend on the matrices \mathbf{S}_j , it can be anticipated that the mean and variance of the system response will depend on the nature (\mathbf{S}_j), as well as the amplitude (c_j), of the loading applied to the subsystems. In fact, for the mean hybrid equations this effect is seen only in the detailed evaluation of P_j , the (ensemble average) external power input to the subsystems which appears in Eq. (6), while the other coefficients in the mean equations, given by Eqs. (7)–(9), are independent of \mathbf{S}_j . As will be shown in what follows, an analogous situation arises for the variance of the response, and thus it is not required to explicitly construct the matrices \mathbf{S}_j ; rather the key point regarding Eq. (13) is that it allows the generality of the *form* of Eq. (6) to be established.

III. RESPONSE VARIANCE

A. Summary of the main assumptions

The subsequent analysis rests on the following main assumptions regarding the properties of the system under investigation:

(i) Each subsystem has random properties, for example: geometric properties, material properties, and boundary conditions. The random properties of each subsystem are statistically independent of those of the other subsystems. The system damping is deterministic, so that the present analysis yields the conditional mean and variance of the response for a specified level of damping; given random damping statistics, the unconditional mean and variance can be obtained through the laws of conditional probability. The comments regarding the damping also apply directly to all of the properties of the master system.

(ii) The response of each subsystem, when viewed across the ensemble of random structures, constitutes a diffuse wave field, so that Eq. (3) is valid. This requires the subsystem properties to be sufficiently random to scatter waves across all directions.

(iii) The statistics of the natural frequencies and mode shapes of each subsystem in isolation conform to the GOE (Gaussian orthogonal ensemble, see, for example, Langley and Brown, 2004). This condition requires the subsystem properties to be sufficiently random that the modal statistics saturate to the GOE. This assumption is not used extensively

here, the main application being to determine the variance of the power input by loading applied directly to the subsystems.

(iv) The subsystems are weakly coupled through the master system, so that the power input to a subsystem from directly applied loading can be approximated to that obtained for an uncoupled subsystem.

(v) The loading applied to the system conforms to a restricted case of Eq. (13). In particular it will be assumed that the loading within each subsystem has a deterministic spatial distribution, with a random amplitude and phase. The reasons for this restriction will be discussed in Sec. III C. A further development of this work will be required to cover spatially random loading and frequency band averaged response quantities.

Certain other properties of the system response arise as a direct consequence of the above assumptions, and each will be discussed at the appropriate point in what follows.

B. Covariance of the subsystem energies

An equation in the form of Eq. (6), or equivalently Eq. (16), can be written for a single member of the ensemble to give

$$\mathbf{C}\bar{\mathbf{E}} = \mathbf{P} + \mathbf{P}_{\text{in}}^{\text{ext}}, \quad (17)$$

$$\bar{E}_j = E_j/n_j, \quad (18)$$

where the quantity \bar{E}_j can be interpreted as the vibrational energy per mode in subsystem j , and the matrix \mathbf{C} and the power inputs \mathbf{P} and $\mathbf{P}_{\text{in}}^{\text{ext}}$ are random across the ensemble of systems. The aim of the present section is to express the ensemble variance of \bar{E}_j in terms of the statistical properties of \mathbf{C} , \mathbf{P} , and $\mathbf{P}_{\text{in}}^{\text{ext}}$, and to this end it can be noted that a very similar problem has been addressed by Langley and Cotoni (2004) in the context of variance prediction within SEA. The detailed analysis of Langley and Cotoni (2004) will not be repeated here, but rather the main steps will be outlined and applied to the present situation (it can be noted that the matrix \mathbf{C} was referred to as \mathbf{D} in that work, but this notation has been changed here to avoid possible confusion with a dynamic stiffness matrix). By expressing each random variable as the sum of an ensemble mean component and a variable component, a first order perturbation expansion can be applied to Eq. (17) to develop a result for the covariance of the subsystem energies. It follows initially from this expansion that to first order in the variable components of \mathbf{C} (Langley and Cotoni, 2004)

$$\mathbf{C}^{-1} \approx \mathbf{E}[\mathbf{C}]^{-1} - \mathbf{E}[\mathbf{C}]^{-1}\{\mathbf{C} - \mathbf{E}[\mathbf{C}]\}\mathbf{E}[\mathbf{C}]^{-1}, \quad (19)$$

$$\mathbf{E}[\mathbf{C}] = \mathbf{C}_0, \quad (20)$$

$$C_{0,jj} = \omega \left(\eta_j + \eta_{d,j} + \sum_{j \neq k} \eta_{jk} \right) n_j, \quad (21a)$$

$$C_{0,jk} = -\omega \eta_{jk} n_j \quad (j \neq k), \quad (21b)$$

where the loss factors and coupling loss factors in Eq. (21) are those used in the mean hybrid equations and given by

Eqs. (7) and (8). It follows from Eqs. (17)–(21) that to leading order the covariance of the subsystem energies \bar{E}_i and \bar{E}_j is given by

$$\begin{aligned} \text{Cov}[\bar{E}_i, \bar{E}_j] = & \sum_k \sum_s C_{0,ik}^{-1} C_{0,js}^{-1} \text{Cov}[P_k + P_{in,k}^{\text{ext}}, P_s + P_{in,s}^{\text{ext}}] \\ & + \sum_k \sum_s \sum_{r \neq k} [(C_{0,ik}^{-1} - C_{0,ir}^{-1}) C_{0,js}^{-1} + (C_{0,jk}^{-1} \\ & - C_{0,jr}^{-1}) C_{0,is}^{-1}] \hat{E}_r \text{Cov}[C_{kr}, P_s + P_{in,s}^{\text{ext}}] \\ & + \sum_k \sum_p \sum_{s \neq k} \sum_{r \neq p} (C_{0,ik}^{-1} - C_{0,is}^{-1}) (C_{0,jp}^{-1} \\ & - C_{0,jr}^{-1}) \hat{E}_s \hat{E}_r \text{Cov}[C_{ks}, C_{pr}], \end{aligned} \quad (22)$$

where $C_{0,jk}^{-1}$ represents the jk th component of \mathbf{C}_0^{-1} , and \hat{E}_s is the ensemble average of \bar{E}_s , so that

$$\hat{\mathbf{E}} = \mathbf{E}[\bar{\mathbf{E}}]. \quad (23)$$

In deriving Eq. (22) it has been noted that the matrix \mathbf{C} must satisfy the power conservation condition (Langley and Cotoni, 2004)

$$\sum_j C_{jk} = \omega(\eta_k + \eta_{d,k}). \quad (24)$$

This equation allows the random parts of diagonal entries of \mathbf{C} to be written in terms of the random parts of the off-diagonal entries, and thus the diagonal entries of \mathbf{C} do not appear explicitly in the Cov terms in Eq. (22). Following Langley and Cotoni (2004) it has been assumed in the above derivation that the subsystem loss factors η_k and $\eta_{d,k}$ are deterministic. In principle the loss factor $\eta_{d,k}$, which arises from energy dissipation in the master system, depends not only on the damping properties of the master system, but also on the random properties of the subsystems. However, for weakly coupled subsystems $\eta_{d,k}$ is generally much smaller than η_k , and can thus be replaced by the ensemble average value without affecting the accuracy of variance equations. Given that the damping has been assumed to be deterministic, the covariance yielded by Eq. (22) is conditional upon prescribed values of the system loss factors and accounts only for ensemble variations in the subsystem mass and stiffness properties. If the loss factors are random with known statistical properties, then the unconditional response variance can be expressed in terms of a standard Bayesian integral.

In general, the main practical interest is with the variance of the subsystem energies rather than with the more general case of the covariances. However, the energy covariances are given in Eq. (22) since these terms are needed to compute the variance of the response of the master system, as discussed in the following subsection. Langley and Cotoni (2004) considered only variance terms on both sides of Eq. (22); as will be shown in the following sections, the presence of a master system (which does not arise in SEA) introduces a range of nonzero covariance terms on the right hand side.

The energies \hat{E}_s that appear in Eq. (22) are by definition the solutions yielded by the mean hybrid equations, Eqs.

(6)–(9), and thus all of the terms on the right hand side of Eq. (22) can be considered to be known, other than the covariance terms. Expressions for these terms are derived in Sec. IV.

C. Variance of the master system response

It follows from Eq. (4) that the master system response for any member of the ensemble has the property

$$\mathbf{q}\mathbf{q}^{*T} = \mathbf{D}_{\text{tot}}^{-1} \left[\mathbf{f} + \sum_k \mathbf{f}_{\text{rev}}^{(k)} \right] \left[\mathbf{f}^{*T} + \sum_k \mathbf{f}_{\text{rev}}^{(k)*T} \right] \mathbf{D}_{\text{tot}}^{-T*}, \quad (25)$$

where the reverberant forces $\mathbf{f}_{\text{rev}}^{(k)}$ are random across the ensemble. It follows from assumption (i) of Sec. III A that the reverberant forces in the various subsystems are uncorrelated; thus if Eq. (3) is employed, it can readily be shown that the ensemble average of Eq. (25) is given by Eq. (10). Thus

$$\begin{aligned} \mathbf{q}\mathbf{q}^{*T} - \mathbf{E}[\mathbf{q}\mathbf{q}^{*T}] = & \mathbf{D}_{\text{tot}}^{-1} \left[\mathbf{f} \sum_k \mathbf{f}_{\text{rev}}^{(k)*T} + \sum_k \mathbf{f}_{\text{rev}}^{(k)} \mathbf{f}^{*T} \right. \\ & + \sum_{k,s} \mathbf{f}_{\text{rev}}^{(k)} \mathbf{f}_{\text{rev}}^{(s)*T} \\ & \left. - \sum_k \left(\frac{4a_k \hat{E}_k}{\omega\pi} \right) \text{Im}\{\mathbf{D}_{\text{dir}}^{(k)}\} \right] \mathbf{D}_{\text{tot}}^{-T*}. \end{aligned} \quad (26)$$

The term a_k that appears in the final contribution to Eq. (26) does not appear in either Eq. (3) or Eq. (10), which are based on the assumption that the subsystem wave fields are perfectly diffuse. It has been shown by Langley (2006) that this may not be the case if the subsystems are excited predominantly by motions of the master system, and the factor a_k accounts for this effect—generally a_k will be close to unity, but in the extreme example of a subsystem which is connected to the master system at a single point, a_k is close to 2. The value of a_k will be discussed further for specific examples in Sec. V A.

Equation (26) can be used to derive an expression for the variance of the response of the master system providing the statistics of the reverberant forces $\mathbf{f}_{\text{rev}}^{(k)}$ are known up to the fourth moment. Given that these forces arise from many random diffuse wave components [assumption (ii) of Sec. III A], it is reasonable to invoke the central limit theorem and assume that, for a prescribed level of the subsystem energy \bar{E}_k , $\mathbf{f}_{\text{rev}}^{(k)}$ is a zero-mean complex Gaussian vector process. This implies that each component of the vector has the form $z(x+iy)$, where x and y are zero-mean Gaussian random variables with $E[x^2] = E[y^2]$ and $E[xy] = 0$, and $z^2 \propto \bar{E}_k$. Given Eq. (3), with the addition of the factor a_k , it can then be deduced that all of the odd moments of $\mathbf{f}_{\text{rev}}^{(k)}$ are zero and that

$$\mathbf{E}[f_i^{(k)} f_j^{(s)*}] = \left(\frac{4a_k \hat{E}_k}{\omega\pi} \right) \delta_{ks} \text{Im}\{\mathbf{D}_{\text{dir},ij}^{(k)}\}, \quad (27)$$

$$\mathbf{E}[f_i^{(k)} f_j^{(s)}] = 0, \quad (28)$$

$$\begin{aligned} E[f_i^{(k)} f_j^{(s)*} f_n^{(p)} f_m^{(q)*}] &= \left(\frac{E[\bar{E}_k \bar{E}_p]}{\hat{E}_k \hat{E}_p} \right) \{E[f_i^{(k)} f_j^{(s)*}] E[f_n^{(p)} f_m^{(q)*}] \\ &+ E[f_i^{(k)} f_m^{(q)*}] E[f_j^{(s)*} f_n^{(p)}]\}, \end{aligned} \quad (29)$$

where the subscript “rev” has been dropped for ease of notation. In assumption (v) of Sec. III A, it was stated that the present analysis excludes spatially random applied forces. The reason for this can now be explained: in Eqs. (14) and (15) the system response quantities are taken to be averaged over the ensemble of the random forces, so that Eq. (13) can be applied. This implies that Eq. (26) must be averaged in this way, which means that products of the form $f_i^{(k)} f_j^{(s)*}$ will be averaged over the loading before the equation is squared to find the variance. If the loading is spatially random, then this “preaveraging” will invalidate Eq. (29), and additional analysis is required. The same type of additional analysis is also required to deal with frequency averaged quantities, and this falls beyond the scope of the present work.

Equations (26)–(29) yield the ensemble variance of the response of the master system in the form

$$\begin{aligned} \text{Var}[(S_{qq})_{ij}] &= 2(\mathbf{D}_{\text{tot}}^{-1} \mathbf{S}_{ff} \mathbf{D}_{\text{tot}}^{-T*})_{ij} \sum_k \hat{E}_k G_{ij}^{(k)} \\ &+ \sum_{k,s} \{2\text{Cov}[\bar{E}_k, \bar{E}_s] + \hat{E}_k \hat{E}_s\} G_{ij}^{(k)} G_{ij}^{(s)}, \end{aligned} \quad (30)$$

where

$$\mathbf{G}^{(k)} = \left(\frac{4a_k}{\omega\pi} \right) \mathbf{D}_{\text{tot}}^{-1} \text{Im}\{\mathbf{D}_{\text{dir}}^{(k)}\} \mathbf{D}_{\text{tot}}^{-T*}. \quad (31)$$

The ensemble average energies \hat{E}_k that appear in Eq. (30) are given by the solution of the mean hybrid equations, while the term $\text{Cov}[\bar{E}_k, \bar{E}_s]$ is given by Eq. (22). The detailed evaluation of the terms on the right hand side of Eq. (22) is considered in the following section.

IV. CONTRIBUTIONS TO THE SUBSYSTEM ENERGY VARIANCE

A. Power input from the forces on the subsystems

The first term on the right hand side of Eq. (22) involves the covariance of the power input to subsystem k , with the contributions P_k and $P_{\text{in},k}^{\text{ext}}$ arising, respectively, from the external forces on the subsystem and the master system. The statistics of the power inputs P_k have been considered previously by Langley and Cotoni (2004); on the basis of assumptions (i) and (iv) of Sec. III A these powers can be taken to be uncorrelated, and the variance of each power can be calculated by using the uncoupled subsystem result due to Langley and Brown (2004). This approach yields

$$\text{Cov}[P_k, P_s] = \delta_{ks} E[P_k]^2 r^2(\alpha_k, m'_k), \quad (32)$$

$$\begin{aligned} r^2(\alpha, m) &= \frac{1}{\pi m} \left\{ \alpha - 1 + \frac{1}{2\pi m} [1 - e^{-2\pi m}] + E_1(\pi m) \right. \\ &\times \left. \left[\cosh(\pi m) - \frac{1}{\pi m} \sinh(\pi m) \right] \right\}. \end{aligned} \quad (33)$$

Here E_1 is the exponential integral, and the term α_k in Eq. (32) depends on the nature of the applied loading. Various loading cases have been considered in detail by Langley and Cotoni (2004) and Cotoni *et al.* (2005); for example, a point load acting on a two-dimensional structure has $\alpha_k=2.7$. The ensemble average power input $E[P_k]$ is readily calculated using standard SEA procedures (Lyon and DeJong, 1995). The term m'_k in Eq. (32) is the effective modal overlap factor of subsystem k , which is defined as

$$m'_k = \omega \eta'_k n_k, \quad (34)$$

$$\eta'_k = 1/(\omega n_k C_{0,kk}^{-1}), \quad (35)$$

where η'_k is the effective in-situ loss factor of the subsystem (Lyon and DeJong, 1995). Equation (32) is based on assumption (iii) of Sec. III A (the statistics of the subsystem natural frequencies and mode shapes are governed by the Gaussian orthogonal ensemble), as detailed by Langley and Brown (2004). It can be noted that for large m , Eq. (33) reduces to the much simpler form $r^2 \approx (\alpha-1)/(\pi m) + 1/(\pi m)^2$, which has a similar functional form to the result derived for Poisson natural frequencies by Lyon (1969): $r^2 = K/(\pi m)$, where K depends on the topological dimension of the subsystem.

Equation (32) completely describes the required statistics of the subsystem power inputs in the absence of any loading on the master system, so that $P_{\text{in},k}^{\text{ext}}=0$. If both P_k and $P_{\text{in},k}^{\text{ext}}$ are present simultaneously then it is necessary to consider (in addition to the statistics of $P_{\text{in},k}^{\text{ext}}$) the correlation between these two terms. A rigorous way to do this is to introduce “master” degrees of freedom at the interior loading points, so that P_k is formally subsumed within $P_{\text{in},k}^{\text{ext}}$; the analysis of the following section, which concerns the statistics of $P_{\text{in},k}^{\text{ext}}$, then covers all aspects of the subsystem input powers. Alternatively, since $P_{\text{in},k}^{\text{ext}}$ is governed by the properties of all the subsystems, while P_k is governed only by the properties of subsystem k , in many cases it will be reasonable to assume that the powers are uncorrelated. In this case

$$\text{Cov}[P_k + P_{\text{in},k}^{\text{ext}}, P_s + P_{\text{in},s}^{\text{ext}}] = \delta_{ks} \text{Var}[P_k] + \text{Cov}[P_{\text{in},k}^{\text{ext}}, P_{\text{in},s}^{\text{ext}}]. \quad (36)$$

The covariance term on the right hand side of this equation is considered in the following section.

B. Power input from the forces on the master system

Equation (6), or equivalently Eq. (16), is a statement of power balance for each subsystem, and the final term on the right hand side of each equation is the power input arising from the forces on the master system. The second-order statistics of these terms can be derived by considering a perturbation solution of the equations of motion of the master system, Eq. (1). These equations can be written in the form

$$\left[\mathbf{D}_{\text{tot}} + \sum_s \mathbf{D}_{\text{ran}}^{(s)} \right] \mathbf{q} = \mathbf{f}, \quad (37)$$

where \mathbf{D}_{tot} is given by Eq. (5) and $\mathbf{D}_{\text{ran}}^{(s)}$ is the contribution to the dynamic stiffness matrix of subsystem s arising from the reverberant field in the subsystem, i.e., the total dynamic stiffness matrix of the subsystem is given by $\mathbf{D}_{\text{dir}}^{(s)} + \mathbf{D}_{\text{ran}}^{(s)}$. Langley (2006) has shown that $\mathbf{D}_{\text{dir}}^{(s)}$ represents the ensemble average of the dynamic stiffness matrix, so that $\mathbf{D}_{\text{ran}}^{(s)}$ has zero mean. An approximate solution to Eq. (37) can be written in the form of a perturbation expansion in the matrices $\mathbf{D}_{\text{ran}}^{(s)}$, so that

$$\mathbf{q} = \mathbf{D}_{\text{tot}}^{-1} \mathbf{f} - \mathbf{D}_{\text{tot}}^{-1} \sum_s \mathbf{D}_{\text{ran}}^{(s)} \mathbf{D}_{\text{tot}}^{-1} \mathbf{f} + \dots = \mathbf{q}^{(0)} + \mathbf{q}^{(1)} + \dots, \quad (38)$$

where the term $\mathbf{q}^{(n)}$ is n th order in the matrices $\mathbf{D}_{\text{ran}}^{(s)}$. The power input to subsystem k is given by the velocity times the force \mathbf{f}_k at the subsystem boundary, i.e., $(1/2)\text{Re}\{i\omega \mathbf{q}^{*T} \mathbf{f}_k\}$, which gives

$$P_{\text{in},k}^{\text{ext}} = (\omega/2) \mathbf{q}^{*T} [\tilde{\mathbf{D}}_{\text{dir}}^{(k)} + \tilde{\mathbf{D}}_{\text{ran}}^{(k)}] \mathbf{q}, \quad (39)$$

where the tilde (\sim) is used as an abbreviated notation to represent the imaginary part of the matrix. Given Eq. (38), Eq. (39) can be expanded as a series in the matrices $\mathbf{D}_{\text{ran}}^{(s)}$ so that

$$P_{\text{in},k}^{\text{ext}} = \Pi_k^{(0)} + \Pi_k^{(1)} + \Pi_k^{(2)} + \dots, \quad (40)$$

where $\Pi_k^{(n)}$ is the n th order contribution to the power input. Each matrix $\mathbf{D}_{\text{ran}}^{(s)}$ has zero mean, and thus it follows that

$$E[P_{\text{in},k}^{\text{ext}}] = E[\Pi_k^{(0)}] + E[\Pi_k^{(2)}] + \dots, \quad (41)$$

$$\text{Cov}[P_{\text{in},k}^{\text{ext}}, P_{\text{in},r}^{\text{ext}}] = E[\Pi_k^{(1)} \Pi_r^{(1)}] + \text{Cov}[\Pi_k^{(1)}, \Pi_r^{(2)}] + \dots \quad (42)$$

The second term in Eq. (42) is of third order, and hence it can be deduced that the power covariance can be estimated correct to second order by considering only the first order contribution to Eq. (40). Now it follows from Eqs. (38) and (39) that the first order contribution has the detailed form

$$\Pi_k^{(1)} = (\omega/2) \text{Im} \left\{ 2i \mathbf{q}^{(0)*T} \tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{D}_{\text{tot}}^{-1} \sum_s \mathbf{f}_1^{(s)} - \mathbf{q}^{(0)*T} \mathbf{f}_1^{(k)} \right\}, \quad (43)$$

$$\mathbf{f}_1^{(s)} = -\mathbf{D}_{\text{ran}}^{(s)} \mathbf{q}^{(0)}, \quad (44)$$

where $\mathbf{q}^{(0)}$ is given by Eq. (38) and $\mathbf{f}_1^{(s)}$ is the first order approximation to the reverberant force arising in subsystem s . Taking this force to satisfy Eq. (27), it follows after some algebra that the covariance of the power inputs to subsystems k and r can be written as

$$\text{Cov}[P_{\text{in},k}^{\text{ext}}, P_{\text{in},r}^{\text{ext}}] = \left(\frac{\omega^2}{8} \right) \sum_s \left(\frac{4\hat{E}_s^{\text{pert}} a_s}{\omega \pi} \right) (\mathbf{q}^{(0)*T} \mathbf{J}_{ksr} \mathbf{q}^{(0)}), \quad (45)$$

where

$$\mathbf{J}_{ksr} = (1/2)[\mathbf{H}_{ksr} + \mathbf{H}_{rsk}], \quad (46a)$$

$$\begin{aligned} \mathbf{H}_{ksr} = & 4\tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{D}_{\text{tot}}^{-1} \tilde{\mathbf{D}}_{\text{dir}}^{(s)} \mathbf{D}_{\text{tot}}^{-T*} \tilde{\mathbf{D}}_{\text{dir}}^{(r)} + \tilde{\mathbf{D}}_{\text{dir}}^{(k)} \delta_{kr} \delta_{sk} \\ & - 2\delta_{sr} i \tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{D}_{\text{tot}}^{-1} \tilde{\mathbf{D}}_{\text{dir}}^{(r)} + 2\delta_{sk} i \tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{D}_{\text{tot}}^{-T*} \tilde{\mathbf{D}}_{\text{dir}}^{(r)} \end{aligned} \quad (46b)$$

and it can be shown that \mathbf{J}_{ksr} is a Hermitian matrix. It should be noted that the term \hat{E}_s^{pert} that appears in Eq. (45), having arisen from Eq. (27), is not the total energy of the subsystem, but rather the first order energy predicted by the current restricted perturbation analysis. Thus Eq. (45) should not be used directly in Eq. (22), but rather the equation should be used to derive a general expression for the relative covariance of the power inputs, which should then be used in conjunction with the mean power inputs. The relative covariance is given by

$$\text{Relcov}[P_{\text{in},k}^{\text{ext}}, P_{\text{in},r}^{\text{ext}}] = \frac{\text{Cov}[P_{\text{in},k}^{\text{ext}}, P_{\text{in},r}^{\text{ext}}]}{E[P_{\text{in},k}^{\text{ext}}]E[P_{\text{in},r}^{\text{ext}}]}. \quad (47)$$

By noting that $E[P_{\text{in},k}^{\text{ext}}] = (\omega/2) \mathbf{q}^{(0)*T} \tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{q}^{(0)}$, and that $\hat{E}_s^{\text{pert}} \approx E[P_{\text{in},s}^{\text{ext}}]/(\omega \eta'_s n_s)$ for power balance, the final result for the relative covariance can be written in the form

$$\begin{aligned} \text{Relcov}[P_{\text{in},k}^{\text{ext}}, P_{\text{in},r}^{\text{ext}}] = & \sum_s \left(\frac{a_s}{\pi m'_s} \right) \\ & \times \left[\frac{(\mathbf{q}^{(0)*T} \mathbf{J}_{ksr} \mathbf{q}^{(0)}) (\mathbf{q}^{(0)*T} \tilde{\mathbf{D}}_{\text{dir}}^{(s)} \mathbf{q}^{(0)})}{(\mathbf{q}^{(0)*T} \tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{q}^{(0)}) (\mathbf{q}^{(0)*T} \tilde{\mathbf{D}}_{\text{dir}}^{(r)} \mathbf{q}^{(0)})} \right], \end{aligned} \quad (48)$$

where m'_s is given by Eq. (34). Given Eq. (38), the above result can be rephrased in terms of \mathbf{S}_{ff} rather than $\mathbf{q}^{(0)}$ if required.

An insight into the implications of Eq. (48) can be obtained by considering the special case of a single subsystem k which is attached to a single degree-of-freedom master system. In this case the dynamic stiffness matrices are scalars, and it can be shown that the relative variance of the power input reduces to

$$\begin{aligned} \text{Relvar}[P_{\text{in},k}^{\text{ext}}] & = \left(\frac{a_k}{\pi m'_k} \right) \\ & \times \left(\frac{[\text{Re}\{D_d\} + \text{Re}\{D_{\text{dir}}^{(k)}\}]^2 + [\text{Im}\{D_d\} - \text{Im}\{D_{\text{dir}}^{(k)}\}]^2}{|D_d + D_{\text{dir}}^{(k)}|^2} \right), \end{aligned} \quad (49)$$

where D_d is the dynamic stiffness of the master system, as in Eq. (1). The real part of D_d will be zero at the resonant frequency of the master system, and thus Eq. (49) can be expected to display a minimum at this frequency; the depth of the minimum will depend on the value of $\text{Re}\{D_{\text{dir}}^{(k)}\}$ and the proximity of $\text{Im}\{D_{\text{dir}}^{(k)}\}$ to $\text{Im}\{D_d\}$. A plate subsystem has $\text{Re}\{D_{\text{dir}}^{(k)}\} = 0$ (in this case $D_{\text{dir}}^{(k)}$ is equal to the point stiffness of an infinite plate), and thus the relative variance will become zero if the resistive part of the dynamic stiffness is “matched” to the master system, in the sense that $\text{Im}\{D_{\text{dir}}^{(k)}\} = \text{Im}\{D_d\}$. This type of behavior is confirmed by the examples considered in Sec. V.

A feature of Eqs. (48) and (49) is that the relative variance is predicted to be proportional to $a_k/(\pi m'_k)$. This result

can be compared with the relative variance of the power input due to loads acting directly on the subsystem, which according to Eq. (32) is given by the function $r^2(\alpha_k, m'_k)$. There is a potential contradiction here: consider, for example, the case of a point load acting on a subsystem. This case is obviously covered directly by Eq. (32), but it can also be covered using Eq. (49) by considering a single degree of freedom master system with $D_d=0$ to lie between the point load and the subsystem. In this case Eq. (49) yields $a_k/(\pi m'_k)$ rather than $r^2(\alpha_k, m'_k)$ and hence we appear to have two different answers for the same problem. A partial resolution of this issue can be obtained by noting that if there is significant modal overlap and $\alpha_k \neq 1$, then Eq. (33) can be written approximately as $r^2 \approx (\alpha_k - 1)/(\pi m'_k)$, and thus the two approaches will agree providing $a_k = \alpha_k - 1$. A full resolution has been given by Langley (2006), who has shown that the discrepancy arises from the assumption of a diffuse wave field which lies behind Eq. (27) and thus also behind Eq. (49). If the subsystem is driven at a single point, then the wave field is not fully diffuse (Langley and Cotoni, 2005), and Eq. (27) must be replaced by an alternative result that leads ultimately to Eq. (32) rather than Eq. (49). The example of a point load is an extreme case, and even in this case the two equations agree providing the modal overlap is not too small; thus Eq. (49) is considered to be sufficiently accurate, and there is no need to consider a much more complex formula based on a more detailed analysis of the nature of the wave field.

C. Entries of the energy matrix C

In addition to the covariance of the power input from the applied forces, Eq. (22) requires the covariance of the entries C_{km} of the matrix \mathbf{C} that appears in Eq. (17). This matrix describes the energy flow between the subsystems: if subsystem m has a prescribed vibrational energy $\bar{E}_m = E_m/n_m$ then the resulting energy flow into subsystem k is given by $-C_{km}\bar{E}_m$. Following a similar approach to that adopted in the previous subsection, the statistical properties of C_{km} can be investigated by considering a perturbation analysis of the system response. Equations (38)–(40) remain valid for this case, providing the excitation force \mathbf{f} is interpreted as $\mathbf{f}_{\text{rev}}^{(m)}$, the force on the master system arising from the prescribed energy in subsystem m . Although the energy \bar{E}_m is prescribed and therefore deterministic, the force $\mathbf{f}_{\text{rev}}^{(m)}$ is random, since it arises from a diffuse field, and one effect of this is to modify Eq. (42) to read

$$\text{Cov}[C_{km}, C_{rp}] = \left(\frac{1}{\bar{E}_m \bar{E}_p} \right) \{ \text{Cov}[\Pi_{km}^{(0)}, \Pi_{rp}^{(0)}] + \text{E}[\Pi_{km}^{(1)} \Pi_{rp}^{(1)}] + \dots \}, \quad (50)$$

where $\Pi_{km}^{(n)}$ represents the n th order power input to subsystem k due to the prescribed energy in subsystem m . The term $\Pi_{km}^{(0)}$ is second order in $\mathbf{f}_{\text{rev}}^{(m)}$ and zero order in the matrices $\mathbf{D}_{\text{ran}}^{(s)}$, while the term $\Pi_{km}^{(1)}$ is second order in $\mathbf{f}_{\text{rev}}^{(m)}$ and first order in the matrices $\mathbf{D}_{\text{ran}}^{(s)}$. The covariance between any two terms

$\Pi_{km}^{(0)}$ and $\Pi_{rp}^{(1)}$ is zero, since the matrices $\mathbf{D}_{\text{ran}}^{(s)}$ have zero mean and are uncorrelated from $\mathbf{f}_{\text{rev}}^{(m)}$, and this fact has been employed in the derivation of Eq. (50).

To derive an expression for the first term on the right hand side of Eq. (50), it can be noted from Eqs. (38)–(40) that

$$\Pi_{km}^{(0)} = (\omega/2) \mathbf{q}^{(0)*T} \tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{q}^{(0)} = (\omega/2) \mathbf{f}_{\text{rev}}^{(m)*T} \mathbf{D}_{\text{tot}}^{-T*} \tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{D}_{\text{tot}}^{-1} \mathbf{f}_{\text{rev}}^{(m)}. \quad (51)$$

Given the properties of $\mathbf{f}_{\text{rev}}^{(m)}$, as described by Eq. (27), it can be deduced that

$$\begin{aligned} \text{E}[\Pi_{km}^{(0)}] &= -\text{E}[C_{km}] \bar{E}_m \\ &= (2a_m \bar{E}_m / \pi) \text{Tr}(\mathbf{D}_{\text{tot}}^{-T*} \tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{D}_{\text{tot}}^{-1} \tilde{\mathbf{D}}_{\text{dir}}^{(m)}), \end{aligned} \quad (52)$$

where Tr represents the trace of the matrix. This result is fully consistent with Eqs. (6) and (8), thus confirming that the ensemble average matrix \mathbf{C} is equivalent to the set of coefficients employed in the existing hybrid method (which takes $a_m = 1$). To aid the following algebra, Eq. (52) can also be written in the abbreviated form

$$\text{E}[\Pi_{km}^{(0)}] = (\omega \bar{E}_m / 2) \text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{G}^{(m)}), \quad (53)$$

where $\mathbf{G}^{(m)}$ is given by Eq. (31). With this notation, it can be shown from Eq. (51) and Eqs. (27)–(29) that

$$\text{Cov}[\Pi_{km}^{(0)}, \Pi_{rp}^{(0)}] = \delta_{mp} \bar{E}_m^2 \left(\frac{\omega}{2} \right)^2 \text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{G}^{(m)} \tilde{\mathbf{D}}_{\text{dir}}^{(r)} \mathbf{G}^{(m)}). \quad (54)$$

An analysis very similar to that leading to Eq. (45) can be employed to calculate the contribution to Eq. (50) arising from the first order terms $\Pi_{km}^{(1)}$ and $\Pi_{rp}^{(1)}$. The power $\Pi_{km}^{(1)}$ is given by Eqs. (43) and (44), with the applied force \mathbf{f} that determines $\mathbf{q}^{(0)}$, via Eq. (38), being identified as $\mathbf{f}_{\text{rev}}^{(m)}$. Considering initially the special case of $\Pi_{km}^{(1)}$ and $\Pi_{rp}^{(1)}$ with $m=p$, Eq. (43) can be used to show that, analogous to Eq. (45)

$$\text{E}[\Pi_{km}^{(1)} \Pi_{rm}^{(1)} | \mathbf{q}^{(0)}] = \left(\frac{\omega^2}{8} \right) \sum_s \left(\frac{4 \hat{E}_s^{\text{pert}} a_s}{\omega \pi} \right) (\mathbf{q}^{(0)*T} \mathbf{J}_{ksr} \mathbf{q}^{(0)}), \quad (55)$$

$$\hat{E}_s^{\text{pert}} = \left(\frac{1}{2 \eta'_s n_s} \right) \mathbf{q}^{(0)*T} \tilde{\mathbf{D}}_{\text{dir}}^{(s)} \mathbf{q}^{(0)}, \quad (56)$$

where the expectation has been taken over the random properties of the subsystems for a prescribed value of $\mathbf{q}^{(0)}$, and it has been noted that the random properties are statistically independent of $\mathbf{q}^{(0)}$. Equation (55) is fourth order in $\mathbf{q}^{(0)}$, and hence fourth order in the reverberant force in the source subsystem m via the relation

$$\mathbf{q}^{(0)} = -\mathbf{D}_{\text{tot}}^{-1} \mathbf{f}_{\text{rev}}^{(m)}. \quad (57)$$

Taking the average of Eq. (55) over $\mathbf{q}^{(0)}$ and employing Eqs. (27)–(29) to give the statistical properties of $\mathbf{f}_{\text{rev}}^{(m)}$ then yields

$$\mathbb{E}[\Pi_{km}^{(1)}\Pi_{rm}^{(1)}] = \bar{E}_m^2 \left(\frac{\omega}{2}\right)^2 \sum_s \left(\frac{a_s}{\pi m'_s}\right) \left\{ \text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(s)} \mathbf{G}^{(m)}) \text{Tr}(\mathbf{J}_{ksr} \mathbf{G}^{(m)}) + \text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(s)} \mathbf{G}^{(m)} \mathbf{J}_{ksr} \mathbf{G}^{(m)}) \right\}. \quad (58)$$

It follows from Eqs. (50), (53), (54), and (58) that the final result for the relative covariance of the entries of the energy matrix \mathbf{C} has the form

$$\text{Relcov}[C_{km}, C_{rm}] = \sum_s \left(\frac{a_s}{\pi m'_s}\right) \left\{ \frac{\text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(s)} \mathbf{G}^{(m)}) \text{Tr}(\mathbf{J}_{ksr} \mathbf{G}^{(m)}) + \text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(s)} \mathbf{G}^{(m)} \mathbf{J}_{ksr} \mathbf{G}^{(m)})}{\text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{G}^{(m)}) \text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(r)} \mathbf{G}^{(m)})} \right\} + \left\{ \frac{\text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{G}^{(m)} \tilde{\mathbf{D}}_{\text{dir}}^{(r)} \mathbf{G}^{(m)})}{\text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{G}^{(m)}) \text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(r)} \mathbf{G}^{(m)})} \right\}. \quad (59)$$

This result can be compared with an expression derived by Langley and Cotoni (2004) for the relative variance of an SEA coupling loss factor, $\text{Relvar}[C_{km}]$. The result obtained was $r^2(\alpha_{km}, m'_k)$, where the function r^2 is given by Eq. (33) and α_{km} is determined by the nature of the coupling between subsystems k and m . If Eq. (59) is applied to the case of two subsystems which are coupled by an M degree of freedom master system, then it can be shown that the first term on the right hand side of the equation is $O(M)$ times the other two terms, and is therefore dominant for large M . Furthermore, if $O(\mathbf{D}_{\text{tot}}) \gg O(\tilde{\mathbf{D}}_{\text{dir}}^{(k,m)})$ then the coupling between the two subsystems is weak, and it can be seen from Eq. (46) that $\mathbf{J}_{ksk} \approx \delta_{sk} \tilde{\mathbf{D}}_{\text{dir}}^{(k)}$. Equation (59) then yields $\text{Relvar}[C_{km}] = a_k / (\pi m'_k)$, and, following the discussion in Sec. IV B, this will be very close to the SEA result if $a_k = \alpha_{km} - 1$. For most types of junction $\alpha_{km} = 2$ (Langley and Cotoni, 2004), and thus the requirement for equality of the two results is simply $a_k = 1$, which conforms with the default value of this parameter.

The calculation of $\mathbb{E}[\Pi_{km}^{(1)}\Pi_{rp}^{(1)}]$ with $m \neq p$ is slightly less straightforward than the case $m=p$ discussed above, since the product includes terms of the form

$$\mathbf{f}_{1m}^{(s)} \mathbf{f}_{1p}^{(w)*T} = \mathbf{D}_{\text{ran}}^{(s)} \mathbf{q}_m^{(0)} \mathbf{q}_p^{(0)*T} \mathbf{D}_{\text{ran}}^{(w)*T}, \quad (60)$$

where the subscripts m and p indicate the source subsystem. The expectation of Eq. (60) over the random properties of the subsystems s and w cannot be evaluated by using Eq. (27), since in the present case the two reverberant forces are caused by different excitation sources. However, the expectation of Eq. (60) can be evaluated by using the methods presented by Langley (2006, Sec. III B), and it can be shown from Eq. (22) of that work that

$$\mathbb{E}[\mathbf{f}_{1m}^{(s)} \mathbf{f}_{1p}^{(w)*T}] \approx \delta_{sw} \left(\frac{2a_s}{\pi m'_s}\right) (\mathbf{q}_p^{(0)*T} \tilde{\mathbf{D}}_{\text{dir}}^{(s)} \mathbf{q}_m^{(0)}) \tilde{\mathbf{D}}_{\text{dir}}^{(s)}. \quad (61)$$

It can readily be verified that this result reduces to Eq. (27) when $m=p$, since in that case Eq. (56) applies and the energy \hat{E}_s is introduced into the equation. Given Eq. (61), the relative covariance of two entries of the coupling matrix that have $m \neq p$ can be evaluated by following the approach used to obtain Eq. (59). The result is

$$\text{Relcov}[C_{km}, C_{rp}] = \sum_s \left(\frac{a_s}{\pi m'_s}\right) \left\{ \frac{\text{Re}[\text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(s)} \mathbf{G}^{(m)} \mathbf{H}_{ksr} \mathbf{G}^{(p)})]}{\text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{G}^{(m)}) \text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(r)} \mathbf{G}^{(p)})} \right\}, \quad m \neq p. \quad (62)$$

Equation (62) has been derived by using Eq. (43) to describe the first order components of the two relevant powers; the product of the two powers is averaged first over the random system properties by employing Eq. (61), and second over the reverberant forces $\mathbf{f}_{\text{rev}}^{(m)}$ and $\mathbf{f}_{\text{rev}}^{(p)}$ by using Eqs. (27)–(29). The zero order powers make no contribution to the covariance in this case. For systems consisting only of directly coupled SEA subsystems, it can readily be shown from Eqs. (61) and (62) that the covariance between different matrix entries is small, and this is consistent with the neglect of these terms by Langley and Cotoni (2004).

D. Covariance between the power input and the energy matrix \mathbf{C}

As shown in the previous subsections, Eq. (43) can be used to represent the first order component of (i) the power input to a subsystem due to forces acting directly on the master system, and (ii) the power input to a subsystem arising from a prescribed energy in another subsystem. In Sec. IV B, representation (i) was used to derive the covariance of the power input, while in Sec. IV C, representation (ii) was used to derive one component of the covariance of the energy matrix. The same approach can be used to derive the covariance between the power input and the matrix entries, and the result can be deduced from Eq. (62) simply by replacing the energy source in subsystem m by a set of forces acting on the master system. This approach gives

$$\text{Relcov}[P_{\text{in},k}^{\text{ext}}, C_{rp}] = \sum_s \left(\frac{a_s}{\pi m'_s}\right) \left\{ \frac{\text{Re}[\text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(s)} \mathbf{q}^{(0)} \mathbf{q}^{(0)*T} \mathbf{H}_{ksr} \mathbf{G}^{(p)})]}{(\mathbf{q}^{(0)*T} \tilde{\mathbf{D}}_{\text{dir}}^{(k)} \mathbf{q}^{(0)}) \text{Tr}(\tilde{\mathbf{D}}_{\text{dir}}^{(r)} \mathbf{G}^{(p)})} \right\}, \quad (63)$$

where $\mathbf{q}^{(0)}$ is given by Eq. (38). Equation (63) completes the set of covariances required for the evaluation of Eq. (22).

V. EXAMPLE APPLICATIONS

A. General overview

In order to validate the forgoing theory, numerical simulations have been performed for a range of structures con-

taining plates. The first three examples reported below (Secs. V B–V D) concern simply supported rectangular plates; in the first example a single plate coupled to an oscillator is considered, while the second and third examples concern three plates which are coupled via springs. Benchmark “exact” results have been obtained by using the Lagrange–Rayleigh–Ritz method to generate the equations of motion of the system, which are then solved for each member of a randomized ensemble. Each plate has been modeled by using the mode shapes of the uncoupled plate as basis functions, considering thin plate theory and out-of-plane motions only. Coupling between any two plates is affected by attaching stiff linear springs between points in the interior of the plates. Although relatively simple structures are considered, this type of system can be used to validate all of the main features of the foregoing theory. Any particular plate can be designed to have a low modal density, and thus form part of the master system, or alternatively a plate can be assigned a high modal density and randomized to form a subsystem. In the final example (Sec. V E) a set of plates which are connected via a framework of beams is considered, and the finite element method has been used to create a reference model including both the in-plane and the out-of-plane motions of the plates. In all cases the plates are made of aluminium with Young’s modulus 7.2×10^{10} N/m², density 2800 kg/m³, and Poisson ratio 0.3. For the first three examples, the randomization of a plate has been achieved by adding 20 point masses in random locations, with each mass having 0.75% of the total mass of the plate, and three rotational springs in random positions on each plate edge. An ensemble of 400 systems has been considered for each example, and the random locations of the masses and springs have been selected from a uniform distribution over the possible geometric range. The main investigated frequency range is from 100 to 300 Hz. Below 100 Hz the system is insufficiently random to justify the present nonparametric modeling approach, i.e., the assumptions listed in Sec. III A are not met, in particular assumptions (ii) and (iii) regarding the statistical properties of the subsystem dynamic response. The randomization method and the frequency range considered are slightly different for the final example, and these are discussed in Sec. V E.

To model the system using the foregoing theory (referred to as the hybrid theory in what follows), randomized plates with high modal density are treated as subsystems, while deterministic components (either plates or beams) with low modal density are considered to be part of the master system. For the master system components, the same mass and stiffness matrices as used in the reference Monte Carlo model are employed in the hybrid model. The hybrid model requires the direct field dynamic stiffness matrices $\mathbf{D}_{\text{dir}}^{(k)}$, and in the present examples these terms are readily constructed from the point stiffness of an infinite plate (Langley and Shorter, 2003); for example, the out-of-plane point stiffness is $8i(mD)^{1/2}\omega$, where m is the mass per unit area and D is the flexural rigidity. If a subsystem has more than one point connection then it is assumed that any coherence effects can be

neglected, and the direct field dynamic stiffness from all connections is computed from that of each connection in isolation.

With regard to the choice of the coefficients a_k which appear in the variance equations, the basic rule is that if a subsystem carries a diffuse wave field then $a_k=1$. Based on the work of Langley and Cotoni (2005) and Langley (2006), the wave field will not be fully diffuse if the subsystem is driven mainly through the low modal density master system—the forces exerted by the master system are approximately fully correlated, and this causes a local “swell” in the diffuse field energy density at the forcing points. In this case an appropriate approximate value is $a_k=2$. The following validation studies confirm the adequacy of this approximate approach, although more exact values could be deduced from the quoted references if required.

B. Oscillator attached to a plate

The first validation case is concerned with a plate coupled to an oscillator. The plate has a very high modal density (the area is 4 m² and the thickness is 1.25 mm) with around 334 modes below 300 Hz, and it is randomized with point masses and springs. The oscillator is deterministic, with a mass of 2 kg and a resonance frequency of 200 Hz. The spring of the oscillator is “earthed” so that the oscillator is attached to the plate via the mass, rather than through the spring. The oscillator and the plate are taken to have the same damping loss factor, and the two values 1% and 0.1% are considered. Over the investigated frequency range (100–300 Hz), the modal overlap of the plate ranges, respectively, from 1 to 3.1 and from 0.1 to 0.31 for the high and low damping cases. A point load is applied either to the oscillator or to the plate at a point away from the oscillator connection point. The hybrid model for this example consists of the single-degree-of-freedom oscillator as the master system, and the plate as the sole subsystem.

Benchmark Monte Carlo results for the case of heavy damping with a force applied to the oscillator are shown in Fig. 1; for the plate the vibrational energy is shown, while for the oscillator the modulus square velocity response is shown. The gray curves are the results for each individual ensemble member, while the black curves are the ensemble average results. It can be seen that the spread of the sample responses around the mean response depends on the proximity to the resonance frequency of the oscillator: near to the resonance frequency, the variance of the plate energy response is very low, while the variance of the oscillator response is high. The opposite trend is seen away from the resonance frequency. It can also be seen that apart from the resonance effect, the variance of the plate energy tends to decrease with increasing frequency. The results yielded by the hybrid method are compared with the Monte Carlo simulations in Fig. 2. In this case results for both low and high damping are presented, and the plots show the mean and relative variance of the oscillator and the plate response. It can be seen that the effect of the oscillator resonance frequency on both the mean and the variance of the system response is captured very well by the hybrid theory. The

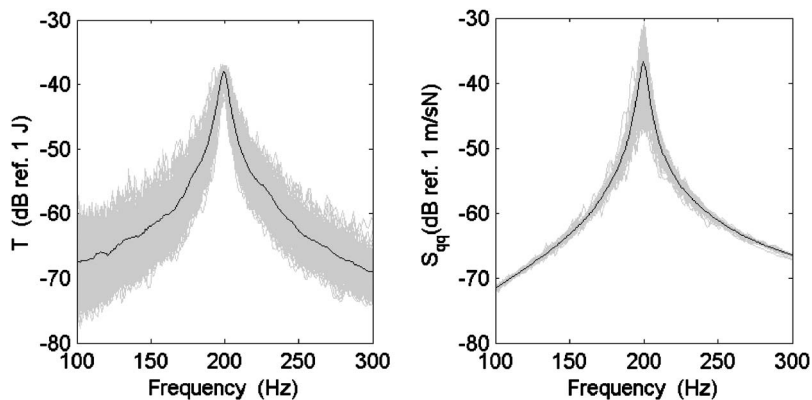


FIG. 1. The response of an oscillator-plate system to forcing on the oscillator. Left: the vibrational energy of the plate. Right: the modulus square velocity of the oscillator. The light curves are the results from each of 400 Monte Carlo simulations, the black curves are the ensemble average results.

mean response can also be predicted by using fuzzy structure theory (Soize, 1993), in which the plate is taken to add a fuzzy impedance to the master system. Including this effect alone leads to the first term on the right hand side of Eq. (10); the second term in the equation arises from the additional reverberant force imposed on the master system by the subsystem. For the lightly damped system, the second term in Eq. (10) has been found to increase the response of the master system at resonance by a factor of 6 above that predicted by the first term. For the highly damped system the increase in the response is 27% at resonance. These results indicate the importance of the second term in Eq. (10), so that, in physical terms, the plate adds both fuzzy damping and reverberant loading to the master system. In Fig. 2, the dip in the plate variance at the oscillator resonance frequency was anticipated in Sec. IV B, and is discussed below Eq. (49). The physics of this dip can be explained as a type of feedback effect: the plate effectively adds damping to the oscillator—if a particular member of the ensemble has a very responsive plate, then the effective damping will be high—

but this will act to reduce the response of the oscillator and hence reduce the level of excitation acting on the plate. Conversely, an unresponsive plate will not add much damping to the oscillator, and thus the response of the oscillator will remain high and the level of excitation acting on the plate will remain high. Hence a responsive plate experiences a low level of excitation, while an unresponsive plate experiences a high level of excitation, and these situations yield a similar plate response, leading to low variance across the ensemble. Conversely, the variance of the oscillator response is high, since the effective damping changes significantly across the ensemble.

Results for the case of a point force applied to the plate are shown in Fig. 3. In this case the plate response is little affected by the oscillator, and the response is very similar to that of a single uncoupled SEA subsystem. The mean response of the oscillator shows a peak at the oscillator resonance frequency, but the relative variance does not. In fact, it can be demonstrated from Eq. (30) that in this case the relative variance of the oscillator at all frequencies is equal to

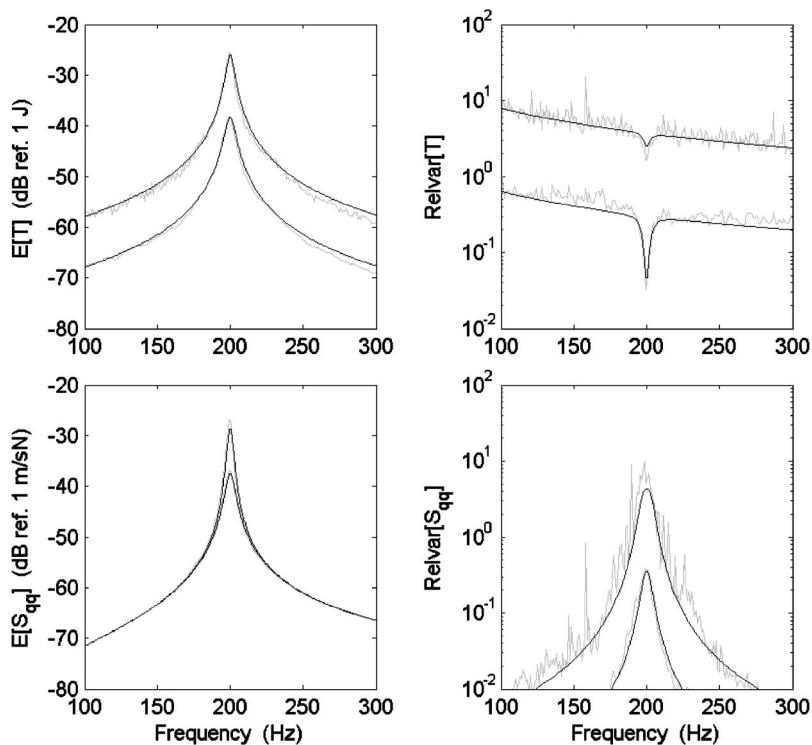


FIG. 2. The response of an oscillator-plate system for forcing on the oscillator, as calculated by Monte Carlo simulations (light curves) and the hybrid method (black curves). Upper figures: the response of the plate. Lower figures: the response of the oscillator. Left figures: the mean response. Right figures: the relative variance. In each figure two sets of curves are presented: the upper curves are for light damping, the lower curves are for heavy damping.

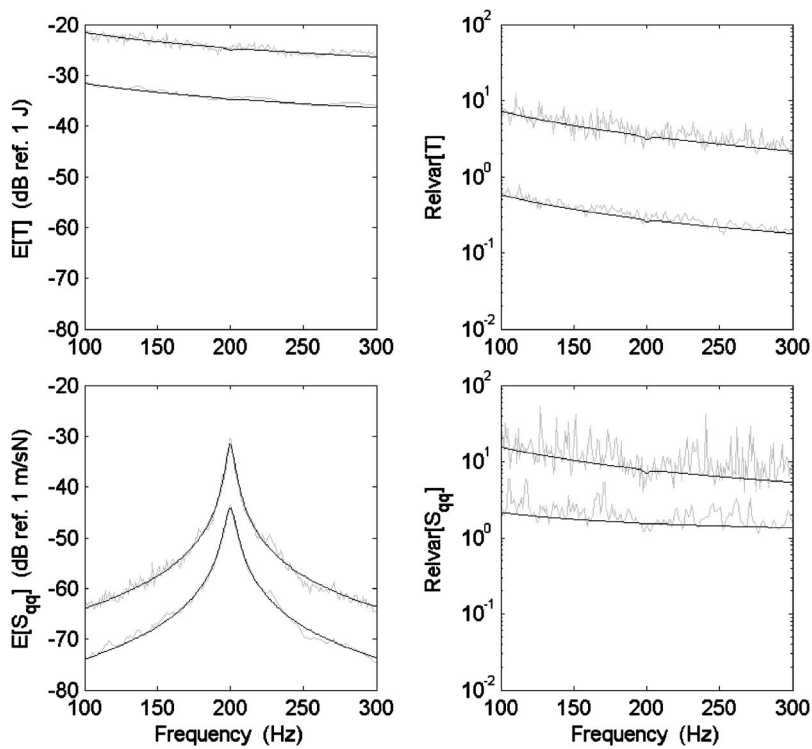


FIG. 3. The response of an oscillator-plate system for forcing on the plate, as calculated by Monte Carlo simulations (light curves) and the hybrid method (black curves). Upper figures: the response of the plate. Lower figures: the response of the oscillator. Left figures: the mean response. Right figures: the relative variance. In each figure two sets of curves are presented: the upper curves are for light damping, the lower curves are for heavy damping.

$1+2\text{Relvar}(E)$, where E is the plate energy. This result is equal to the relative variance of the plate response at a single point (Cotoni *et al.*, 2005), which is fully consistent with the physics of the situation, since the oscillator is driven by the response of the plate at the attachment point.

C. Random plate-master plate-random plate

The second validation example is concerned with a low modal density master plate which lies between two high modal density randomized plates. The aim of this example is to exercise the variance terms arising from the coupling between subsystems, as given by Eq. (59). The master plate has area $A_m=1\text{ m}^2$ (dimensions $1.05\text{ m}\times 0.95\text{ m}$) and thickness $h_m=5\text{ mm}$, and has around 19 modes below 300 Hz. The properties of the two randomized plates are $A_{r1}=4A_m$, $h_{r1}=h_m/4$, $A_{r2}=6.25A_m$, and $h_{r2}=h_m/3$, and each has around 330 modes below 300 Hz. All the plates have the same damping loss factor, which is taken to be either 1% or 0.1%. The coupling between the plates is affected by attaching three stiff springs between the master plate and randomized plate 1, and three stiff springs between the master plate and randomized plate 2. The springs connected to randomized plate 1 are connected to the master plate at the coordinates (0.464,0.762), (0.832,0.525), and (0.161,0.268), which are measured in metres relative to one corner of the plate. The corresponding coordinates for the springs connected to randomized plate 2 are (0.809,0.209), (0.402,0.181) and (0.764,0.723). The hybrid model for this system consists of a deterministic model of the master plate and the connecting springs, which is coupled to two subsystems, representing the randomized plates. The excitation consists of a point force applied to one of the subsystems.

The ensemble mean and relative variance of the energy of the subsystems are shown in Fig. 4, for each of the two

values of damping. Results from the hybrid method are compared against those of the benchmark Monte Carlo simulations and a good level of agreement can be seen. Corresponding results for the ensemble mean and relative variance of the velocity squared response at two different points on the master plate are shown in Fig. 5; the two points have coordinates (0.255,0.142) and (0.760,0.503). Again very good agreement between the hybrid method and the Monte Carlo simulations is obtained.

It can be seen in Fig. 4 that the mean and relative variance of the response of the driven subsystem is a fairly smooth function of frequency, as previously observed in Fig. 3 for the oscillator-plate example. This is because the coupling with the master system is weak, and so the response is similar to that of an uncoupled subsystem. In contrast, the response of the nondriven subsystem is not smooth, since the plate is excited through the master system. Peaks in the mean response (which occur at the resonances of the master system) tend to be accompanied by dips in the relative variance, as observed in the oscillator-plate example. Overall, the relative variance of the nondriven subsystem is higher than that of the driven subsystem, the reason being that the variance of the driven subsystem, as predicted by Eq. (22), arises mainly from the variance of the external power input; in contrast the nondriven subsystem is affected by both this variance and the variance of the coupling. This situation is particularly marked for the high damping case, where there is a significant difference in the mean subsystem energies, and the first term in Eq. (22) is strongly dominant for the driven subsystem. At the lower damping level the final term in Eq. (22) makes a more significant contribution to the driven subsystem, thus raising the relative variance beyond that due to the input power alone.

In Fig. 5 it can be seen that the mean response of the

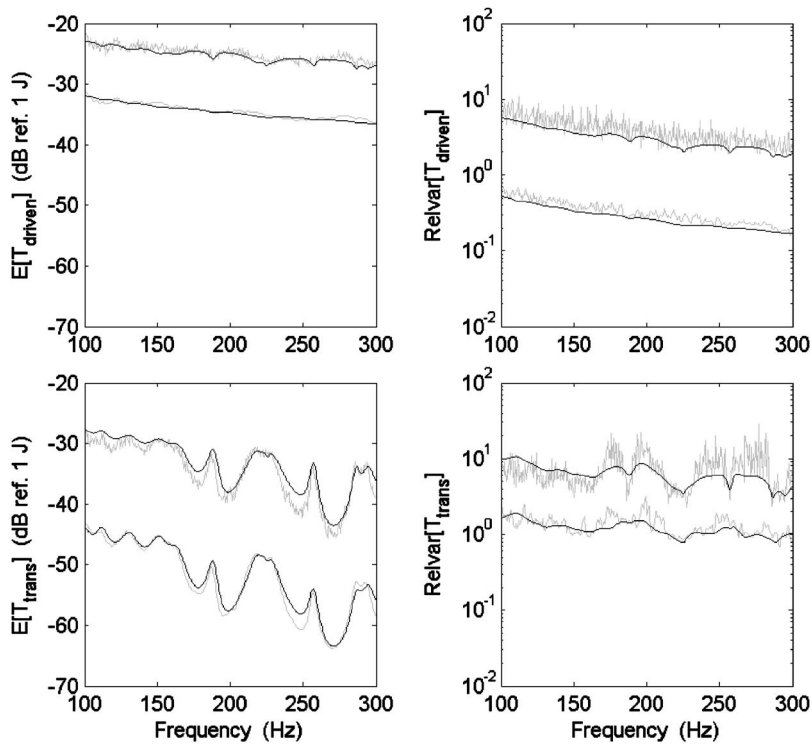


FIG. 4. The response of a random-master-random system for forcing on one of the random systems, as calculated by Monte Carlo simulations (light curves) and the hybrid method (black curves). Upper figures: the response of the driven subsystem. Lower figures: the response of the nondriven subsystem. Left figures: the mean response. Right figures: the relative variance. In each figure two sets of curves are presented: the upper curves are for light damping, the lower curves are for heavy damping.

master system displays peaks at the master system resonances, and varies significantly with the location of the observation point. Both of these features are less evident in the relative variance, and this behavior is consistent with the results displayed in Fig. 3 for the oscillator-plate example. At the lower level of damping, the relative variance shown in Fig. 5 becomes less smooth due to the increasing influence of the energy of the nondriven subsystem in Eq. (30).

D. Master plate-random plate-random plate

This example is similar to the previous example, but the topology of the three plates is changed: the master plate is coupled to the first randomized plate by three stiff springs (attached to the master system in the same location as before), and the two randomized plates are then coupled by three stiff springs. In this way the second randomized plate is

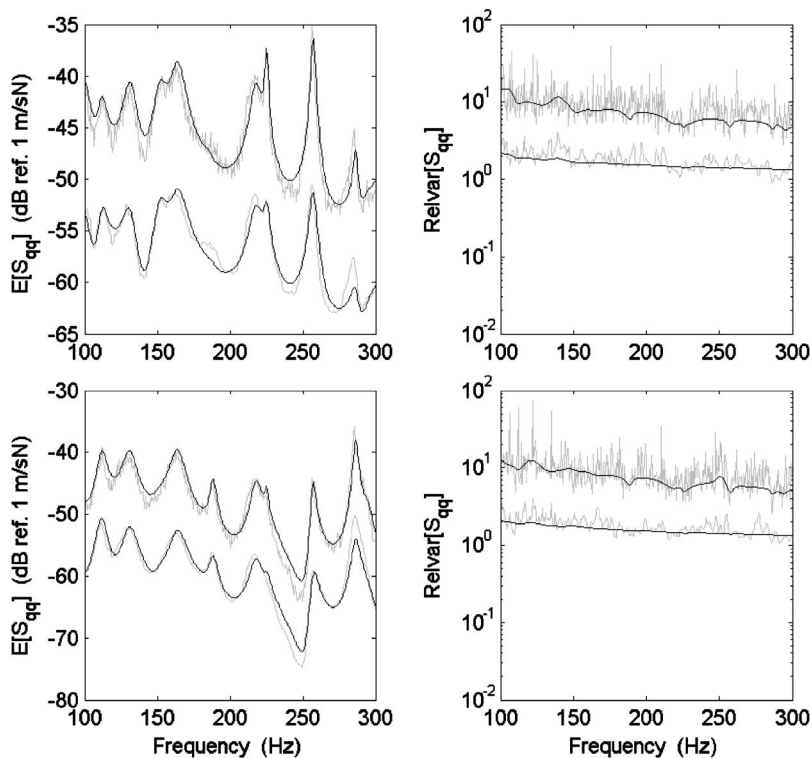


FIG. 5. The response of a random-master-random system for forcing on one of the random systems, as calculated by Monte Carlo simulations (light curves) and the hybrid method (black curves). Upper figures: the response at a point on the master system. Lower figures: the response at a different point on the master system. Left figures: the mean response. Right figures: the relative variance. In each figure two sets of curves are presented: the upper curves are for light damping, the lower curves are for heavy damping.

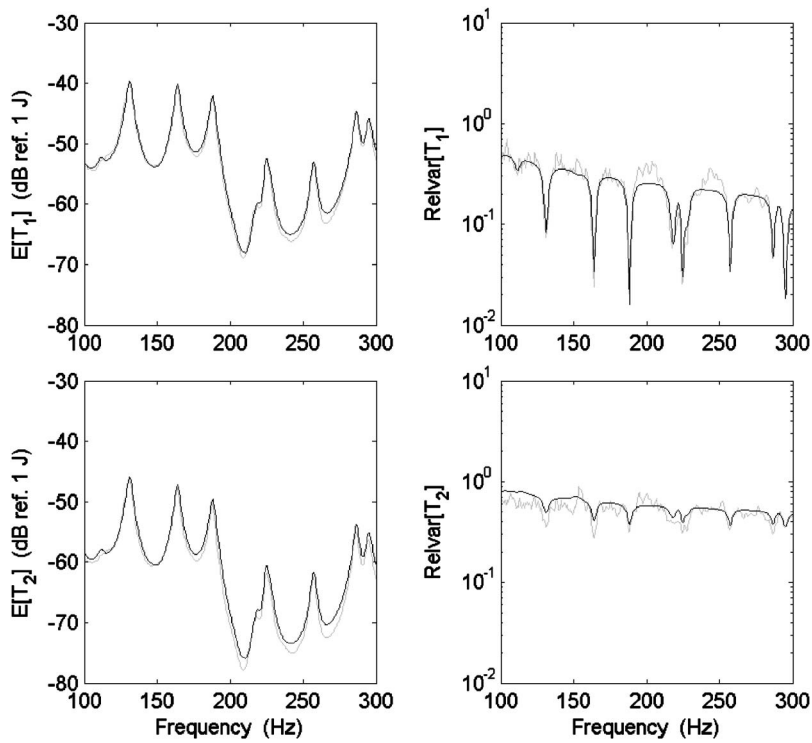


FIG. 6. The response of a master-random-random system for forcing on the master system, as calculated by Monte Carlo simulations (light curves) and the hybrid method (black curves). Upper figures: the response of subsystem 1, directly connected to the master system. Lower figures: the response of subsystem 2. Left figures: the mean response. Right figures: the relative variance.

not directly coupled to the master plate. To apply the formalism of the hybrid method, additional master freedoms must be introduced to describe the coupling between the two subsystems, as discussed in Sec. II A. In total six degrees of freedom are added, representing the displacements at each end of the three coupling springs. The springs then become part of the master system, and the analysis proceeds by direct application of the hybrid equations. In the following ex-

amples, the loss factor of each plate is taken to be 1% and a point load is applied to one of the three plates.

Results for the case of a point load applied to the master system at coordinates (0.338,0.449) are shown in Figs. 6 and 7. The response points chosen on the master system for Fig. 7 (and also Fig. 9 below) were (0.553,0.631) and (0.255,0.142). This example is significant, since it illustrates the difference between the present hybrid analysis and a di-

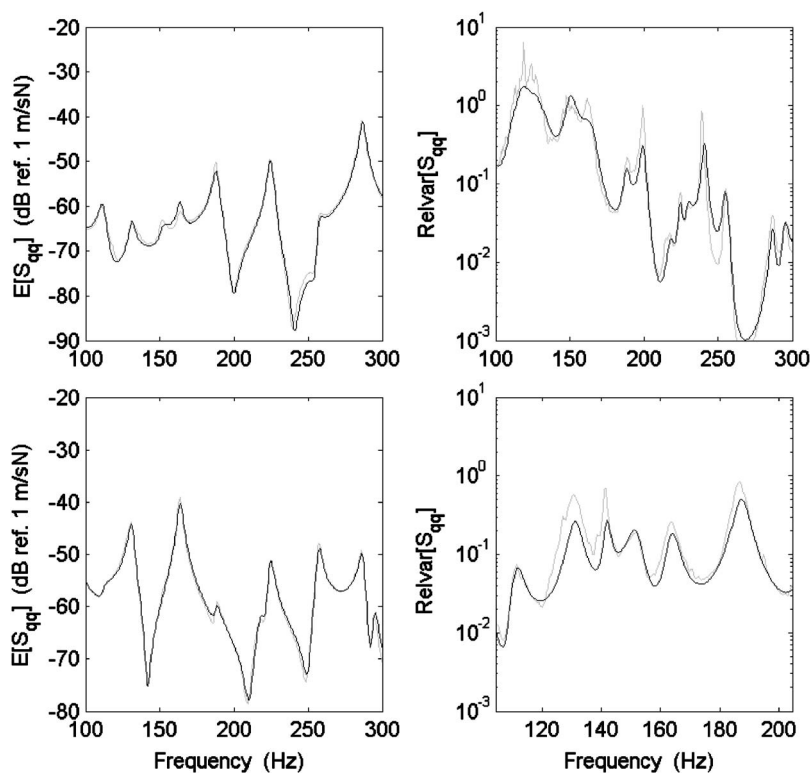


FIG. 7. The response of a master-random-random system for forcing on the master system, as calculated by Monte Carlo simulations (light curves) and the hybrid method (black curves). Upper figures: the response at a point on the master system. Lower figures: the response at a different point on the master system. Left figures: the mean response. Right figures: the relative variance.

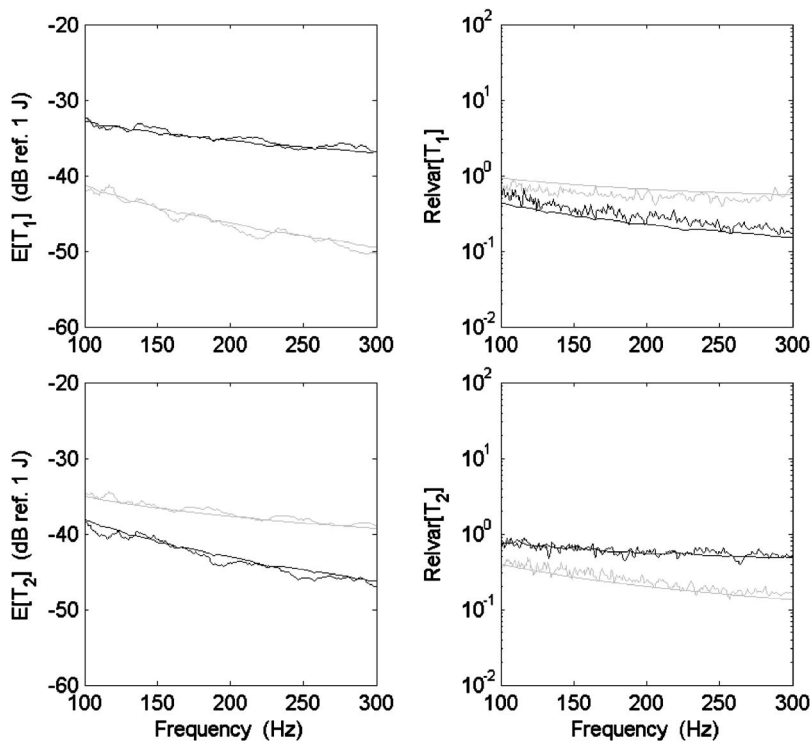


FIG. 8. The response of a master-random-random system for forcing on one of the subsystems, as calculated by Monte Carlo simulations (irregular curves) and the hybrid method (smooth curves). Upper figures: the response of subsystem 1, directly connected to the master system. Lower figures: the response of subsystem 2. Left figures: the mean response. Right figures: the relative variance. In each figure two sets of curves are presented: the black curves are for forcing on subsystem 1, the light curves are for forcing on subsystem 2.

direct first order perturbation solution of the governing equations. Equation (38) represents a perturbation analysis of the response for this case, and it can readily be shown that retaining terms to first order in $\mathbf{D}_{\text{dir}}^{(k)}$ yields a zero response for the second randomized plate. This situation does not arise with the hybrid approach, since the second randomized plate receives nonzero excitation via the energy coupling matrix \mathbf{C} . The results shown in Fig. 6 demonstrate that both the mean and variance of the energies in each subsystem are well

predicted by the method. The results for subsystem 1 show the same features as were obtained for the oscillator-plate example in Fig. 2, the main difference being that the master system now has several resonant frequencies. Subsystem 2 has a lower mean energy than subsystem 1 and a higher relative variance; both of these features are due to the fact that the subsystem is more remote from the source than subsystem 1—the mean energy is reduced due to the increased transmission path, and the relative variance is increased due

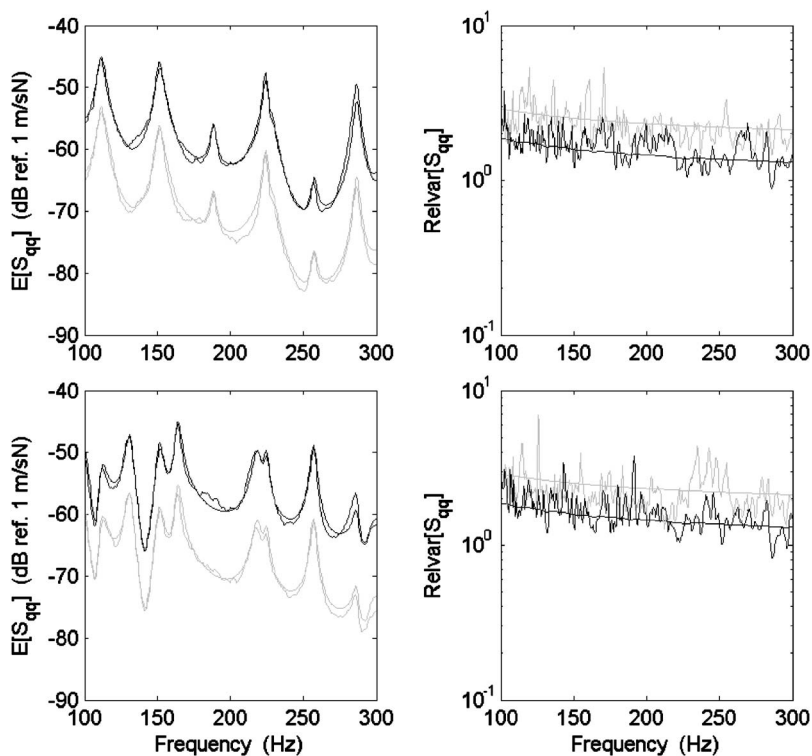


FIG. 9. The response of a master-random-random system for forcing on one of the subsystems, as calculated by Monte Carlo simulations (irregular curves) and the hybrid method (smooth curves). Upper figures: the response at a point on the master system. Lower figures: the response at a different point on the master system. Left figures: the mean response. Right figures: the relative variance. In each figure two sets of curves are presented: the black curves are for forcing on subsystem 1 (directly connected to the master system), the light curves are for forcing on subsystem 2.

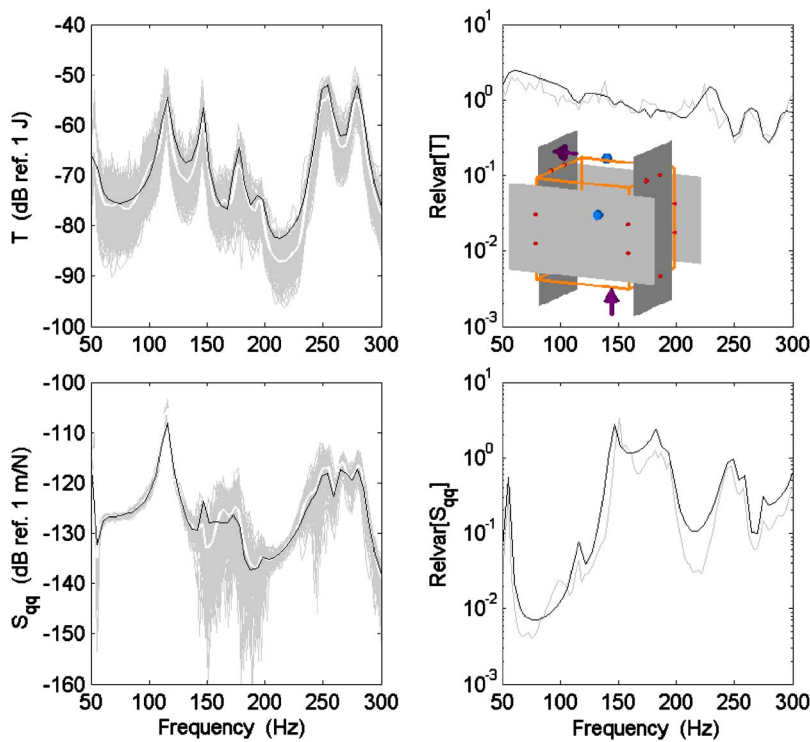


FIG. 10. (Color online) The response of a beam framework and plate assembly to a force applied to the framework, as calculated by Monte Carlo simulations and the hybrid method. Upper figures: the response of a panel. Lower figures: the response at a point on the framework. Left figures: the mean response (the white curve is the Monte Carlo result, the black curve is the hybrid prediction), together with the ensemble of Monte Carlo simulations (gray curves). Right figures: the relative variance (the black curve is the hybrid prediction, the gray curve is the Monte Carlo result).

to the additional random elements in the transmission path. In terms of Eq. (22), the first term dominates for subsystem 1, while the other terms are important for subsystem 2.

Results for point loading on one of the two subsystems are shown in Figs. 8 and 9. These results are fully consistent with physical considerations: the driven subsystem always has the highest mean energy and the lowest relative variance, while the master system has a lower mean response and a higher relative variance when the loading is on the more remote subsystem, subsystem 2. In all cases the response is well predicted by the hybrid method. In particular, it can be noted that the mean response on the master system is very sensitive to the location of the response point, while the relative variance is fairly insensitive, and this is captured by the present approach.

E. Beam framework and plate assembly

The final validation case is concerned with a beam framework to which four thin panels are attached. Each panel is connected to the framework at four points, as shown in Fig. 10. The beams are each 0.7 m long, with a square hollow section of external side width 25.4 mm and wall thickness 3.2 mm. They are connected at right angles to form the edges of a cube. The area of each panel is 0.66 m², and the thickness is 1 mm, which results in about 70 modes per panel below 300 Hz. The damping loss factor is 2% for all components. A point force is applied to one of the panels or to the framework; the forcing and response points considered are shown Fig. 10, where an arrow represents a possible force location, and a circular “sensor” represents a response location. When the sensor is located on the framework, the response quantity of interest is the square of the velocity, while the response quantity for a panel is the vibrational energy.

Benchmark results for this structure have been obtained by using a detailed finite element model, created using the VA One software package (ESI Group, 2006). The model consists of 10 481 degrees of freedom, with 96 beam elements for the framework and 3354 triangular shell elements for the panels. The structure was randomized by adding ten masses to each panel, each having 2% of the mass of the bare panel, and 100 Monte Carlo realizations were computed. For the hybrid model of this system, the framework and the in-plane motion of the panels were modeled by finite elements, constituting the master system, while the bending motion of each panel was modeled as a subsystem. The finite element part of this model has 1184 degrees of freedom. Full details of the finite element model, together with detailed information regarding the location of the forces and sensors, can be found in Cotoni *et al.* (2007).

Results for the case of a point load applied to the framework are shown in Fig. 10. The results show the mean and relative variance of the energy of one of the panels, and also of the velocity squared response of the framework at a point remote from the drive point. The response of each member of the random ensemble is added to the plot of the mean predictions, to indicate the spread in the results. The response of the frame can be seen to be almost deterministic in some frequency regions (e.g., between 60 and 130 Hz), and highly sensitive to random perturbation of the system in other regions (e.g., between 140 and 190 Hz). This behavior is accurately predicted by the hybrid method.

Results for the case of a point load applied to one of the panels are shown in Fig. 11. In this case the results are shown for one of the nondriven panels and for a point on the framework. As observed in the earlier examples, for this type of situation the relative variances show little sensitivity to the

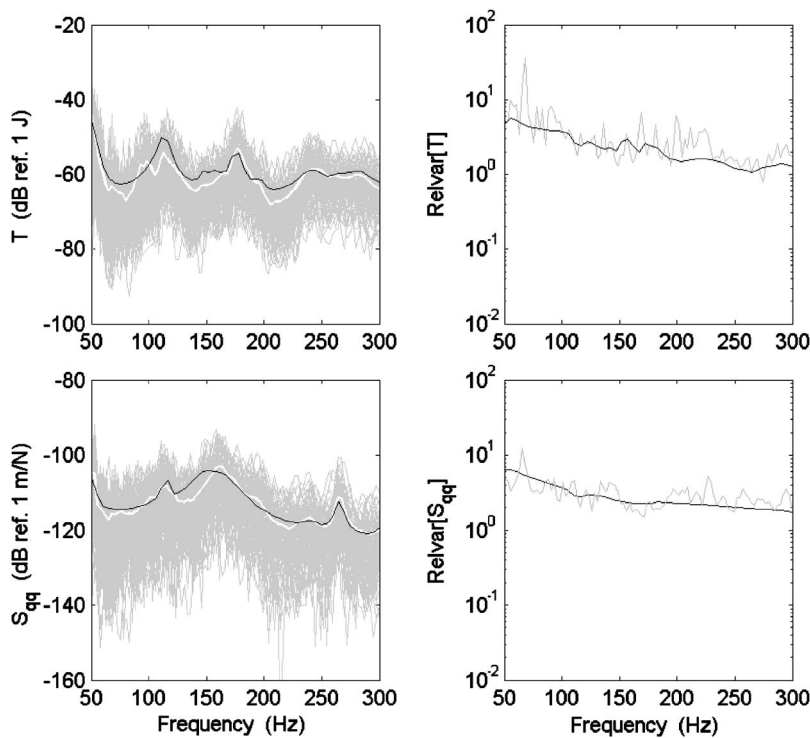


FIG. 11. The response of a beam framework and plate assembly to a force applied to a panel, as calculated by Monte Carlo simulations and the hybrid method. Upper figures: the response of a nondriven panel. Lower figures: the response at a point on the framework. Left figures: the mean response (the white curve is the Monte Carlo result, the black curve is the hybrid prediction), together with the ensemble of Monte Carlo simulations (gray curves). Right figures: the relative variance (the black curve is the hybrid prediction, the gray curve is the Monte Carlo result).

resonances of the master system. Again, the response of the system is captured well by the hybrid method.

VI. CONCLUDING COMMENTS

In this paper the hybrid FE-SEA analysis method presented by Shorter and Langley (2005a) has been extended to predict the variance of the response of an ensemble of random systems. The main results are Eq. (22) for the variance of the energy of the SEA subsystems, and Eq. (30) for the variance of the response cross spectrum of the FE model. The various covariance terms that appear on the right hand side of Eq. (22) are detailed in Sec. IV. The validation examples considered here have been fairly simple structures, but the variance theory can be applied to the same wide range of structures as the original hybrid theory.

The present approach to modeling uncertainties is non-parametric, in the sense that a detailed model of the uncertain system parameters is not required. Rather, it has been assumed that the system is sufficiently random for the response of the subsystems to display a universal statistical behavior. For the majority of the analysis it has been sufficient to assume that the subsystems carry a diffuse wave field, so that Eq. (3) is valid, although a set of coefficients, a_k , have been introduced into the present work to counter possible deficiencies in the diffuse field assumption. The relation between the diffuse field assumption and variance predictions based on the GOE has been studied by Langley (2006), and GOE based considerations can be used to inform the choice of a_k as discussed in Secs. IV B, IV C, and V A. As part of this discussion, it was shown in Sec. IV C that under the appropriate conditions the present work reduces to the SEA variance theory of Langley and Cotoni (2004), which is based on GOE statistics.

In addition to the foregoing assumptions regarding the statistics of the subsystems, it has been assumed in the present work that the system damping and all the properties of the master system are deterministic. This assumption could be relaxed by combining the present approach with a parametric model of uncertainties in the master system and/or the damping, and this parametric model could be statistical or based on fuzzy distributions or interval analysis. Also in terms of future developments, the present analysis does not consider spatially random loading or frequency band averaging; both of these effects require a modification of Eq. (29) and the subsequent equations which use this result. Finally, it was assumed in Sec. III B that the subsystem loss factor $\eta_{d,k}$, which arises from damping in the master system, can reasonably be replaced by its ensemble average value. For a very unusual system in which this might not be the case, the variance of the term can readily be found by a modified version of the analysis presented in Sec. IV C.

- Charpentier, A., Cotoni, V., and Fukui, K. (2006). "Using the hybrid FE-SEA method to predict structure-borne noise in a car body-in-white," in *Proceedings of the Inter-Noise 2006*, Honolulu, Hawaii.
- Cotoni, V., Gardner, B., Shorter, P. J., Loewenstein, M., Fuller, C., and Carneal, J. (2006). "Modelling the vibro-acoustic response of commercial aircraft using the hybrid FE-SEA method," in *Proceedings of the Inter-Noise 2006*, Honolulu, Hawaii.
- Cotoni, V., Langley, R. S., and Kidner, M. R. F. (2005). "Numerical and experimental validation of variance prediction in the statistical energy analysis of built-up systems," *J. Sound Vib.* **288**, 701–728.
- Cotoni, V., Shorter, P. J., and Langley, R. S. (2007). "Numerical and experimental validation of a hybrid finite element-statistical energy analysis method," *J. Acoust. Soc. Am.* **122**, 259–270.
- Cotoni, V., Shorter, P. J., and Loewenstein, M. (2006). "Improving SEA models of structure-borne noise transmission with the use of FE," in *Proceedings of the Inter-Noise 2006*, Honolulu, Hawaii.
- Cremer, L., Heckl, M., and Ungar, E. E. (1990). *Structure Borne Sound: Structural Vibrations and Sound Radiation at Audio Frequencies*, 2nd ed. (Springer-Verlag, Berlin).

- Elishakoff, I. (1995). "Essay on uncertainties in elastic and viscoelastic structures: From A.M. Freudenthal's criticisms to modern convex modeling." *Comput. Struct.* **56**, 871–895.
- ESI Group (2006). *VA One User's Guide* (ESI Group, Rungis, France).
- Kompella, M. S., and Bernhard, B. J. (1993). "Measurement of the statistical variation of structural-acoustic characteristics of automotive vehicle," in *Proceedings of the SAE Noise and Vibration Conference*, Warrendale, PA.
- Langley, R. S. (2007). "On the diffuse field reciprocity relationship and vibrational energy variance in a random subsystem at high frequencies," *J. Acoust. Soc. Am.* **121**, 913–921.
- Langley, R. S., and Brown, A. W. M. (2004). "The ensemble statistics of the energy of a random system subjected to harmonic excitation," *J. Sound Vib.* **275**, 823–846.
- Langley, R. S., and Cotoni, V. (2004). "Response variance prediction in the statistical energy analysis of built-up systems," *J. Acoust. Soc. Am.* **115**, 706–718.
- Langley, R. S., and Cotoni, V. (2005). "The ensemble statistics of the vibrational energy density of a random system subjected to single point harmonic excitation," *J. Acoust. Soc. Am.* **118**, 3064–3076.
- Langley, R. S., and Shorter, P. J. (2003). "The wave transmission coefficients and coupling loss factors of point connected structures," *J. Acoust. Soc. Am.* **113**, 1947–1964.
- Lobkis, O. I., Weaver, R. L., and Rozhkov, I. (2000). "Power variances and decay curvature in a reverberant system," *J. Sound Vib.* **237**, 281–302.
- Lyon, R. H. (1969). "Statistical analysis of power injection and response in structures and rooms," *J. Acoust. Soc. Am.* **45**, 545–565.
- Lyon, R. H., and DeJong, R. G. (1995). *Theory and Application of Statistical Energy Analysis*, Butterworth-Heinemann, Boston.
- Mace, B. R., and Shorter, P. J. (2001). "A local modal/perturbational method for estimating frequency response statistics of built-up structures with uncertain properties," *J. Sound Vib.* **242**, 793–811.
- Mehta, M. L. (1991). *Random Matrices* (Academic, San Diego).
- Moens, D., and Vandepitte, D. (2005). "A fuzzy finite element procedure for the calculation of uncertain frequency-response functions of damped structures: Part I—procedure," *J. Sound Vib.* **288**, 431–462.
- Schuëller, G. I. (2006). "Developments in stochastic structural mechanics," *Arch. Appl. Mech.* **75**, 755–773.
- Shorter, P. J., and Langley, R. S. (2005a). "Vibro-acoustic analysis of complex systems," *J. Ultrastruct. Res.* **288**, 669–700.
- Shorter, P. J., and Langley, R. S. (2005b). "On the reciprocity relationship between direct field radiation and diffuse reverberant loading," *J. Sound Vib.* **117**, 85–95.
- Soize, C. (1993). "A model and numerical method in the medium frequency range for vibroacoustic predictions using the theory of structural fuzzy," *J. Acoust. Soc. Am.* **94**, 849–865.
- Soize, C. (2000). "A nonparametric model of random uncertainties for reduced matrix models in structural dynamics," *Probab. Eng. Mech.* **15**, 277–294.
- Soize, C. (2005). "A comprehensive overview of a non-parametric probabilistic approach of model uncertainties for predictive models in structural dynamics," *J. Sound Vib.* **288**, 623–652.
- Weaver, R. L. (1989). "On the ensemble variance of reverberation room transmission functions, the effect of spectral rigidity," *J. Sound Vib.* **130**, 487–491.

Generating sensor signals in isotropic noise fields

Emanuël A. P. Habets^{a)}

School of Engineering, Bar-Ilan University, Ramat-Gan, Israel, and Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel

Sharon Gannot

School of Engineering, Bar-Ilan University, Ramat-Gan, Israel

(Received 14 June 2007; revised 27 September 2007; accepted 27 September 2007)

Researchers in the signal processing community often require sensor signals that result from a spherically or cylindrically isotropic noise field for simulation purposes. Although it has been shown that these signals can be generated using a number of uncorrelated noise sources that are uniformly spaced on a sphere or cylinder, this method is seldom used in practice. In this paper algorithms that generate sensor signals of an arbitrary one- and three-dimensional array that result from a spherically or cylindrically isotropic noise field are developed. Furthermore, the influence of the number of noise sources on the accuracy of the generated sensor signals is investigated. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2799929]

PACS number(s): 43.50.Ed, 43.60.Fg [EJS]

Pages: 3464–3470

I. INTRODUCTION

A spherically isotropic 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 car. Cylindrically isotropic 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,^{2,3} adaptive noise cancellation,^{4,5} and source localization. From a physical point of view the noise signals can be generated using a number of uncorrelated noise sources that are uniformly spaced on a sphere^{4,6} or cylinder. This method is, however, seldom used in practice. Some researchers, for example, convolve two uncorrelated noise signals with a room impulse response (without direct path) to generate the sensor signals that result from a spherically isotropic noise field. However, using this method the spherical noise field is not accurately simulated.

In this paper we develop algorithms that generate sensor signals of an arbitrary one-dimensional (1D) and 3D array that result from a spherical or cylindrical noise field. Furthermore, the influence of the number of noise sources on the accuracy of the generated sensor signals is investigated.

In Sec. II we show that an isotropic noise field can be generated using equally spaced noise sources on a sphere and cylinder to generate 3D and 2D diffuse noise fields, respectively. The algorithms that can be used to efficiently generate the sensor signals are developed in Sec. III. In Sec. IV we compare the spatial coherence that results from the generated sensor signals with the theoretical spatial coherence. In Sec.

V we demonstrate the use of the generated sensor signals by analyzing the directivity index of a filter and sum beamformer. Conclusions are presented in Sec. VI.

II. BACKGROUND THEORY

The noise fields of interest in this paper are composed of a superposition of uncorrelated plane waves arriving at omnidirectional sensors from various directions. The spatial coherence can be calculated by integrating the effect of a single plane wave. Therefore, we begin by considering the effect of a single plane wave. The sensor signals that result from a plane wave arriving from angle ϕ (see Fig. 1) are related by

$$x_2(t) = x_1\left(t - \frac{\Delta}{c}\right), \quad (1)$$

where c denotes the sound velocity in ms^{-1} and $\Delta = d \cos(\phi)$ the path difference of the plane wave, where d denotes the distance between the sensors. The isotropic assumption implies that the power spectrum densities of the signals are independent of the location, i.e.,

$$S_{x_1}(\omega) = S_{x_2}(\omega). \quad (2)$$

Hence, the cross-power spectrum density is given by

$$S_{x_1x_2}(\omega) = S_{x_1}(\omega)e^{-j\omega/cd \cos(\phi)}. \quad (3)$$

The spatial coherence can now be calculated by taking the integral over all plane waves that originate from a surface area A , i.e.,

$$\gamma_{x_1x_2}(\omega) = \frac{\oint_A S_{x_1x_2}(\omega) dA}{\oint_A \sqrt{S_{x_1}(\omega)S_{x_2}(\omega)} dA}, \quad (4)$$

^{a)}Author to whom correspondence should be addressed. Electronic mail: e.habets@ieee.org

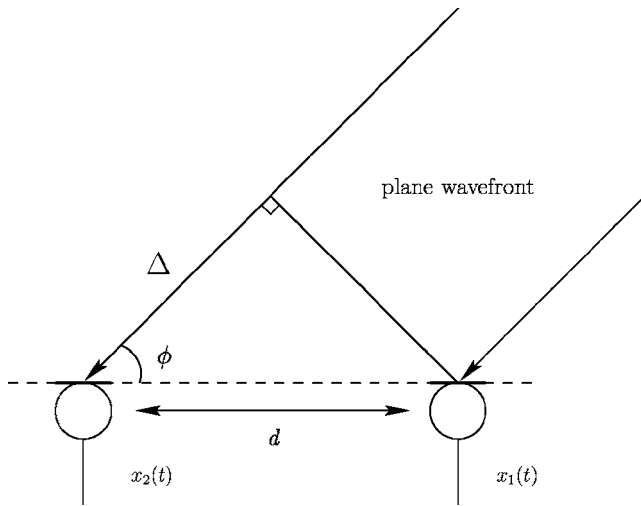


FIG. 1. Plane wave front impinging on an array of two sensors with an angle ϕ .

where $d\mathbb{A}$ denotes an infinitesimal area on a surface. Using Eqs. (2) and (3) we obtain

$$\gamma_{x_1 x_2}(\omega) = \frac{1}{A} \oint_{\mathbb{A}} e^{-j\omega/cd \cos \phi} d\mathbb{A}, \quad (5)$$

where A denotes the total surface area.

A. Spherically isotropic noise field

In case the sources are uniformly distributed on the surface of the sphere the integral in Eq. (5) can be evaluated over the surface area \mathbb{A} of the sphere. Note that the plane wave assumption holds if the radius of the sphere r is much larger than the sensor distance d . Without loss of generality it is assumed that the sensors are positioned on the x axis. The infinitesimal area on the sphere $d\mathbb{A} = r^2 \sin(\phi) d\phi d\theta$, and the surface of the sphere $A = 4\pi r^2$. In terms of the spherical coordinates $\phi \in [0, \pi]$ and $\theta \in [0, 2\pi)$ (see Fig. 2) we then obtain

$$\begin{aligned} \gamma_{x_1 x_2}(\omega) &= \frac{1}{4\pi r^2} \int_0^{2\pi} \int_0^\pi e^{-j\omega/cd \cos \phi} r^2 \sin(\phi) d\phi d\theta \\ &= \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi e^{-j\omega/cd \cos \phi} \sin(\phi) d\phi d\theta. \end{aligned} \quad (6)$$

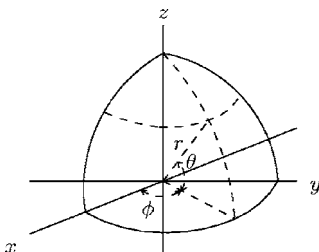


FIG. 2. Part of a sphere with radius r , azimuth ϕ , and elevation θ .

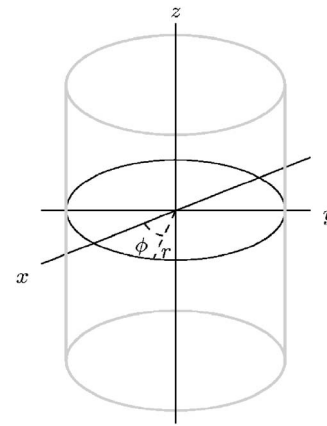


FIG. 3. Cylinder with radius r and cylindrical angle ϕ .

Now, using the substitution $g = \omega/cd \cos(\phi)$, we have

$$\gamma_{x_1 x_2}(\omega) = \frac{1}{2\omega d/c} \int_{-\omega d/c}^{\omega d/c} e^{-jg} dg = \frac{\sin(\omega d/c)}{\omega d/c}, \quad (7)$$

which is the well-known theoretical spatial coherence function for spherically isotropic noise and omni-directional sensors.⁷

B. Cylindrically isotropic noise field

In a cylindrically isotropic noise field the sources are uniformly distributed on the surface of a cylinder. It should be noted that all rings along the z axis of the cylinder have an equal contribution to the generated sensor signals. To derive the spatial coherence function in a cylindrical noise field the integral in Eq. (5) can be evaluated by integrating in one dimension, the cylindrical angle $\phi \in [0, 2\pi)$ (see Fig. 3). Note that $d\mathbb{A} = rd\phi$ and $A = 2\pi r$. We then obtain⁸

$$\gamma_{x_1 x_2}(\omega) = \frac{1}{2\pi} \int_0^{2\pi} e^{-j\omega/cd \cos \phi} d\phi = J_0(\omega d/c), \quad (8)$$

where $J_0(\cdot)$ is the zero-order Bessel function of the first kind.

III. IMPLEMENTATION

In this section efficient algorithms are developed to generate the sensor signals of an arbitrary 3D and 1D array that result from a spherically and cylindrically isotropic noise field.

In the implementation we approximate the integral in Eq. (5) by a summation, i.e.,

$$\hat{\gamma}_{x_1 x_2}(\omega) = \frac{1}{N} \sum_{n=0}^{N-1} e^{-j\omega/cd \cos \phi_n}. \quad (9)$$

Note that the number of uncorrelated noise sources N should be sufficiently large to obtain a good approximation of the integral. We elaborate on the number of noise sources in Sec. IV.

The time delay (Δ/c) between the sensor signals depends on the location of the noise source. To efficiently implement this (fractional) delay the signals are generated in the Fourier domain. In this domain the delay can be modeled using a simple phase shift.

In the sequel it is assumed that the radius of the sphere or cylinder is much larger than the span of the sensor array such that all waves can be assumed to be plane waves.

A. Spherically isotropic noise field

The basic requirement is that the noise sources are uniformly distributed on the surface of the sphere, i.e., the probability that a noise source exists in each infinitesimal area should be equal. In that sense it should be understood that a uniform distribution of the noise sources with the spherical coordinates $\phi \in [0, \pi]$ and $\theta \in [0, 2\pi)$ does not result in the desired uniform distribution on the surface of the sphere. The probability that a noise source exists in each infinitesimal area on the sphere with spherical coordinates (ϕ, θ) is given by

$$P_r(\phi \leq \tilde{\phi} \leq \phi + d\phi, \theta \leq \tilde{\theta} \leq \theta + d\theta) = \frac{dA}{A} = \frac{1}{4\pi} \sin(\phi) d\phi d\theta. \quad (10)$$

The probability density function (pdf) of the surface area A can thus be expressed in terms of ϕ and θ as

$$p_A(\theta, \phi) = \frac{1}{4\pi} \sin(\phi). \quad (11)$$

This pdf can be factorized into two independent densities for ϕ and θ , such that

$$p_\phi(\phi) = \frac{1}{2} \sin(\phi) \text{ and } p_\theta(\theta) = \frac{1}{2\pi}. \quad (12)$$

Using the *inverse transform sampling method*⁹ we can generate ϕ and θ with the desired distributions using their cumulative densities, which are given by

$$P_\phi(\phi) = \frac{1}{2}(1 - \cos(\phi)) \text{ and } P_\theta(\theta) = \frac{\theta}{2\pi}, \quad (13)$$

respectively. Let $v_1 = P_\phi(\phi)$ and $v_2 = P_\theta(\theta)$ be independent uniform random variables on $[0, 1]$ and $[0, 1)$. Then, if we solve ϕ and θ we obtain

$$\phi = \arccos(1 - 2v_1) \quad (14)$$

and

$$\theta = 2\pi v_2, \quad (15)$$

which will have the desired pdfs given in Eq. (12).

1. 3D array

The M sensor positions, relative to the first sensor position, are stacked into a matrix \mathbf{P} such that

$$\mathbf{P} = \begin{bmatrix} 0 & x_2 & \dots & x_M \\ 0 & y_2 & \dots & y_M \\ 0 & z_2 & \dots & z_M \end{bmatrix}. \quad (16)$$

Each contributing noise signal is generated directly in the frequency domain. The path difference $\Delta(m, \phi, \theta)$ denotes the difference between the path lengths from the incident plane wave with direction (ϕ, θ) to the m th sensor and the first sensor. Its value is calculated by projecting the position vector of the m th sensor $\mathbf{P}(:, m)$ on the normal \mathbf{v} of the plane wave, i.e., $\Delta(m, \phi, \theta) = \mathbf{v}^T \mathbf{P}(:, m) / \|\mathbf{v}\|^2$. Since $\|\mathbf{v}\|^2 = 1$ we have

$$\Delta(m, \phi, \theta) = \mathbf{v}^T \mathbf{P}(:, m). \quad (17)$$

The M sensor signals of length L , denoted by \mathbf{Z} , that result from $N = N_\phi N_\theta$ uniformly distributed noise sources can be efficiently generated using Algorithm 1.

2. 1D array

For an arbitrary 1D array, i.e., with equally or non-equally spaced sensors, the algorithm can be simplified by exploiting the symmetry of the array. Without loss of generality it is assumed that the sensors of the array are positioned on the x axis. The sensor positions on the x axis relative to the first sensor are

$$\mathbf{p}_x = [0 \ x_2 \ \dots \ x_M]. \quad (18)$$

Since the sensors are located on the x axis the path difference Δ only depends on the azimuth $\tilde{\phi}$, i.e., all noise sources that lie on a ring with spherical coordinates $\phi = \tilde{\phi}$ and $\theta \in [0, 2\pi)$ result in the same path difference Δ . The path difference Δ is given by $d \cos(\phi)$, where d is the distance of the sensor with respect to the origin. For the m th sensor d is equal to $\mathbf{p}_x(m)$.

The M sensor signals of length L that result from N uniformly distributed noise sources can be generated using Algorithm 2.

B. Cylindrically isotropic noise field

In the previous section we dealt with spherically isotropic noise fields. However, it has been proposed that some room acoustic fields may be more closely modeled as a cylindrically isotropic noise field. Therefore, we develop an algorithm to generate the sensor signals that result from a cylindrical noise field.

Note that in this section the variable ϕ denotes the cylindrical angle $\phi \in [0, 2\pi)$.

Algorithm 1: Creating sensor signals for an arbitrary 3D array that result from a spherical noise field.

Data: \mathbf{P} , M , L , N_ϕ , N_θ , f_s

Result: \mathbf{Z}

$$L' = 2^{\lceil \log_2(L) \rceil};$$

$$\omega = \pi f_s \left[0 : \frac{1}{L'} : 1 \right];$$

$$\theta = 2\pi \left[0 : \frac{1}{N_\theta} : \frac{N_\theta - 1}{N_\theta} \right];$$

$$\phi = \arccos \left(1 - 2 \left[0 : \frac{1}{N_\phi - 1} : 1 \right] \right);$$

for $k=1: N_\phi$ **do**

for $l=1: N_\theta$ **do**

$$\mathbf{X}' = \text{randn}(1, L' + 1) + i \text{randn}(1, L' + 1);$$

$$\mathbf{X}(1, :) = \mathbf{X}(1, :) + \mathbf{X}';$$

$$\mathbf{v} = \begin{bmatrix} \cos(\theta(l)) \sin(\phi(k)) \\ \sin(\theta(l)) \sin(\phi(k)) \\ \cos(\phi(k)) \end{bmatrix};$$

for $m=2: M$ **do**

$$\Delta = \mathbf{v}^T \mathbf{P}(:, m);$$

$$\mathbf{X}(m, :) = \mathbf{X}(m, :) + \mathbf{X}' \exp(-j\Delta\omega/c);$$

end

end

end

$$\mathbf{X} = [\mathbf{X}(:, 1:L' + 1), \text{conj}(\mathbf{X}(:, L' : -1:2))];$$

$$\mathbf{Z} = \text{ifft}(\mathbf{X}, 2L', 2);$$

$$Z_{\max} = \max(\max(\text{abs}(\mathbf{Z}(:, 1:L)))));$$

$$\mathbf{Z} = \mathbf{Z}(:, 1:L) / Z_{\max};$$

1. 3D array

The M sensor positions, relative to the first sensor position, are stacked into a matrix \mathbf{P} as defined in Eq. (23).

Algorithm 2: Creating sensor signals for an arbitrary 1D array that result from a spherical noise field.

Data: \mathbf{p}_x , M , L , N , f_s

Result: \mathbf{Z}

$$L' = 2^{\lceil \log_2(L) \rceil};$$

$$\omega = \pi f_s \left[0 : \frac{1}{L'} : 1 \right];$$

$$\phi = \arccos \left(1 - 2 \left[0 : \frac{1}{N-1} : 1 \right] \right);$$

for $k=1: N$ **do**

$$\mathbf{X}' = \text{randn}(1, L' + 1) + i \text{randn}(1, L' + 1);$$

$$\mathbf{X}(1, :) = \mathbf{X}(1, :) + \mathbf{X}';$$

for $m=2: M$ **do**

$$\Delta = \mathbf{p}_x(m) \cos(\phi(k));$$

$$\mathbf{X}(m, :) = \mathbf{X}(m, :) + \mathbf{X}' \exp(-j\Delta\omega/c);$$

end

end

$$\mathbf{X} = [\mathbf{X}(:, 1:L' + 1), \text{conj}(\mathbf{X}(:, L' : -1:2))];$$

$$\mathbf{Z} = \text{ifft}(\mathbf{X}, 2L', 2);$$

$$Z_{\max} = \max(\max(\text{abs}(\mathbf{Z}(:, 1:L)))));$$

$$\mathbf{Z} = \mathbf{Z}(:, 1:L) / Z_{\max};$$

The path difference $\Delta(m, \phi)$ denotes the difference between the path lengths from the incident plane wave with direction ϕ to the m th sensor and the first sensor. Its value is calculated by projecting the position vector of the m th sensor on the x - y plane, and subsequently projecting the resulting position vector on the normal \mathbf{v} of the plane wave, i.e., $\Delta(m, \phi) = \mathbf{v}^T \mathbf{P}(:, m) / \|\mathbf{v}\|^2$. Since $\|\mathbf{v}\|^2 = 1$ we have

$$\Delta(m, \phi) = \mathbf{v}^T \mathbf{P}(:, m). \quad (19)$$

It should be noted that according to the model the height of the sensor does not influence the path difference.

The M sensor signals of length L that result from N uniformly distributed noise sources can be generated using Algorithm 3.

Algorithm 3: Creating sensor signals for an arbitrary 1D and 3D array that result from a cylindrical noise field.

Data: \mathbf{P} , M , L , N , f_s

Result: \mathbf{Z}

$$L' = 2^{\lceil \log_2(L) \rceil};$$

$$\omega = \pi f_s \left[0 : \frac{1}{L'} : 1 \right];$$

$$\phi = 2\pi \left[0 : \frac{1}{N} : \frac{N-1}{N} \right];$$

for $k=1: N$ **do**

$$\mathbf{X}' = \text{randn}(1, L' + 1) + i \text{randn}(1, L' + 1);$$

$$\mathbf{X}(1, :) = \mathbf{X}(1, :) + \mathbf{X}';$$

$$\mathbf{v} = \begin{bmatrix} \cos(\phi(k)) \\ \sin(\phi(k)) \\ 0 \end{bmatrix};$$

for $m=2: M$ **do**

$$\Delta = \mathbf{v}^T \mathbf{P}(:, m);$$

$$\mathbf{X}(m, :) = \mathbf{X}(m, :) + \mathbf{X}' \exp(-j\Delta\omega/c);$$

end

end

$$\mathbf{X} = [\mathbf{X}(:, 1:L' + 1), \text{conj}(\mathbf{X}(:, L' : -1:2))];$$

$$\mathbf{Z} = \text{ifft}(\mathbf{X}, 2L', 2);$$

$$Z_{\max} = \max(\max(\text{abs}(\mathbf{Z}(:, 1:L)))));$$

$$\mathbf{Z} = \mathbf{Z}(:, 1:L) / Z_{\max};$$

2. 1D array

For an arbitrary 1D array the algorithm in Algorithm 3 can be simplified. Without loss of generality it is assumed that the sensors of the array are positioned on the x axis. All sensor positions are relative to the first sensor. The positions are stored in the vector \mathbf{p}_x as defined in Eq. (18).

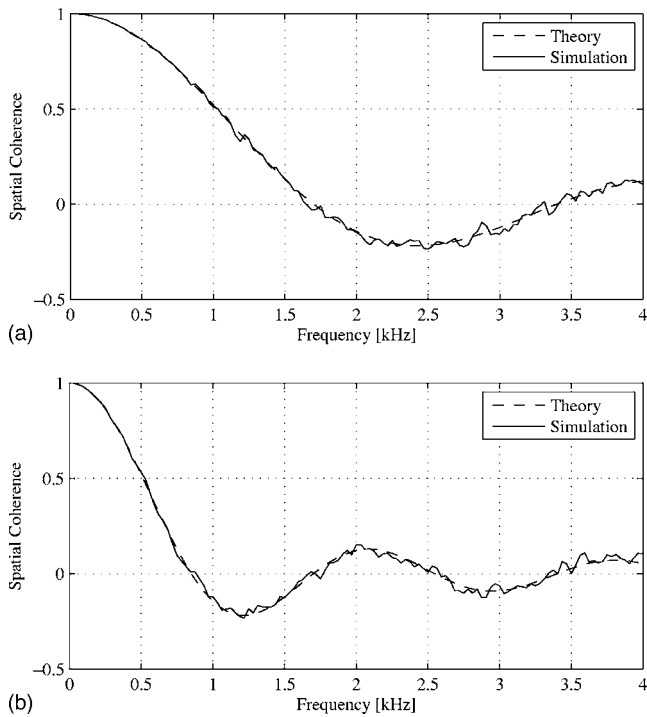


FIG. 4. Spatial coherence between two sensors with distance (a) $d=10$ cm and (b) $d=20$ cm, in a spherically isotropic noise field.

The M sensor signals of length L that result from N uniformly distributed noise sources can be generated by replacing the matrix \mathbf{P} by \mathbf{p}_x and by using

$$\Delta = \mathbf{p}_x(m)\cos(\phi(k)) \quad (20)$$

in Algorithm 3. Furthermore, the line containing the calculation of \mathbf{v} should be omitted.

IV. RESULTS

In this section we analyze the generated sensor signals. First, the obtained spatial coherence between two sensors in a spherical and cylindrical noise field is shown, and the number of noise sources N is investigated. Second, the number of noise sources N_ϕ and N_θ that are required for generating the sensor signals for a 3D array in a spherical noise field are investigated.

A. Using two sensors

First, the algorithm was used to generate the sensor signals in a spherically and cylindrically isotropic noise field using $N=64$ sources. We generated two sensor signals of $L=2^{18}$ samples and inter sensor distance $d=\{10, 20\}$ cm. The coherence between the two sensor signals was estimated using Welch's averaged periodogram method.¹⁰ We used a fast Fourier transform of length $K=256$, a Hanning window, and 75% overlap.

The simulation and theoretical results for the spherical and cylindrical noise fields are shown in Figs. 4 and 5, respectively. From the results shown in these figures we can see that the spatial coherence of the generated sensor signals closely matches the theoretical value.

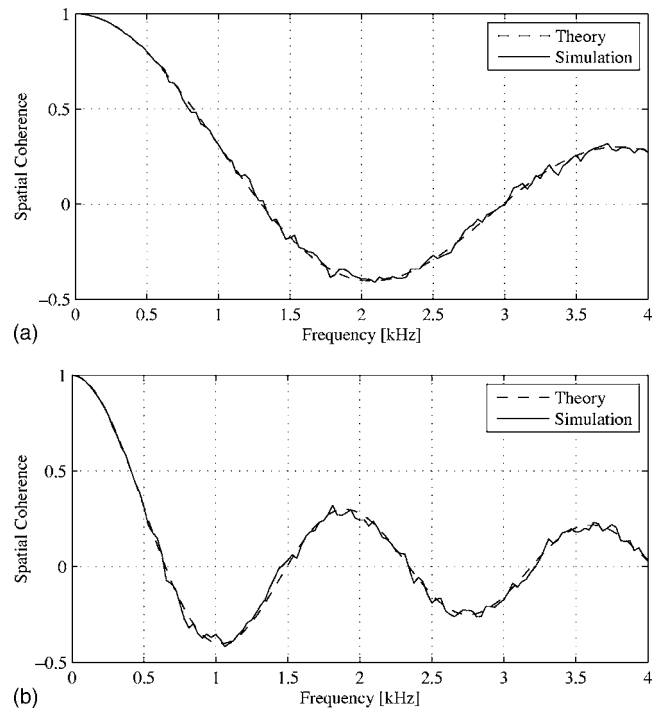


FIG. 5. Spatial coherence between two sensors with distance (a) $d=10$ cm and (b) $d=20$ cm, in a cylindrically isotropic noise field.

In practice only a finite set of noise sources can be used. Therefore, the theoretical value can only be approximated. The error between the spatial coherence of two generated sensor signals and theoretical spatial coherence is determined by i) the spectrum estimation error, and ii) the fact that a finite number of noise sources is used. The number of noise sources that is required will now be investigated. The error between the spatial coherence of two generated signals and the theoretical spatial coherence is defined by the normalized mean square error (MSE) between these two values, i.e.,

$$\text{MSE}(N) = \frac{\sum_{k=0}^{K/2} (\hat{\gamma}_{x_1 x_2}(k; N) - \gamma_{x_1 x_2}(k))^2}{\sum_{k=0}^{K/2} (\gamma_{x_1 x_2}(k))^2}, \quad (21)$$

where k denotes the discrete frequency index, and $\hat{\gamma}_{x_1 x_2}(k; N)$ denotes the estimated spatial coherence obtained using N noise sources. The results for a spherical and cylindrical noise field are shown in Figs. 6(a) and 6(b), respectively. For large N the MSE asymptotically reaches a level determined by the power spectral density estimation method. In case the number of noise sources is larger than approximately 64, the theoretical spatial coherence is well approximated.

B. Using three sensors

In this section the number of noise sources that is required for generating the sensor signals of a 3D array in a spherical noise field is investigated. The number of noise sources is specified by $N_\phi N_\theta$, where N_ϕ denotes the number of noise sources on each ring and N_θ denotes the number of rings. The average error between all unique spatial coherence

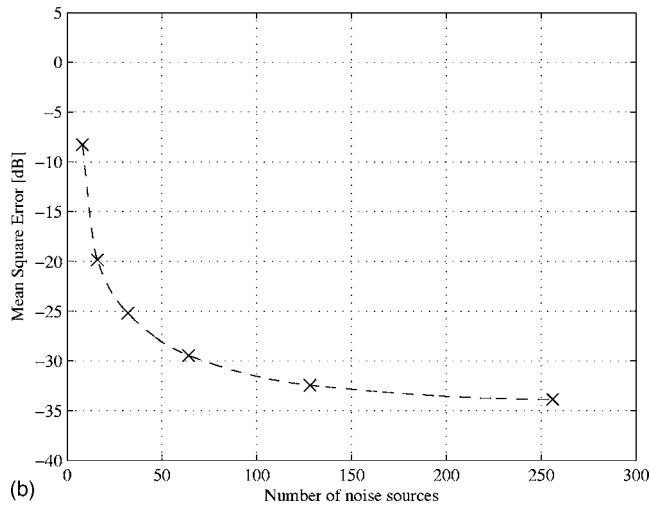
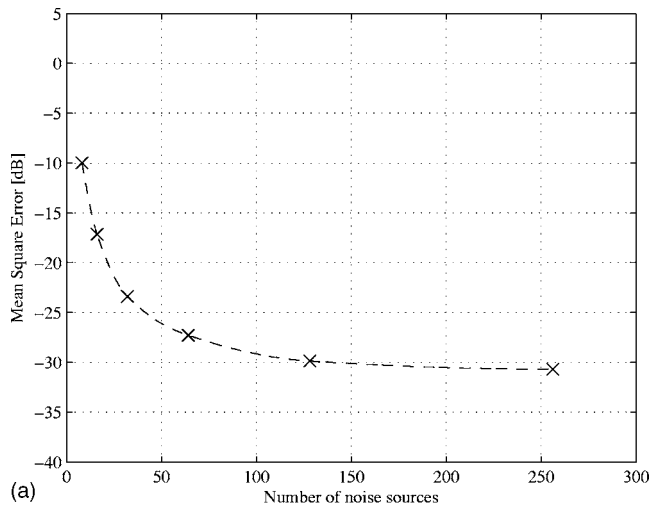


FIG. 6. MSE between the spatial coherence of the generated sensor signals and the theoretical value using N uncorrelated noise sources.

pairs of M generated sensor signals (obtained using Algorithm 1) and the corresponding theoretical spatial coherence values is obtained by

$$\text{MSE}(N_\phi, N_\theta) = \frac{2(M-2)!}{M!} \sum_{i=1}^M \sum_{j=i+1}^M \frac{\sum_{k=0}^{K/2} (\hat{\gamma}_{x_i x_j}(k; N_\phi, N_\theta) - \gamma_{x_i x_j}(k))^2}{\sum_{k=0}^{K/2} (\gamma_{x_i x_j}(k))^2}. \quad (22)$$

The position matrix of the three sensors that was used in this experiment is

$$\mathbf{P} = \begin{bmatrix} 0 & 0.2 & 0 \\ 0 & 0 & 0.2 \\ 0 & 0 & 0 \end{bmatrix}. \quad (23)$$

The contour plot of the average MSE obtained for different values N_ϕ and N_θ is shown in Fig. 7. From the results shown in this figure it can be concluded that both N_ϕ and N_θ should be sufficiently large to accurately generate the signals. In general $N_\phi \geq 96$ and $N_\theta \geq 32$ yield accurate results ($\text{MSE} < -25$ dB).

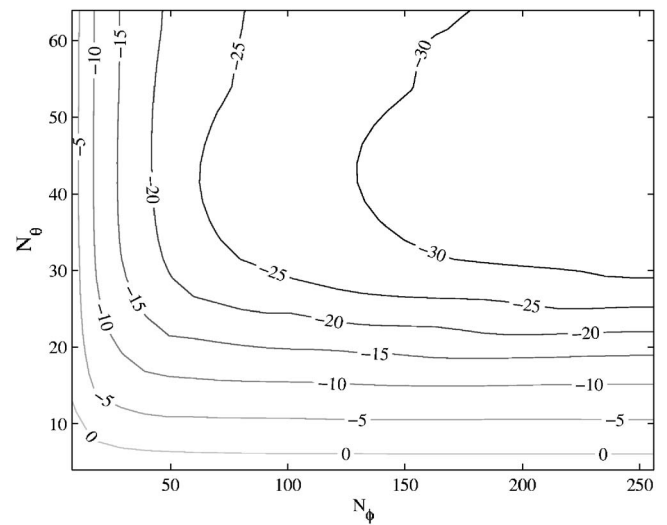


FIG. 7. Contour plot of $\text{MSE}(N_\phi, N_\theta)$ using $N_\phi N_\theta$ uncorrelated noise sources.

V. ACOUSTIC SIGNAL PROCESSING EXAMPLE

To demonstrate the applicability of the proposed simulator we use the generated sensor signals to verify the theoretical gain of a filter and sum beamformer in a spherically isotropic noise field.

The array gain of the filter and sum beamformer in a spherically isotropic noise field is equal to the directivity index (DI) of the array, which is given by¹¹

$$\text{DI}(k) = 10 \log_{10} \left(\frac{|\mathbf{w}(k)^H \mathbf{d}(k)|^2}{\mathbf{w}(k)^H \mathbf{\Gamma}_{\text{diffuse}}(k) \mathbf{w}(k)} \right) [\text{dB}], \quad (24)$$

where the numerator represents the power of the signal at the output of the beamformer, the denominator represents the power of the noise at the output of the beamformer, $\mathbf{\Gamma}_{\text{diffuse}}(k)$ denotes the spatial noise covariance matrix, $\mathbf{d}(k)$ denotes the (frequency dependent) array steering vector, and

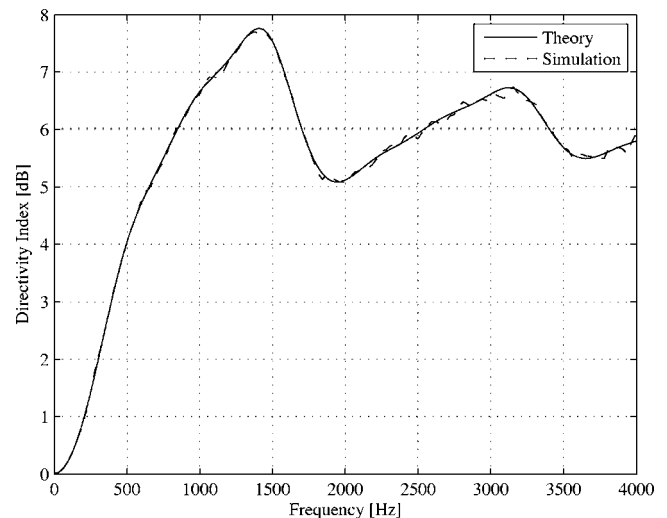


FIG. 8. Theoretical and simulated directivity index (DI) of a filter and sum beamformer and a uniform linear microphone array that consists of four microphones and an inter-microphone spacing of 0.2 m.

$\mathbf{w}(k)$ contains the frequency dependent weights of the beamformer.

Let us assume that the array steering vector of length M is $\mathbf{d}(k)=[1 \dots 1]^T \forall k$, and that $\mathbf{w}(k)=[1 \dots 1]^T \forall k$. It is easy to verify that for low frequencies ($k \rightarrow 0$) DI is equal to $10 \log_{10}(M^2/M^2)=0$ dB, and for high frequencies ($k \rightarrow K/2$) DI is equal to $10 \log_{10}(M^2/M)=10 \log(M)$ dB.

For this experiment a uniform linear microphone array was used. The number of microphones M equals 4, and the inter-microphone distance was 0.2 m. In Fig. 8 the theoretical DI, and the DI that was calculated using the generated sensor signals is shown. From the depicted results we can see that the generated sensor signals are applicable for verifying the theoretical performance of the beamformer.

VI. CONCLUSIONS

In this paper we have developed efficient algorithms to generate the sensor signals of a 1D and 3D array that are observed in a spherically or cylindrically isotropic noise field. The MATLAB implementation of the developed algorithms is available online.¹² It should be noted that the developed algorithms can be extended to a more general case in which directional sensors are used. This can be done by weighting the level of each of the source signals.^{1,5,7}

¹G. Elko, *Acoustic Signal Processing for Telecommunication*, edited by S. Gay and J. Benesty (Kluwer Academic, Hingham, MA, 2000), Chap. 10, pp. 181–237.

²I. McCowan and H. Bourlard, “Microphone array post-filter for diffuse noise field,” in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP’02)*, 2002, Vol. 1, pp. 905–908.

³S. Gannot and I. Cohen, “Speech enhancement based on the general transfer function GSC and postfiltering,” *IEEE Trans. Speech Audio Process.* **12**, 561–571 (2004).

⁴N. Dal-Degan and C. Prati, “Acoustic noise analysis and speech enhancement techniques for mobile radio applications,” *Signal Process.* **18**, 43–56 (1988).

⁵G. Elko, *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. Brandstein and D. Ward (Springer, New York, 2001), Chap. 4, pp. 61–85.

⁶M. G. J. Bird, “Speech enhancement for mobile telephony,” *IEEE Trans. Veh. Technol.* **39**, 316–326 (1990).

⁷H. Cox, “Spatial correlation in arbitrary noise fields with application to ambient sea noise,” *J. Acoust. Soc. Am.* **54**, 1289–1301 (1973).

⁸R. 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).

⁹L. Devroye, *Non-Uniform Random Variate Generation* (Springer-Verlag, New York, 2005), Chap. 2, p. 29.

¹⁰P. Stoica and R. Moses, *Spectral Analysis of Signals* (Prentice-Hall, Englewood Cliffs, NJ, 2005).

¹¹H. V. Trees, *Optimum Array Processing*, Detection, Estimation and Modulation Theory (Wiley, New York, 2002).

¹²E. Habets and S. Gannot, “Matlab implementation for: Generating sensor signals in isotropic noise fields,” url: <http://home.tiscali.nl/ehabets/ginfs.html> (last viewed September, 2007).

Analytic mode matching for a circular dissipative silencer containing mean flow and a perforated pipe

Ray Kirby^{a)}

School of Engineering and Design, Mechanical Engineering, Brunel University, Uxbridge, Middlesex, UB8 3PH, United Kingdom

Francisco D. Denia

Departamento de Ingeniería Mecánica y de Materiales, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain

(Received 18 January 2007; revised 22 June 2007; accepted 9 September 2007)

An analytic mode matching scheme that includes higher order modes is developed for a straight-through circular dissipative silencer. Uniform mean flow is added to the central airway and a concentric perforated screen separates the mean flow from a bulk reacting porous material. Transmission loss predictions are compared with experimental measurements and good agreement is demonstrated for three different silencers. Furthermore, it is demonstrated that, when mean flow is present, the axial kinematic matching condition should equate to that chosen for the radial kinematic boundary condition over the interface between the airway and the material. Accordingly, if the radial matching conditions are continuity of pressure and displacement, then the axial matching conditions should also be continuity of pressure and displacement, rather than pressure and velocity as previously thought. When a perforated screen is present the radial pressure condition changes, but the radial kinematic condition should always remain equivalent to that chosen for the axial kinematic matching condition; here, results indicate that continuity of displacement should be retained when a perforated screen is present. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2793614]

PACS number(s): 43.50.Gf, 43.20.Mv, 43.20.Bi [KA]

Pages: 3471–3482

I. INTRODUCTION

Dissipative silencers are often used to attenuate broadband noise emanating from fluid moving devices such as fans and internal combustion engines. When the velocity of the mean gas flow generated by a device is high, relative to the ambient speed of sound, then the acoustic performance of a dissipative silencer may be altered significantly when compared to that found in the absence of mean flow. Automotive dissipative silencers are often required to perform in environments in which the mean flow Mach number reaches values of up to 0.2; under such circumstances the effect of the mean flow should be accounted for when measuring and/or predicting silencer performance. Accordingly, this paper investigates the effect of mean flow on the acoustic performance of a typical “straight-through” bulk reacting dissipative silencer that contains a perforated screen separating the mean gas flow from the absorbing material. Theoretical predictions are obtained using an analytic approach that assumes a uniform and incompressible mean gas flow, and predictions are compared with experimental measurement and the finite element method.

Mean flow has long been known to affect dissipative silencer performance, although relatively few studies take this into account. Early work on automotive dissipative silencers focused on predicting modal attenuation in infinite

silencers; for example, Nilsson and Brander,¹ and later Cummings and Chang² obtained a number of eigenmodes for a circular silencer when mean flow was present in the central airway. Here, roots of the governing eigenequation were found using an appropriate analytic¹ or numerical root finding technique.² Nilsson and Brander¹ also included a perforated screen in their model, although later work on discontinuities (with potential for application to silencers of finite length) relied on the Wiener-Hopf technique, which is, arguably, a complex and difficult method to generalize for dissipative silencers. Cummings and Chang² omitted a perforated screen but added mean flow in the liner itself, on the basis that axial static pressure drops generated at the interface between the airway and the material induced a mean flow in the material. It is unlikely, however, that these effects will be relevant when a perforated screen is present, as the screen will significantly reduce frictional effects at the interface between the airway and the material and so reduce the axial static pressure gradient over the silencer section. An alternative method was proposed by Astley and Cummings,³ who used finite elements to compute the silencer eigenmodes for a general cross section, although they omitted a perforated screen from the analysis. A finite element based method has the advantage that it does not rely on root finding techniques (that are known to be susceptible to missing modes) in order to find the required eigenmodes for the silencer, although this is normally at the expense of extra computational effort.

Dissipative silencer models that are based on modal attenuation alone are of limited use for predicting overall si-

^{a)}Author to whom correspondence should be addressed. Electronic mail: ray.kirby@brunel.ac.uk

lencer performance as they neglect sound scattering over the inlet and outlet planes of the silencer. Moreover, measurements of silencer performance are normally reported in terms of the sound power difference across the (finite length) silencer. In view of this, Cummings and Chang⁴ used the silencer eigenmodes in an analytic mode matching scheme that enforced continuity of pressure and axial velocity over the inlet and outlet planes of the silencer. Sound power reduction across the silencer (the silencer transmission loss) was then calculated for mean flow Mach numbers of up to 0.196. This method depends on finding a sufficient number of eigenmodes to generate a converged solution: For silencers of a relatively modest size, Cummings and Chang found that up to six modes were necessary. A fast and efficient alternative to Cummings and Chang's method was proposed by Peat,⁵ and later by Kirby,⁶ who both developed closed form analytic solutions based on the attenuation of the fundamental mode only. Kirby⁶ included a perforated screen and, by adding extra terms to the series expansions of the Bessel and Neumann functions, achieved a more accurate model when compared to that proposed by Peat.⁵ The methods of Peat⁵ and Kirby⁶ are, however, accurate only over a limited frequency range for a given silencer geometry and/or material parameters. Such limitations apply also to other methods based on the fundamental mode, for example those methods investigated by Panigrahi and Munjal.⁷

An alternative approach for predicting the transmission loss of dissipative silencers with mean flow is the finite element based method of Peat and Rathi.⁸ Here, mean flow is included in the central airway, but a three-dimensional induced flow field is also introduced into the liner. This approach extends the one-dimensional assumed flow field of Cummings and Chang,⁴ and is also capable of modeling silencers of arbitrary shape (provided the geometry of the central airway is uniform). In common with Cummings and Chang, Peat and Rathi omitted a perforated screen, although when the two methods were compared agreement between them was poor at low frequencies. Peat and Rathi were, rightly, surprised by this and they tentatively proposed that these discrepancies in the transmission loss predictions were caused by the use of differing material parameters. One disadvantage of Peat and Rathi's method is that it is time consuming to implement, as it is based on the finite element method and in order to find the silencer's four poles one must solve the problem twice, using two different axial boundary conditions. For a uniform dissipative silencer (of arbitrary cross section), Kirby⁹ proposed using a point collocation technique that is, essentially, a numerical mode matching method. This reduces the dimensions of the problem by one and so speeds up solution time, although this method is still based on the finite element method. Kirby⁹ also added a perforated screen to his model and neglected an induced mean flow field in the liner; predictions were presented for two elliptical dissipative silencers and good agreement with experiment was reported.

In general it is desirable to include higher order modes in any predictive scheme for dissipative silencers. Moreover, when mean flow is present a perforated screen is almost always used to reduce static pressure losses over the silencer

and to prevent egress of the lining material. To date, the only approach in the literature that accommodates mean flow, a bulk reacting absorbent, and a perforated screen is the point collocation method of Kirby.⁹ However, this approach depends on the use of the finite element method and the writing of dedicated finite element software, which is not always the favored option for researchers. Furthermore, Kirby's method⁹ is similar to Cummings and Chang's method⁴ in that it uses a modal expansion to represent the sound field in the silencer, and then matches acoustic pressure and axial velocity over the inlet and outlet planes of the silencer. Problems with Cummings and Chang's⁴ mode matching predictions were noted by Peat and Rathi⁸ and, although Kirby⁹ reported good agreement between prediction and experiment, those problems with Cummings and Chang's method have yet to be resolved. Therefore, until these problems have been addressed, mode matching cannot yet be considered fully validated when mean flow is present, and this is true also for the numerical matching scheme of Kirby.⁹ Accordingly, this paper intends to review the appropriate axial matching conditions when mean flow is present and this is most readily achieved using an analytic mode matching approach based on Cummings and Chang's method.⁴ Moreover, although this approach requires solutions to be found for the governing silencer eigenequation, analytic methods are also popular for the ease in which computer code can be written and applied. Evidence for this latter point may be found in the large number of recent articles that rely on analytic mode matching in order to study the effect of a perforated screen on dissipative silencers without mean flow; see, for example, Refs. 10 and 11. This paper will extend this work to include mean flow.

This work begins by obtaining the eigenmodes in the silencer chamber and the inlet/outlet pipes. Here, the appropriate radial matching conditions for a perforated screen subjected to grazing mean flow and backed by a porous material are discussed and a root finding technique is also discussed in detail. The appropriate axial matching conditions over the inlet/outlet planes of the silencer are then reviewed and conclusions drawn after comparing predictions with the finite element calculations of Peat and Rathi⁸ in the absence of a perforated screen, and with point collocation predictions generated using the method of Kirby.⁹ Finally, a comparison is made between the new analytic mode matching method and experimental measurements for three different silencers with mean flow Mach numbers of 0.15, and two impedance models for the perforated screen are also investigated.

II. THEORY

The dissipative silencer is assumed to have a uniform circular cross section and to contain a uniform mean gas flow of Mach number M in the central channel (see Fig. 1). In the silencer chamber a perforated screen separates the central channel (region 2) from a bulk reacting (isotropic) porous absorbent (region 3). A plane wave propagates in the positive x direction in region 1, and the outlet pipe (region 4) is terminated anechoically. The pipe walls in regions 1 and 4, and the walls of the silencer chamber, are assumed to be rigid and impervious to sound. The analysis proceeds by as-

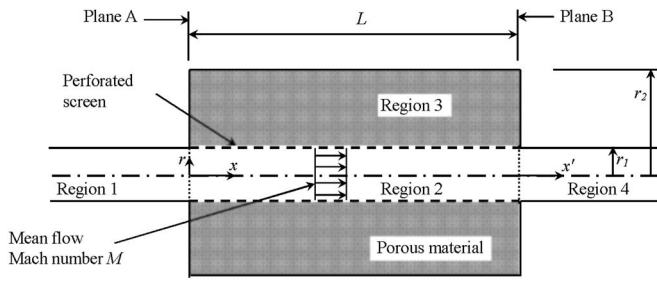


FIG. 1. Geometry of silencer.

suming that the acoustic fields in the inlet/outlet pipes, and also the silencer chamber, may be expanded as an infinite sum over the pipe/silencer eigenmodes. On finding the pipe/silencer eigenfunctions and associated wave numbers, the modal amplitudes are computed by matching analytically over the two axial discontinuities, after suitable truncation of each modal sum.

A. Sound field in the inlet/outlet pipes

The acoustic wave equation in region 1 (or region 4, which is identical) is given by

$$\frac{1}{c_0^2} \frac{D^2 p_1'}{Dt^2} - \nabla^2 p_1' = 0, \quad (1)$$

where c_0 is the isentropic speed of sound in air, p_1' is the acoustic pressure, and t is time. For an axisymmetric silencer, and assuming a time dependence of $e^{i\omega t}$ (where $i = \sqrt{-1}$ and ω is the radian frequency), Eq. (1) may be rewritten as

$$(1 - M^2) \frac{\partial^2 p_1'}{\partial x^2} + \frac{1}{r} \frac{\partial p_1'}{\partial r} + \frac{\partial^2 p_1'}{\partial r^2} - 2iMk_0 \frac{\partial p_1'}{\partial x} + k_0^2 p_1' = 0, \quad (2)$$

where $k_0 = \omega/c_0$ and M is the mean flow Mach number. The sound pressure is now written as an expansion over the pipe eigenmodes to give

$$p_1'(x, r) = \sum_{n=0}^{\infty} F^n \Phi_i^n(r) e^{-ik_0 \lambda_i^n x} + \sum_{n=0}^{\infty} A^n \Phi_r^n(r) e^{-ik_0 \lambda_r^n x}, \quad (3)$$

and

$$p_4'(x', r) = \sum_{n=0}^{\infty} D^n \Phi_i^n(r) e^{-ik_0 \lambda_i^n x'} + \sum_{n=0}^{\infty} E^n \Phi_r^n(r) e^{-ik_0 \lambda_r^n x'}. \quad (4)$$

Here, F^n , A^n , D^n , and E^n are the modal amplitudes, λ_i^n are the incident, and λ_r^n the reflected axial wave numbers, and Φ_i^n are the incident and Φ_r^n the reflected eigenfunctions, in regions 1 and 4, respectively. The wave numbers and eigenfunctions for the incident and reflected sound fields are found by substituting Eqs. (3) and (4) into Eq. (2). For the axial wave numbers, this yields

$$\lambda_{i,r}^n = \left\{ -M \mp \sqrt{1 - (1 - M^2) \left\{ \frac{a_1^n}{k_0 r_1} \right\}^2} \right\} / (1 - M^2). \quad (5)$$

Here, the incident wave is calculated using the minus, and the reflected wave using the plus, before the square root; r_1 is the radius of the pipe in region 1, and a_1^n are solutions of the rigid wall boundary condition $J_1(a_1) = 0$ (J_m is a Bessel function of the first kind and order m). The eigenfunctions are given as

$$\Phi_{i,r}^n(r) = J_0(\gamma_{i,r}^n r), \quad (6)$$

where $\gamma_{i,r}$ is a radial wave number given by $\gamma_{i,r}^n = a_1^n / r_1$.

B. Sound field in the chamber

The acoustic wave equation in region 2 is given by Eq. (2). For region 3, the acoustic wave equation may be written⁶ as

$$\frac{\partial^2 p_3'}{\partial x^2} + \frac{1}{r} \frac{\partial p_3'}{\partial r} + \frac{\partial^2 p_3'}{\partial r^2} - \Gamma^2 p_3' = 0, \quad (7)$$

where Γ is the propagation constant for the porous material. The sound pressure fields in regions 2 and 3 are coupled and written as an expansion over the chamber eigenmodes to give

$$p_c'(x, r) = \sum_{n=0}^{\infty} B^n \Psi_i^n(r) e^{-ik_0 k_i^n x} + \sum_{n=0}^{\infty} C^n \Psi_r^n(r) e^{-ik_0 k_r^n x}. \quad (8)$$

Here, B^n and C^n are the modal amplitudes, k_i^n are the incident and k_r^n the reflected axial wavenumbers, and Ψ_i^n are the incident and Ψ_r^n the reflected eigenfunctions. The substitution of Eq. (8) into Eqs. (2) and (7) allows the radial pressure for a positive traveling wave and for eigenmode n , to be written as

$$p_{2,i}'(r) = P_{2,i}^n J_0(\alpha_i^n r) \quad (9)$$

and

$$p_{3,i}'(r) = P_{3,i}^n \left[J_0(\beta_i^n r) - \frac{Y_0(\beta_i^n r) J_1(\beta_i^n r_2)}{Y_1(\beta_i^n r_2)} \right]. \quad (10)$$

Here, two identical expressions may also be written for a reflected wave. The terms $P_{2,i}^n$ and $P_{3,i}^n$ are constants, r_2 is the radius of the silencer chamber, and Y_m denotes a Neumann function of order m . The radial wave numbers α and β are given by

$$\alpha_{i,r}^n = k_0 \sqrt{1 - 2Mk_{i,r}^n - (1 - M^2)(k_{i,r}^n)^2}, \quad (11)$$

and

$$\beta_{i,r}^n = k_0 \sqrt{-\tilde{\Gamma}^2 - (k_{i,r}^n)^2}, \quad (12)$$

where $\tilde{\Gamma} = \Gamma/k_0$. The eigenequation for the chamber is found by eliminating $P_{2,i}^n$ and $P_{3,i}^n$ from Eqs. (9) and (10) using the appropriate radial boundary conditions over the perforated screen.

In the absence of a perforated screen the accepted (distributed) boundary conditions when uniform mean flow is

present are continuity of pressure and displacement. However, when a perforated screen is present, identifying the correct distributed boundary conditions is more problematic because the screen is treated as an infinitely thin distributed boundary, whereas in reality experimental measurements are taken for a single (discrete) orifice of finite thickness.¹² Moreover, when grazing mean flow is present, impedance measurements include the effects of a viscous boundary layer adjacent to one side of the perforated screen, whereas the model assumes an infinitely thin boundary layer. The pressure condition over the perforated screen is normally written so that the change in pressure (Δp) over the perforated screen is expressed in terms of the measured impedance ζ and the acoustic velocity in the orifice U , to give $\Delta p = \zeta U$. The difficulty lies in identifying the second continuity condition and Kirby^{6,9} retained continuity of displacement, a decision based on the earlier work of Nilsson and Brander;¹ however, this means that the choice of U is ambiguous when defining the pressure condition. Recently, Aurégan and Leroux¹³ used a discrete plane wave approach to model a reactive silencer and examined the second continuity condition under sheared grazing mean flow. They conclude that the correct condition lies somewhere between continuity of displacement and continuity of velocity. Dokumaci¹⁴ proposed using a “slip” velocity to accommodate the departure from an assumed no slip condition at the wall of the perforated screen, and also a term that accounts for the nonuniformity of the actual mean flow profile. Using a distributed model of the perforated screen, Dokumaci¹⁴ demonstrates the effect of these two parameters on transmission loss predictions for a straight through reactive silencer, although no comparisons were made with experimental measurements and so no conclusions regarding appropriate values for these parameters were forthcoming. Clearly, work on reactive silencers illustrates some difficulty in successfully identifying the appropriate second continuity condition for the perforated screen. Furthermore, the model presented here exam-

ines a dissipative silencer and also includes higher order modes, and so it is debatable whether the analysis for reactive silencers is directly applicable to the present case. Nevertheless, in view of the problems seen for reactive silencers the following condition (for eigenmode n , incident or reflected wave) over the perforated screen is specified:

$$\tilde{\rho}[1 - Mk^n]^{-T} \frac{\partial p'_2}{\partial r}(x, r_1) = \frac{\partial p'_3}{\partial r}(x, r_1). \quad (13)$$

Here T is a constant, where $1 \leq T \leq 2$; $T=1$ corresponds to continuity of velocity, and $T=2$ to continuity of displacement. The density of air in region 2 is denoted by ρ_0 and in region 3 the effective (complex) density is denoted by $\rho(\omega)$,⁶ where $\tilde{\rho} = \rho(\omega)/\rho_0$. The pressure condition over the perforated screen is written as

$$p'_2(x, r_1) - p'_3(x, r_1) = \rho_0 c_0 \zeta u'_r(x, r_1), \quad (14)$$

where u'_r is the radial acoustic velocity in region 3. Equation (14) is written in terms of the velocity in region 3, rather than region 2, in order to reflect the acoustic velocity used in the measurements of impedance under the influence of grazing mean flow, see for example Refs. 12 and 15. Thus, although theoretically the choice of velocity here is still ambiguous if one assumes $T=2$, the above-presented discussions suggest that it is appropriate always to choose the velocity used when measuring, and hence defining, the impedance.

Equations (13) and (14) are now combined with Eqs. (9) and (10) to give the following eigenequation for the chamber (for eigenmode n , incident or reflected wave):

$$\left\{ [1 - Mk]^{-T} \frac{J_0(\alpha r_1)}{J_1(\alpha r_1)} + \frac{i\zeta\alpha}{k_0} \right\} - \frac{\tilde{\rho}\alpha}{\beta} \left[\frac{J_1(\beta r_1)Y_1(\beta r_2) - Y_1(\beta r_1)J_1(\beta r_2)}{J_0(\beta r_1)Y_1(\beta r_2) - Y_0(\beta r_1)J_1(\beta r_2)} \right] = 0. \quad (15)$$

The transverse eigenfunction is given by

$$\Psi(r) = \begin{cases} J_0(\alpha r), & 0 \leq r \leq r_1 \\ \frac{\tilde{\rho}\alpha}{\beta} \frac{J_1(\alpha r_1)}{[1 - Mk]^{-T}} \frac{[J_0(\beta r)Y_1(\beta r_2) - Y_0(\beta r)J_1(\beta r_2)]}{[J_1(\beta r_1)Y_1(\beta r_2) - Y_1(\beta r_1)J_1(\beta r_2)]}, & r_1 \leq r \leq r_2 \end{cases}. \quad (16)$$

The incident and reflected wave numbers for the chamber $k_{i,r}^n$ are found by solving eigenequation (15). Here, the roots are found using the Newton-Raphson method, which is discussed in detail in Appendix A.

C. Analytic mode matching

The modal amplitudes in Eqs. (3), (4), and (8) are found by enforcing matching conditions over the inlet ($x=0$) and the outlet ($x=L$) of the silencer. The appropriate axial matching conditions are continuity of pressure and a kinematic

continuity condition that is written here in terms of the constant T , specified in Eq. (13). Accordingly, continuity of pressure over planes A and B yields

$$p'_1(0, r) = p'_2(0, r), \quad 0 \leq r \leq r_1, \quad (17)$$

and

$$p'_2(L, r) = p'_4(0, r), \quad 0 \leq r \leq r_1. \quad (18)$$

The kinematic matching condition yields

$$[1 - M\lambda^n]^{-T} \frac{\partial p'_1}{\partial x}(0, r) = [1 - Mk^n]^{-T} \frac{\partial p'_c}{\partial x}(0, r), \quad 0 \leq r \leq r_1, \quad (19a)$$

$$0 = \frac{1}{\tilde{\rho}} \frac{\partial p'_c}{\partial x}(0, r), \quad r_1 \leq r \leq r_2, \quad (19b)$$

and

$$[1 - M\lambda^n]^{-T} \frac{\partial p'_4}{\partial x}(0, r) = [1 - Mk^n]^{-T} \frac{\partial p'_c}{\partial x}(L, r), \quad 0 \leq r \leq r_1, \quad (20a)$$

$$0 = \frac{1}{\tilde{\rho}} \frac{\partial p'_c}{\partial x}(L, r), \quad r_1 \leq r \leq r_2. \quad (20b)$$

Here, the wave numbers λ and k are written in general; a distinction between the incident and reflected waves is introduced only after differentiating the general expression for pressure.

Mode matching proceeds by choosing a weighting function for each matching condition and then integrating over the cross-sectional area of the inlet/outlet pipe and the silencer chamber. Here, the incident eigenfunction in region 1 is used as the weighting function for Eqs. (17) and (18); the incident eigenfunction in the silencer chamber is used as the weighting function for Eqs. (19) and (20). Accordingly, substituting Eqs. (3), (4), and (8) into the axial matching conditions, applying the weighting functions, and then integrating over the relevant cross section, yields

$$A^n I_{11_r}^{mn} - B^n I_{1C_i}^{mn} - \tilde{C}^n e^{ik_0 k_r^n L} I_{1C_r}^{mn} = -F^0 I_{11_i}^{m0}, \quad (21)$$

$$\begin{aligned} A^n \frac{\lambda_r^n}{[1 - M\lambda_r^n]^{-T}} I_{C1_r}^{mn} - B^n k_i^n I_{CC_i}^{mn} - \tilde{C}^n k_r^n e^{ik_0 k_r^n L} I_{CC_r}^{mn} \\ = -F^0 \frac{\lambda_i^0}{[1 - M\lambda_i^0]^{-T}} I_{C1_i}^{m0}, \end{aligned} \quad (22)$$

$$B^n e^{-ik_0 k_i^n L} I_{1C_i}^{mn} + \tilde{C}^n I_{1C_r}^{mn} - D^n I_{11_i}^{mn} = 0, \quad (23)$$

$$B^n k_i^n e^{-ik_0 k_i^n L} I_{CC_i}^{mn} + \tilde{C}^n k_r^n I_{CC_r}^{mn} - D^n \frac{\lambda_i^n}{[1 - M\lambda_i^n]^{-T}} I_{C1_i}^{mn} = 0. \quad (24)$$

Here,

$$I_{11_{i,r}}^{mn} = \int_0^{r_1} r \Phi_i^m(r) \Phi_{i,r}^n(r) dr, \quad (25)$$

$$I_{C1_{i,r}}^{mn} = I_{1C_{i,r}}^{mn} = \int_0^{r_1} r \Psi_i^m(r) \Phi_{i,r}^n(r) dr, \quad (26)$$

TABLE I. Data for silencers.

Silencer	Length L (mm)	Diameter $2r_2$ (mm)	Absorbent
A	315	152.4	E glass
B	330	203.2	E glass
C	450	152.4	A glass

$$\begin{aligned} I_{CC_{i,r}}^{mn} = [1 - Mk_{i,r}^n]^{-T} \int_0^{r_1} r \Psi_i^m(r) \Psi_{i,r}^n(r) dr \\ + \tilde{\rho}^{-1} \int_{r_1}^{r_2} r \Psi_i^m(r) \Psi_{i,r}^n(r) dr, \end{aligned} \quad (27)$$

and $\tilde{C}^n = C^n e^{-ik_0 k_r^n L}$, for a silencer of length L . These integrals are computed in Appendix B. Equations (21)–(24) assume that a plane wave is incident in region 1 (so that $F^n=0$, for $n>0$, and $F^0=1$) and that region 4 is terminated anechoically (so that $E^n=0$, for all n). This system of equations forms a complete set of $2(N_1+N_C)$ equations and corresponding unknowns (the modal amplitudes), where N_1 and N_C represent the number of modes at which the sums are truncated in regions 1 (or 4) and the chamber, respectively. Finally, the transmission loss (TL) for the silencer is given by

$$\text{TL} = -20 \log_{10} |D^0|, \quad (28)$$

assuming that a plane wave propagates in region 4. It is common also to express silencer performance in terms of four poles so that predictions may be integrated with other components in an exhaust system. Here, the four poles are defined by

$$\begin{bmatrix} p'_1(0) \\ u'_1(0) \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} p'_4(0) \\ u'_4(0) \end{bmatrix}. \quad (29)$$

This gives

$$T_{11} = 2(1 + A^0)/D^0, \quad (30a)$$

$$T_{12} = -Z(1 + A^0)/D^0, \quad (30b)$$

and

$$T_{21} = 2(1 - A^0)/ZD^0, \quad (31a)$$

$$T_{22} = -(1 - A^0)/D^0, \quad (31b)$$

where $Z = \rho_0 c_0$.

III. EXPERIMENT

Experimental measurements are reported here for three straight-through circular dissipative silencers; see Table I for the silencer dimensions ($2r_1=74$ mm). The silencers, and the transmission loss measurements reported here, are identical to those studied by Kirby⁶ and so the reader is referred to Ref. 6 for a detailed description of the experimental methodology. It should be noted, however, that the impulse technique is used here for reasons of available laboratory space, and that the transmission loss measurements are valid only over a limited frequency range of approximately 150–1500

TABLE II. Values of material constants.

Constant	E glass	A glass
a_1	0.2202	0.2251
a_2	-0.5850	-0.5827
a_3	0.2010	0.1443
a_4	-0.5829	-0.7088
a_5	0.0954	0.0924
a_6	-0.6687	-0.7177
a_7	0.1689	0.1457
a_8	-0.5707	-0.5951
Θ (MKS R/m)	30716	5976
Ω	0.952	0.952
q_0^2	5.49	3.77
ξ_0	0.005	0.025

Hz. At low frequencies, experimental errors appear because of reflections from the end of the apparatus. At high frequencies, flow noise increases and is seen to cause fluctuations in the measured data because the signal to noise ratio has dropped. This problem is exacerbated by a reduction in the sound power delivered by a rectangular pulse at higher frequencies coinciding with the frequency range in which the silencer works most effectively. It is possible that these problems may be reduced by using an alternative method, such as two-microphone method (see, for example, Ref. 15), and by using error analysis to suppress flow noise.¹⁶ Problems with flow noise are, however, still evident in the measurements of Lee and Ih,¹⁵ and obtaining accurate transmission loss predictions over a wide frequency range when mean flow is present still represents a considerable challenge, which is reflected in the relative dearth of experimental data for dissipative silencers with mean flow in the literature.

The bulk acoustic properties of the materials in each silencer are the same as those in Ref. 6, and so the normalized propagation constant is given by

$$\tilde{\Gamma} = a_1 \xi^{-a_2} + i(1 + a_3 \xi^{-a_4}), \quad (32)$$

and the normalized complex density by

$$\tilde{\rho} = -\tilde{\Gamma}[a_5 \xi^{-a_6} + i(1 + a_7 \xi^{-a_8})]. \quad (33)$$

Here, the constants a_1, \dots, a_8 are Delany and Bazley coefficients that are given for each material in Table II; ξ is a nondimensional frequency parameter given by $\xi = \rho_0 f / \Theta$, where Θ is the material flow resistivity (see Table II). A low frequency correction, discussed in detail by Kirby^{6,9} is also used here, which requires the material porosity Ω , the “steady flow” tortuosity q_0^2 , and a so-called transition value for ξ , known as ξ_0 ; values for these parameters are also listed in Table II.

Values for the impedance of the perforated screen depend on experimental measurements. Kirby^{6,9} used the semi-empirical model proposed by Kirby and Cummings¹² to quantify the impedance of a perforated screen under grazing flow. These results have since been reviewed by other authors and it is apparent that values for the impedance may depend on the degree to which the viscous boundary layer in the test rig has developed by the time the flow reaches a

perforation. The impedance data proposed by Kirby and Cummings were measured on the same test rig as the transmission loss predictions presented here, and so it appears sensible to retain these results in this current study. However, it is interesting also to review how other measured impedance data affect the transmission loss predictions and so the data of Lee and Ih¹⁵ will also be reviewed here. Kirby and Cummings¹² also measured the effect of a fibrous material on the impedance of the perforated screen and suggested adding a correction, based on replacing the density of air with the effective complex density of the fibrous material, to the expression for the impedance. Lee *et al.*¹⁷ recently investigated this correction by using detailed experimental measurements without mean flow, and concluded that this is indeed capable of capturing the effect of the absorbing material on the perforate impedance. Accordingly, this correction is retained here, although Denia *et al.*¹⁸ suggest modifying this correction by introducing the parameter $F(\sigma)$, where

$$F(\sigma) = 1 - 1.06\sigma^{0.5} + 0.17\sigma^{1.5}. \quad (34)$$

Here, σ is the open area porosity of the perforated screen and the normalized impedance ζ is then given as

$$\zeta = [\zeta' + i0.425k_0d(\tilde{\rho} - 1)F(\sigma)]/\sigma, \quad (35)$$

where d is the diameter of the hole and ζ' is the orifice impedance measured in the absence of a porous material. In view of the good correlation between prediction and experiment observed by Denia *et al.*,¹⁸ $F(\sigma)$ will be retained here and all transmission loss calculations that include a perforated screen will use Eq. (35). Values for ζ' measured by Kirby and Cummings¹² (see also Refs. 6 and 9) and Lee and Ih¹⁵ will be used here with a perforated screen that has a thickness of 1 mm, a hole diameter of 3.5 mm, and an open area porosity of 26.3%.

IV. RESULTS AND DISCUSSION

A. Axial continuity condition

Here, changing the axial kinematic matching condition from continuity of velocity (the method used by Cummings and Chang⁴) to continuity of displacement is investigated. In principle, both conditions should deliver identical predictions when used in a mode matching scheme applied to the silencers studies here, even when mean flow is present; however, the results presented will demonstrate that problems arise with Cummings and Chang’s method. In Ref. 9 it is argued that mode matching depends on finding a convergent system of equations that form a transfer matrix, \mathbf{S} , whose elements S_{ij} decay rapidly with increasing i, j . Moreover, without such a property, one cannot guarantee that the final solution will reflect the physics of the problem. When mean flow is neglected, Eq. (27) provides an orthogonality relation that guarantees a convergent system of equations; however, when mean flow is present the silencer eigenfunctions are not orthogonal and Eq. (27) no longer provides a true orthogonality relation. Accordingly, in the absence of an orthogonality relation when mean flow is present, one must be careful first to establish a convergent system of equations before one can be sure the solution reflects the physics of the problem. To

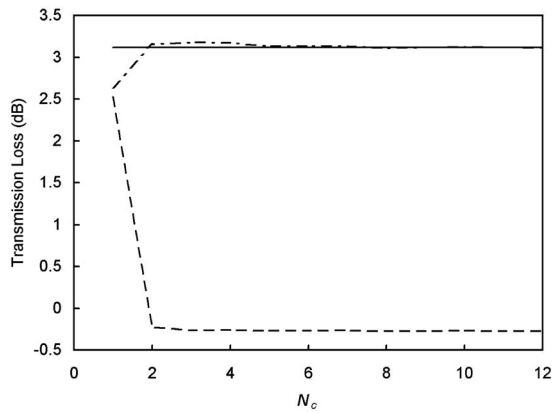


FIG. 2. Predicted transmission loss for Xu *et al.* (Ref. 10) silencer, for $F = 40$ Hz and $M=0.15$ ($N_1=1$): (—) Finite element method (Ref. 8); (---) Cummings and Chang's method (Ref. 4); (-·-) current method with $T=2$.

investigate this issue for dissipative silencers, a number of different geometries are examined here. First, in Fig. 2 convergence is examined for the silencer studied by Xu *et al.*¹⁰ For this silencer, the inner radius of the porous material is coincident with the radius of the inlet/outlet pipe, the excitation frequency is 40 Hz, the mean flow Mach number is 0.15, and a perforated screen is omitted. Transmission loss predictions using the new approach (with $T=2$) are compared with Cummings and Chang's method⁴ and the finite element method⁸ (with 1545 degrees of freedom); for the latter two models internal mean flow is neglected. A problem with convergence for Cummings and Chang's method is clearly evident in Fig. 2, and nonphysical predictions occur when higher order modes are added. In contrast, the new method is seen to converge smoothly toward the (benchmark) finite element predictions. In Fig. 3, transmission loss predictions for the silencer of Xu *et al.* are presented over a range of frequencies and the new mode matching predictions (with $N_c=8$) compare very well with the finite element method, especially at low frequencies. Conversely, problems with Cummings and Chang's method are obvious in the form of predictions that are too low (in some instances negative)

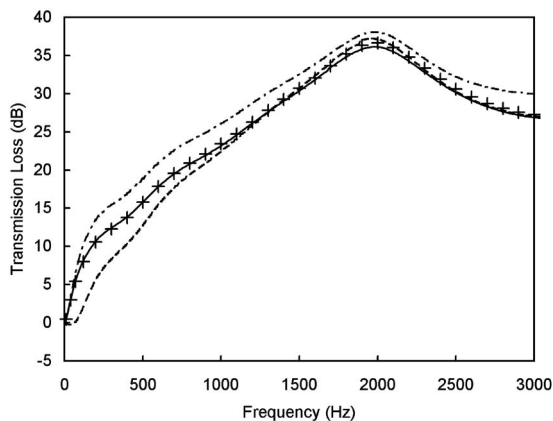


FIG. 3. Predicted transmission loss for Xu *et al.* (Ref. 10) silencer: (+) Finite element method (Ref. 8), $M=0.15$; (---) current method, $M=0.15$; (-·-) Cummings and Chang's method (Ref. 4), $M=0.15$; (---) current method $M=0$.

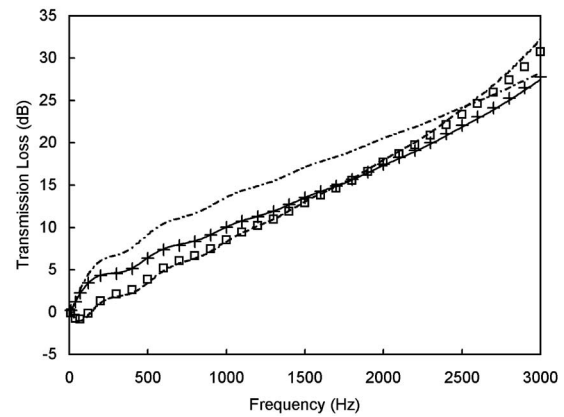


FIG. 4. Predicted transmission loss for Cummings and Chang's (Ref. 4) silencer (with no internal flow): (+) Finite element method (Ref. 8), $M=0.196$; (---) current method, $M=0.196$; (-·-) Cummings and Chang's method (Ref. 4), $M=0.196$; (---) current method $M=0$; (□) point collocation method (Ref. 9), $M=0.196$.

at low frequencies, although better agreement with the finite element predictions is observed at higher frequencies.

The problems observed with Cummings and Chang's method in Figs. 2 and 3 are thought to be caused by the absence of a convergent system of equations when different axial and radial kinematic matching conditions are applied at the silencer edges (at $r=r_1$, for $x=0$ and $x=L$). On enforcing the same kinematic boundary condition, it is evident that predictions now agree well with finite elements and this is thought to be because the new boundary condition delivers a transfer matrix, \mathbf{S} , whose elements S_{ij} decay rapidly with increasing i, j . Further evidence to support this is provided in Fig. 4, which compares transmission loss predictions for Cummings and Chang's silencer (for $M=0.196$, perforated screen and internal mean flow omitted). Here, Cummings and Chang's method is again seen to provide predictions that are too low over an important part of the frequency range, whereas the new method agrees well with the finite element predictions. Also included in Fig. 4 are point collocation predictions obtained using the method of Kirby.⁹ The point collocation method is a numerical version of mode matching and the method presented in Ref. 9 is based on the same axial kinematic matching condition used by Cummings and Chang. Accordingly, one would expect the two methods to provide similar predictions and in Fig. 4 this is seen to be the case, even at low frequencies. Here, the failure of Kirby's method at low frequencies is thought to be for the same reasons as that seen for Cummings and Chang's method, namely the failure to deliver a convergent system of equations. This does not, however, explain why the predictions presented by Kirby⁹ provided good agreement with experimental data and why agreement improves at higher frequencies in Figs. 3 and 4. To investigate this further, point collocation predictions are presented in Fig. 5 for silencer A ($M=0.15$, $u_*=2.56$ m/s), and these are shown to agree well with predictions generated using the new method ($T=2$), even at low frequencies (note that the point collocation predictions overlay those obtained using Cummings and Chang's method). This good agreement is thought to be because Cummings and Chang's transfer matrix is sufficiently

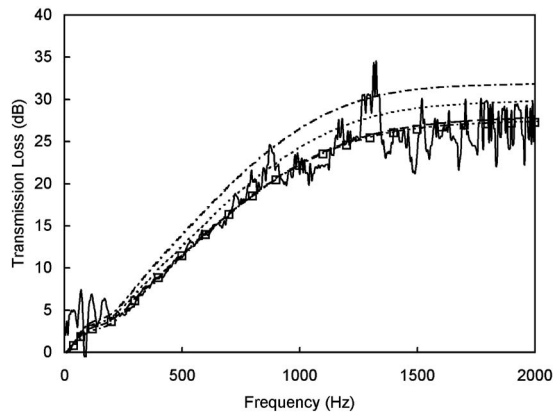


FIG. 5. Predicted and measured transmission loss for silencer A, with $M=0.15$: (—) Experiment; (---) prediction with $T=2$; (···) prediction with $T=1.5$; (-·-) prediction with $T=1$; (□) point collocation (Ref. 9); (-·-·) Cummings and Chang method (Ref. 4).

convergent to provide sensible predictions for those materials with a high flow resistivity. Furthermore, similar convergent conditions also exist for the silencer transfer matrix when materials of low flow resistivity are present, but only at high frequencies (see Figs. 3 and 4). Similar behavior is observed, as one would expect, for the point collocation results in Figs. 4 and 5, and this indicates that the reason for the good agreement between prediction and measurement is the high material flow resistivity used in the silencers studied by Kirby.⁹ It appears, therefore, that in order properly to reflect the physics of the problem over a wide range of parameters, a true orthogonality relation is not required provided that one carefully chooses the kinematic matching conditions in order to realize a convergent system of equations. Accordingly, when mean flow is present, the results presented here indicate that if continuity of displacement is used as a radial boundary condition, then the correct axial continuity condition is also displacement. But, if the radial boundary condition changes, say to continuity of velocity, then the axial boundary condition must also change. This conclusion is based on numerous numerical experiments, although only a sample is shown in this paper. In the current study, equality of the kinematic matching conditions is enforced by the use of the constant T , and all further transmission loss predictions presented here will use this new approach.

B. Boundary conditions for the perforated screen

The boundary conditions for the perforated screen are defined by a pressure condition, Eq. (14), and a kinematic continuity condition, Eq. (13). The constant T allows for the alteration of the kinematic continuity condition and in Figs. 5–7 mode matching predictions are presented (with $N_1=1$, $N_c=8$) for values of T of 1, 1.5, and 2, for silencers A, B, and C, respectively. Here, $M=0.15$ and a perforated screen with the impedance specified by Kirby and Cummings¹² is used in Eq. (35) (with $u_*=2.56$ m/s). In Figs. 5–7 it is evident that changing the axial and radial kinematic boundary conditions has a significant effect on transmission loss: In each case, predictions for continuity of velocity ($T=1$) are higher than those for continuity of displacement ($T=2$), and, not surpris-

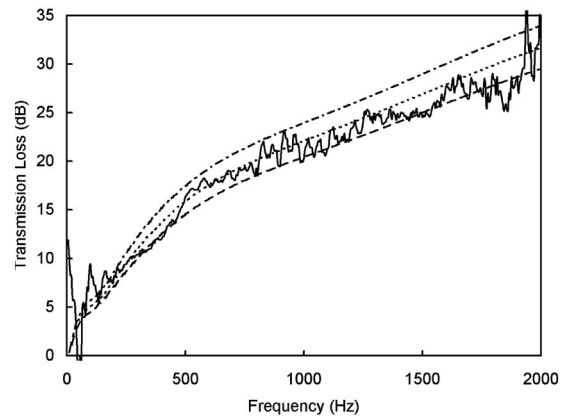


FIG. 6. Predicted and measured transmission loss for silencer B, with $M=0.15$: (—) Experiment; (---) prediction with $T=2$; (···) prediction with $T=1.5$; (-·-) prediction with $T=1$.

ingly, predictions for $T=1.5$ bisect these two extremes. For silencers A and C, continuity of displacement provides good agreement with measurement, although for silencer B some underprediction is evident. In contrast, continuity of velocity consistently overpredicts the transmission loss for all three silencers. It is evident that by altering the value for T , improvements in the correlation between prediction and experiment may be obtained for silencer B; however, there does not appear to be a consistent value for T that provides good agreement for all silencers. Thus, in view of these results, and the successful use of continuity of displacement in other dissipative silencer models that assume uniform mean flow, it appears prudent here to retain continuity of displacement as the kinematic boundary condition. Accordingly, this condition is used in the following section, which reviews two different impedance models. We note, however, that more supporting evidence in the form of further transmission loss measurements is required before one can be certain of the correct boundary conditions for the perforated screen.

C. Predicted and measured transmission loss

This section presents comparisons between transmission loss measurements and predictions found using the impedance data for a perforated screen measured by Kirby and

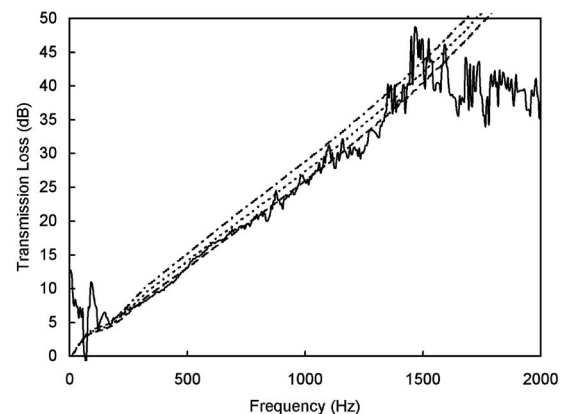


FIG. 7. Predicted and measured transmission loss for silencer C, with $M=0.15$: (—) Experiment; (---) prediction with $T=2$; (···) prediction with $T=1.5$; (-·-) prediction with $T=1$.

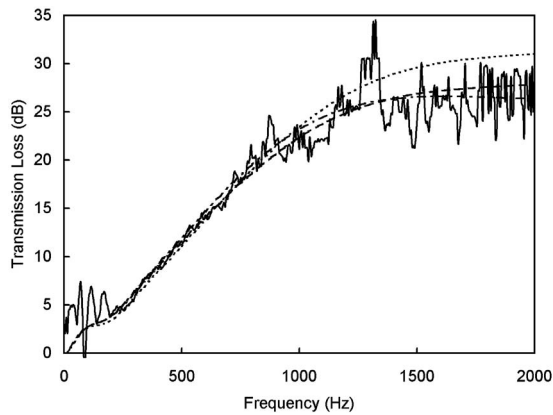


FIG. 8. Predicted and measured transmission loss for silencer A, with $M=0.15$: (—) Experiment; (---) predictions using Kirby and Cummings' perforate data (Ref. 12); (-·-) predictions using Lee and Ih's perforate data (Ref. 15); (···) prediction with no perforated screen.

Cummings,¹² and Lee and Ih.¹⁵ In both cases, a Mach number of $M=0.15$ is present (with $u_s=2.56$ m/s for Ref. 12), $T=2$, $N_1=1$, and $N_c=8$. In Fig. 8 a comparison is presented for silencer A, which has a screen of porosity of $\sigma=0.263$. This porosity lies outside Lee and Ih's suggested limits of $0.0279 \leq \sigma \leq 0.223$; however, the difference here is small and, as the influence of the perforate will reduce at higher porosities, any inaccuracies associated with this extrapolation are likely to have a negligible effect on the transmission loss predictions. In general, good agreement between prediction and experiment is observed in Fig. 8, and it is noticeable that the two impedance models give largely similar predictions over the frequency range shown. The effect of the perforated screen is clearly evident, especially at higher frequencies, and for the silencers studied here an improvement in the agreement between prediction and experiment is observed when a perforated screen is included. It is, however, difficult to arrive at conclusions regarding the most appropriate impedance model, especially as one could argue that the data here are biased toward the method of Kirby and Cummings since the transmission loss measurements were obtained under the same mean flow conditions as the impedance measurements. Further observations regarding an appropriate choice for impedance data awaits additional silencer measurements, especially measurements that are reliable at much higher frequencies. Nevertheless, the new mode matching approach is demonstrated here to work well, and it is noticeable that there is no underprediction of transmission loss at low frequencies, further supporting the change in the axial boundary condition.

V. CONCLUSIONS

When mean flow is added to the central airway of a straight-through circular dissipative silencer, the analytic mode matching schemes developed without flow may readily be extended to include flow. The mode matching scheme presented here uses the silencer eigenfunctions as weighting functions and good agreement is observed among the mode matching predictions, experimental data, and benchmark finite element calculations. The main result of this study is that

if mean flow and higher order modes are included in the matching scheme, the axial kinematic continuity condition must be the same as the radial kinematic boundary condition chosen for the interface between the airway and the material (regardless of whether a perforated screen is present or not). That is, if continuity of pressure and displacement are chosen for the radial boundary conditions then the axial matching conditions should also be continuity of pressure and displacement, rather than pressure and velocity as previously thought. In this study, equality for the axial and radial kinematic boundary conditions is enforced using a constant T , where using $T=2$ enforces continuity of displacement and $T=1$ continuity of velocity. On adding a perforated screen, the appropriate radial kinematic boundary condition is reviewed by varying T . Here, good agreement between prediction and experiment is observed using continuity of displacement, whereas the transmission loss tends to be over-predicted when using continuity of velocity. Finally, two different sets of impedance data for the perforated screen were reviewed and it is demonstrated that similar transmission loss predictions are obtained at low to medium frequencies when using the data of Kirby and Cummings¹² and Lee and Ih.¹⁵

APPENDIX A: ROOT FINDING

The roots of Eq. (15) provide the incident and reflected axial wave numbers in the silencer chamber. There are numerous methods available for finding these roots, although in the design of dissipative silencers three methods have found favor: the Newton-Raphson method,² Muller's method,¹⁹ and the Secant method.¹¹ Each method has its own advantages and disadvantages; however, a problem common to all root finding techniques is the possibility of missing roots. This happens because each method depends on inputting an initial guess, or guesses, and then iterating toward the desired root. If the root is complex (as it is for dissipative silencers) one cannot guarantee that the initial guess is close enough to locate a desired root. This problem is further exacerbated when attempting to track roots over a frequency sweep, as it can be readily shown that roots may jump position in the complex plane. Selamet *et al.*¹¹ adopt the Secant method in order to find the roots of an eigenequation that is similar (but without mean flow) to Eq. (15) and use initial guesses based on limiting values at high frequencies. This procedure was found to succeed for the silencer geometries and material parameters chosen in their study, but in general an iterative method does not, and cannot, guarantee to find all desired roots over a specified region in the complex plane unless one first uses the Argument Principle²⁰ to see if any roots have been missed, and then laboriously searches smaller segments of the complex plane, again with the aid of the Argument Principle, in order to locate a missing root. This procedure must be repeated for every root and every time a parameter, such as frequency, is changed. This issue is discussed in more detail by Lawrie and Kirby,²¹ who avoid root finding altogether, although this method depends on writing an orthogonality relation for the system and so is not currently applicable when mean flow is present. The salient point,

however, is that all root finding techniques suffer from the problem of missing roots and one must be wary of this when using an analytic mode matching scheme such as the one reported here. The authors' preference is to use the Argument Principle as a check to make sure all desired roots have been found, but if roots have been missed then to refine the initial guess(es) (for example, by increasing the number of guesses) rather than laboriously searching the complex plane for the missing root. A further complication in the current analysis is the presence of mean flow. Here, the authors have found that it is harder to track and locate roots when mean flow is present. The authors' preferred method for finding roots is the Newton-Raphson method, as this method converges faster than the secant method and requires only one initial guess, unlike Muller's method. The Newton-Raphson method does, however, require the derivative of the governing eigenequation to be found. To facilitate this, Eq. (15) is recast here to give

$$G(k) = \frac{\beta}{\alpha} [1 - Mk]^T (b_1 - b_4) + \frac{i\zeta\beta}{k_0} (h_1 - h_2) - \tilde{\rho} (h_3 - h_6). \quad (\text{A1})$$

The derivative is then given by

$$G'(k) = G_1(k) - G_2(k), \quad (\text{A2})$$

where

$$G_1(k) = \frac{k_0^2}{\alpha^3 \beta} [1 - Mk]^T (g_{11} - g_{12}) - \frac{\beta}{\alpha} TM [1 - Mk]^{T-1} (b_1 - b_4), \quad (\text{A3})$$

and

$$G_2(k) = \frac{i\zeta k_0}{\alpha^2 \beta} (g_{21} - g_{22}) + \frac{\tilde{\rho} k_0^2}{\alpha^2 \beta^2} (g_{31} - g_{32}). \quad (\text{A4})$$

Here

$$g_{11} = \alpha^2 k b_1 + s(b_1 + \alpha r_1 h_1) - \alpha^2 \beta k (r_1 b_2 + r_2 b_3), \quad (\text{A5})$$

$$g_{12} = \alpha^2 k b_4 + s(b_4 + \alpha r_1 h_2) - \alpha^2 \beta k (r_2 b_5 + r_1 b_6), \quad (\text{A6})$$

$$g_{21} = -\alpha^2 k h_1 + s(\alpha r_1 b_1 - h_1) + \alpha^2 \beta k (r_1 h_3 + r_2 h_4), \quad (\text{A7})$$

$$g_{22} = -\alpha^2 k h_2 + s(\alpha r_1 b_4 - h_2) + \alpha^2 \beta k (r_2 h_5 + r_1 h_6), \quad (\text{A8})$$

$$g_{31} = \alpha^2 k (\beta [r_1 h_1 - r_2 h_7] + h_3) - s(\alpha r_1 b_2 - h_3), \quad (\text{A9})$$

$$g_{32} = \alpha^2 k (\beta [r_1 h_2 - r_2 h_8] + h_6) - s(\alpha r_1 b_6 - h_6); \quad (\text{A10})$$

and

$$s = \beta^2 [M + (1 - M^2)k], \quad (\text{A11})$$

$$b_1 = J_0(\alpha r_1) J_1(\beta r_1) Y_1(\beta r_2), \quad (\text{A12})$$

$$b_2 = J_0(\alpha r_1) J_0(\beta r_1) Y_1(\beta r_2), \quad (\text{A13})$$

$$b_3 = J_0(\alpha r_1) J_1(\beta r_1) Y_0(\beta r_2), \quad (\text{A14})$$

$$b_4 = J_0(\alpha r_1) J_1(\beta r_2) Y_1(\beta r_1), \quad (\text{A15})$$

$$b_5 = J_0(\alpha r_1) J_0(\beta r_2) Y_1(\beta r_1), \quad (\text{A16})$$

$$b_6 = J_0(\alpha r_1) J_1(\beta r_2) Y_0(\beta r_1), \quad (\text{A17})$$

$$h_1 = J_1(\alpha r_1) J_1(\beta r_1) Y_1(\beta r_2), \quad (\text{A18})$$

$$h_2 = J_1(\alpha r_1) J_1(\beta r_2) Y_1(\beta r_1), \quad (\text{A19})$$

$$h_3 = J_1(\alpha r_1) J_0(\beta r_1) Y_1(\beta r_2), \quad (\text{A20})$$

$$h_4 = J_1(\alpha r_1) J_1(\beta r_1) Y_0(\beta r_2), \quad (\text{A21})$$

$$h_5 = J_1(\alpha r_1) J_0(\beta r_2) Y_1(\beta r_1), \quad (\text{A22})$$

$$h_6 = J_1(\alpha r_1) J_1(\beta r_2) Y_0(\beta r_1), \quad (\text{A23})$$

$$h_7 = J_1(\alpha r_1) J_0(\beta r_1) Y_0(\beta r_2), \quad (\text{A24})$$

$$h_8 = J_1(\alpha r_1) J_0(\beta r_2) Y_0(\beta r_1). \quad (\text{A25})$$

Roots of Eq. (A1) are found by supplying a number of initial guesses to the Newton-Raphson scheme in order to locate N eigenmodes for the incident or reflected wave. The methodology used here begins at a low frequency, say at 20 Hz, and starts with two initial guesses based on the low frequency model of Kirby.⁶ Then $8N$ further guesses are added, and these are:

$$k_{i,r}^n = \left\{ -M \mp \sqrt{\left[1 - (1 - M^2) \left\{ \frac{a_0^n}{k_0 r_1} \right\}^2 \right]} \right\} / (1 - M^2), \quad (\text{A26})$$

$$k_{i,r}^n = \left\{ -M \mp \sqrt{\left[1 - (1 - M^2) \left\{ \frac{a_0^n}{k_0 r_2} \right\}^2 \right]} \right\} / (1 - M^2), \quad (\text{A27})$$

$$k_{i,r}^n = \pm \sqrt{-\tilde{\Gamma}^2 + \left\{ \frac{a_0^n}{k_0 r_1} \right\}^2}, \quad (\text{A28})$$

$$k_{i,r}^n = \pm \sqrt{-\tilde{\Gamma}^2 + \left\{ \frac{a_0^n}{k_0 r_2} \right\}^2}, \quad (\text{A29})$$

$$k_{i,r}^n = \pm \sqrt{-\tilde{\Gamma}^2 + \left\{ \frac{b_0^n}{k_0 r_1} \right\}^2}, \quad (\text{A30})$$

$$k_{i,r}^n = \pm \sqrt{-\tilde{\Gamma}^2 + \left\{ \frac{b_0^n}{k_0 r_2} \right\}^2}, \quad (\text{A31})$$

$$k_{i,r}^n = \pm \left(2M - i \frac{b_0^n}{k_0 r_1} \right), \quad (\text{A32})$$

$$k_{i,r}^n = \pm \left(2M - i \frac{b_0^n}{k_0 r_2} \right). \quad (\text{A33})$$

Here, a_0^n and b_0^n are solutions of $J_0(a_0)=0$ and $Y_0(b_0)=0$, respectively. After finding all the required roots (N incident, N reflected) at the starting frequency, the frequency is then incremented upwards, say by 10 Hz, and those roots found at the starting frequency are then used as the initial guesses for the following frequency. Crucially, however, the additional $8N$ guesses previously mentioned are also used, thus increasing the number of initial guesses to $9N$. For frequencies that follow, those N roots found at the previous frequency, along with the $8N$ guesses in Eqs. (A26)–(A33), are used as initial guesses. Accordingly, this approach adopts the tactic of using a large number of initial guesses to provide a high probability that all required roots are found. Obviously, this requires a method that converges quickly, hence the use of Newton-Raphson, and repeated roots must also be filtered out. The

authors do not claim that this approach represents an optimum, in terms of number and/or value of the guesses chosen; instead, this approach has been arrived at using trial and error for a large number of silencer configurations (including differing mean flow Mach numbers and materials) and is deliberately overspecified in order to give the best possible chance of locating all required roots. On locating N incident and reflected axial wave numbers $k_{x,i,r}$, these are then sorted in ascending order of the imaginary part for use in the mode matching scheme that follows.

APPENDIX B: EIGENFUNCTION INTEGRALS

The integrals defined in Eqs. (25)–(27) are given by

$$I_{11,i,r}^{mn} = \begin{cases} 0, & m \neq n \\ \frac{r_1^2}{2} J_0(\gamma_i^m r_1) J_0(\gamma_{i,r}^m r_1), & m = n \end{cases}, \quad (\text{B1})$$

$$I_{1C,i,r}^{mn} = \begin{cases} \frac{r_1}{(\alpha_{i,r}^n)^2 - (\gamma_i^n)^2} [\alpha_{i,r}^n J_1(\alpha_{i,r}^n r_1) J_0(\gamma_i^n r_1)], & (\alpha_{i,r}^n)^2 \neq (\gamma_i^n)^2 \\ \frac{r_1^2}{2} J_0(\alpha_{i,r}^n r_1) J_0(\gamma_i^n r_1), & (\alpha_{i,r}^n)^2 = (\gamma_i^n)^2 \end{cases}. \quad (\text{B2})$$

For $m=n$,

$$I_{CC,i}^{mn} = \frac{r_1^2 [J_0^2(\alpha_i^n r_1) + J_1^2(\alpha_i^n r_1)]}{2[1 - Mk_i^n]^T} + \frac{\tilde{\rho} r_2^2 (\alpha_i^n)^2}{2 (\beta_i^n)^2 [1 - Mk_i^n]^{2T}} \left\{ t_3^n - \frac{r_1^2}{r_2^2} t_4^n \right\}, \quad (\text{B3})$$

and

$$I_{CC,r}^{mn} = \begin{cases} I_{CC,i}^{mn}, & M = 0 \\ \frac{t_{1,r}^{mn} r_1}{[1 - Mk_r^n]^T} - \frac{t_{2,r}^{mn} r_1}{[(\beta_r^n)^2 - (\beta_i^n)^2]}, & M > 0 \end{cases}. \quad (\text{B4})$$

For $m \neq n$,

$$I_{CC,i,r}^{mn} = \frac{t_{1,i,r}^{mn} r_1}{[1 - Mk_{i,r}^n]^T} - \frac{t_{2,i,r}^{mn} r_1}{[(\beta_{i,r}^n)^2 - (\beta_i^m)^2]}. \quad (\text{B5})$$

Here,

$$t_{1,i,r}^{mn} = |\alpha_{i,r}^n J_1(\alpha_{i,r}^n r_1) J_0(\alpha_i^m r_1) - \alpha_i^m J_0(\alpha_{i,r}^n r_1) J_1(\alpha_i^m r_1)| [(\alpha_{i,r}^n)^2 - (\alpha_i^m)^2], \quad (\text{B6})$$

$$t_{2,i,r}^{mn} = \frac{\alpha_{i,r}^n J_1(\alpha_{i,r}^n r_1) J_0(\alpha_i^m r_1)}{[1 - Mk_{i,r}^n]^T} - \frac{\alpha_i^m J_0(\alpha_{i,r}^n r_1) J_1(\alpha_i^m r_1)}{[1 - Mk_i^n]^T}, \quad (\text{B7})$$

$$t_3^n = \frac{[J_0(\beta_i^n r_2) Y_1(\beta_i^n r_2) - J_1(\beta_i^n r_2) Y_0(\beta_i^n r_2)]^2}{[J_1(\beta_i^n r_1) Y_1(\beta_i^n r_2) - J_1(\beta_i^n r_2) Y_1(\beta_i^n r_1)]^2}, \quad (\text{B8})$$

$$t_4^n = 1 + \frac{[J_0(\beta_i^n r_1) Y_1(\beta_i^n r_2) - J_1(\beta_i^n r_2) Y_0(\beta_i^n r_1)]^2}{[J_1(\beta_i^n r_1) Y_1(\beta_i^n r_2) - J_1(\beta_i^n r_2) Y_1(\beta_i^n r_1)]^2}. \quad (\text{B9})$$

¹B. Nilsson and O. Brander, "The propagation of sound in cylindrical ducts with mean flow and bulk reacting lining. I. Modes in an infinite duct," *J. Inst. Math. Appl.* **26**, 269–298 (1980).

²A. Cummings and I. J. Chang, "Internal mean flow effects on the characteristics of bulk-reacting liners in circular ducts," *Acustica* **64**, 170–178 (1987).

³R. J. Astley and A. Cummings, "A finite element scheme for attenuation in ducts lined with porous material: Comparison with experiment," *J. Sound Vib.* **116**, 239–263 (1987).

⁴A. Cummings and I. J. Chang, "Sound attenuation of a finite length dissipative flow duct silencer with internal mean flow in the absorbent," *J. Sound Vib.* **127**, 1–17 (1988).

⁵K. S. Peat, "A transfer matrix for an absorption silencer element," *J. Sound Vib.* **146**, 353–360 (1991).

⁶R. Kirby, "Simplified techniques for predicting the transmission loss of a circular dissipative silencer," *J. Sound Vib.* **243**, 403–426 (2001).

⁷S. N. Panigrahi and M. L. Munjal, "Comparison of various methods for analyzing lined circular ducts," *J. Sound Vib.* **285**, 905–923 (2005).

⁸K. S. Peat and K. L. Rathi, "A finite element analysis of the convected acoustic wave motion in dissipative silencers," *J. Sound Vib.* **184**, 527–545 (1995).

⁹R. Kirby, "Transmission loss predictions for dissipative silencers of arbitrary cross section in the presence of mean flow," *J. Acoust. Soc. Am.* **114**, 200–209 (2003).

¹⁰M. B. Xu, A. Selamet, I. J. Lee, and N. T. Huff, "Sound attenuation in dissipative expansion chambers," *J. Sound Vib.* **272**, 1125–1133 (2004).

¹¹A. Selamet, M. B. Xu, I. J. Lee, and N. T. Huff, "Analytical approach for

- sound attenuation in perforated dissipative silencers," *J. Acoust. Soc. Am.* **115**, 2091–2099 (2004).
- ¹²R. Kirby and A. Cummings, "The impedance of perforated plates subjected to grazing mean flow and backed by porous media," *J. Sound Vib.* **217**, 619–636 (1988).
- ¹³Y. Aurégan and M. Leroux, "Failures in the discrete models for flow duct with perforations: An experimental investigation," *J. Sound Vib.* **265**, 109–121 (2003).
- ¹⁴E. Dokumaci, "Effect of sheared grazing mean flow on acoustic transmission in perforated pipe mufflers," *J. Sound Vib.* **283**, 645–663 (2005).
- ¹⁵S. H. Lee and J. G. Ih, "Empirical model of the acoustic impedance of a circular orifice in grazing mean flow," *J. Acoust. Soc. Am.* **114**, 98–113 (2003).
- ¹⁶M. Åbom and H. Bodén, "Error analysis of two-microphone measurements in ducts with flow," *J. Acoust. Soc. Am.* **83**, 2429–2438 (1988).
- ¹⁷I. Lee, A. Selamet, and N. T. Huff, "Impact of perforation impedance on the transmission loss of reactive and dissipative silencers," *J. Acoust. Soc. Am.* **120**, 3706–3713 (2006).
- ¹⁸F. D. Denia, A. Selamet, F. J. Fuenmayor, and R. Kirby, "Acoustic attenuation performance of perforated dissipative mufflers with empty inlet/outlet extensions," *J. Sound Vib.* **302**, 1000–1017 (2007).
- ¹⁹A. Cummings and N. Sormaz, "Acoustic attenuation in dissipative splitter silencers containing mean fluid flow," *J. Sound Vib.* **168**, 209–227 (1993).
- ²⁰R. Kirby and J. B. Lawrie, "A point collocation approach to modelling large dissipative silencers," *J. Sound Vib.* **286**, 313–339 (2005).
- ²¹J. B. Lawrie and R. Kirby, "Mode matching without root finding: Application to a dissipative silencer," *J. Acoust. Soc. Am.* **119**, 2050–2061 (2006).

Evaluating noise in social context: The effect of procedural unfairness on noise annoyance judgments

Eveline Maris,^{a,b)} Pieter J. Stallen,^{a,c)} Riel Vermunt, and Herman Steensma
Faculty of Social and Behavioral Sciences, Universiteit Leiden, Section of Social and Organizational Psychology, P.O. Box 9555, 2300 RB Leiden, The Netherlands

(Received 23 January 2007; revised 31 July 2007; accepted 25 September 2007)

General dosage-response curves typically over- or underestimate the actual prevalence of noise annoyance for specific groups of individuals. The present study applies a social psychological approach to noise annoyance that helps to understand and predict collective deflections from the curve. The approach holds that being exposed to man-made sound is more than mere exposure; it is a social experience, too: You expose Me. In effect, social aspects of the situation, like the evaluation of the sound management procedure, influence the evaluation of sound. The laboratory experiment ($N=90$) investigates the effect of procedural unfairness on noise annoyance. The sound management procedure is varied systematically: Participants are promised they will listen to the sound of their choice (i.e., bird song, radio sound, or aircraft sound) but receive aircraft sound despite their expressed preference (unfair procedure), or they are simply told they will listen to aircraft sound (neutral procedure). All are exposed to aircraft sound (50 or 70 dBA Leq). A collective rise in noise annoyance is predicted in the unfair relative to the neutral procedure conditions. Results show that noise annoyance ratings are significantly higher in the unfair relative to the neutral conditions. Consequences for theory and practice are discussed.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799901]

PACS number(s): 43.50.Qp [BSF]

Pages: 3483–3494

I. INTRODUCTION

In the field of noise annoyance, it is not uncommon to find that descriptive dosage-response curves over- or underestimate actual noise annoyance levels. Nevertheless, governmental decisions on the location of airports and highway infrastructures, as well as on the award of large amounts of money to mitigate their noise impacts, rest on such dosage-response curves. On average, annoyance with aircraft sound is underestimated by general transportation noise curves by over 5 dB (Green and Fidell, 1991, p. 241). A thorough re-examination of the often-used FICON-curve indicates a systematic underestimation of annoyance with aircraft sound, particularly for the range of sound exposure levels that are of practical value (Fidell, 2003).

In the literature, several explanations for the systematic underestimation of noise annoyance (“excess annoyance”) have been given. With regard to the technical aspects of the Schultz curve “the functional form of the relationship, the range of values over which the relationship was developed, and its lack of source-specificity” have been blamed (Fidell, 2003, p. 3010; see also Miedema and Vos, 1998). Attempts to solve the problem have been made by fitting different, or source specific, functions on the data, but still much variance remains unexplained (Fidell *et al.*, 1991; Miedema and Vos, 1998).

In addition to these technical curve-related explanations, psychological explanations have been given. Noise annoyance is related to a range of variables besides the purely acoustical parameters. Analysis of survey data shows that much variance in annoyance scores is attributable to non-acoustic variables [e.g., noise sensitivity, and attitudinal variables like perceived mal- or misfeasance, distrust in authorities, uncertainty regarding the (future) noise environment (e.g., Sörenson, 1970; Staples, 1996; Guski, 1999; Fields, 1993; Job, 1988)]. Considered from a psychological perspective, noise annoyance is not a function of solely the characteristics of the acoustic stimulus, it is a function of a dynamic cognitive process in which the acoustic stimulus *and* a diversity of nonacoustic variables, including the attribution of semantic features, play a role [for an illustration of the cardinal role played by cognition in the evaluation of complex sounds, see Dubois *et al.* (2006); for an experiment illustrating semantic influences on sound evaluations, see Guastavino (2007)]. The World Health Organization (WHO) defines annoyance as “a feeling of discomfort which is related to adverse influencing of an individual or a group by any substances or circumstances” (WHO, 2004, p. 3), but the influence of circumstances is not incorporated in the dosage-response curves.

The causal direction of the relationship between non-acoustic variables and noise annoyance is not clear. It is plausible (Job, 1988; Cederlöf *et al.*, 1967) that variations in certain nonacoustic variables have a causal relationship with variations in noise annoyance. Although nonacoustic variables are likely related to stable person-related factors like personality or genetic make-up, it has been argued that they may to a certain extent be influenced by situational factors.

^{a)}Current affiliation: Faculty of Social and Behavioral Sciences, Universiteit Leiden, Section of Cognitive Psychology.

^{b)}Electronic mail: email@evelinmaris.nl

^{c)}Electronic mail: stallen@fsw.leidenuniv.nl

For instance, during situations of change, the heightened publicity and media attention may make residents more aware of the (effects of) noise than would be expected in a steady-state situation. Also, resentment with regard to perceived unfairness of the decision-making process (Fields *et al.*, 2000) has been speculated to increase the likelihood of residents' reporting annoyance (e.g., Green and Fidell, 1991). For the prediction of noise annoyance levels it is interesting to know whether a person's sensitivity or attitude toward sound can be influenced by situational variables.

If the majority of people in a community respond in largely the same way to the situational factors that affect noise annoyance, this will cause collective deflections from the dosage-response curve, rather than random variance in individual annoyance scores. There are indications that indeed deflections from the dosage response curve are to some extent collective. Research has shown that descriptive models of noise annoyance explain more variance in annoyance scores when they are enriched with one free parameter to account for collective nonacoustic differences: Up to an additional 47% of variance in annoyance scores is accounted for when a nonacoustic parameter, correcting for data-set-specific elevations or depressions in noise annoyance that cannot be explained by acoustical variables, is included in the mathematical model (Green and Fidell, 1991, p. 237; Fidell *et al.*, 1988; Miedema and Vos, 1999).

Notwithstanding the noteworthy improvement in the power of such enriched models, these models are purely descriptive and therefore neither improve the prediction of deflections from the curve, nor further our theoretical understanding of the psychology of annoyance in general, and the influence of nonacoustic factors in particular (Fidell, 2003). This is unfortunate, as both are needed to improve the abatement and prevention of noise problems (Staples, 1997, 1996). Given the financial and political costs of inaccurate predictions of annoyance and the apparent bother experienced by residents, it is important to further the theoretical understanding of noise annoyance in order to improve the accuracy of its prediction.

In this paper, it is argued that collective deflections from the curve can be understood and predicted when the social nature of noise annoyance is taken into account. Man-made environmental sound is rarely perceived in a social vacuum. People associate a sound they hear with its source (Guastavino, 2006), and in the case of man-made sound they may hold the operators of the source responsible for their exposure. Being exposed to man-made sound is a social experience: You expose Me (Stallen, 1999). The social process that characterizes a social experience influences the evaluation of that social experience as well as its associated outcomes (e.g., Lind and Tyler, 1988). When "You expose Me" to sound I do not like to hear, my judgment of the social process between us likely influences how far I will be annoyed by the fact that you are exposing me, and in how far I will be annoyed by the sound you make (Stallen, 1999; Van Gunsteren, 1999). Hence, the social process can be a situational factor that influences the cognitive process of sound evaluation. This is interesting, because the quality of the social process is, to a certain extent, tractable to the people who

are causing the noise. For example, if a person planning a late-night party at their home checks with their neighbors whether the timing of the party matches the neighbors' plans for the weekend, their annoyance with the party noise will likely be lower. Not only the acoustical properties of the sound can be a source of annoyance and discomfort, the social process instigated by the operators of the noise source can be a cause for dissatisfaction, too.

Many dimensions of the social process may be important in determining people's reactions to social experiences and related outcomes. One dimension has dominated social psychological research: The perceived fairness of the social process, in particular the fairness of the procedures used (Lind and Tyler, 1988, p. 1). In social psychology, a procedure is considered fair when people judge it to be fair.¹ Social psychological theories of justice describe an array of procedure characteristics that enhance fairness judgments, identified over years of experimental and survey research. Among other criteria, procedures are generally judged to be fairer when they (1) are transparent; (2) offer opportunities for participation in the decision-making process (e.g., "voice"); (3) are applied consistently across time and across persons; and (4) are applied in a respectful manner [Lind and Tyler, 1988; Mikula, 2001; Greenberg, 1993, for a concise review and meta-analysis of 25 years of social justice research, see Colquitt *et al.*, 2001]. Theories of social justice distinguish between one and four dimensions of justice (or fairness). Distributive justice (the fairness of outcomes relative to a certain standard, e.g., Adams, 1965; Deutsch, 1975; Leventhal, 1976) is distinguished from procedural justice (the fairness of the processes whereby outcomes are allocated, e.g., Thibaut and Walker, 1975; Folger, 1977; Tyler and Lind, 1992). Interpersonal justice "reflects the degree to which people are treated with politeness, dignity, and respect" (Colquitt *et al.*, 2001, p. 427), and informational justice "focuses on the explanations provided to people that convey information about why procedure were used in a certain way or why outcomes were distributed in a certain fashion" (Colquitt *et al.*, 2001, p. 427; Greenberg, 1993; Bies and Moag, 1986). Research has indicated that the four dimensions of fairness can have interactive effects (Colquitt *et al.*, 2001). In practice, however, these theoretical dimensions may overlap.

Fair procedures have been found to increase outcome satisfaction (the "fair process effect") (e.g., Lind and Tyler, 1988), and to decrease psychological stress (Vermunt and Steensma, 2001, 2003, 2005). The fair process effect is stronger when the outcomes are negative, or when physical stress is experienced (Tepper, 2001; Vermunt and Steensma, 2003). Research has also indicated that fair management procedures enhance feelings of trust in authorities, and increase people's support for policies (e.g., Mikula, 2001). Sound management activities by the operators of a sound source can be considered procedures (i.e., the operator allocates the sound to the residents). Hence, it can be expected that the perceived fairness of sound management procedures will influence people's evaluation of the sound. Results from several studies demonstrate the influence of procedures on reactions to noise, but the perceived fairness of these procedures has not

been assessed (Glass and Singer, 1972; Cederlöf *et al.*, 1967; Maziul and Vogt, 2002). A theoretical framework has not been proposed that explains or predicts effects of social processes on noise annoyance.

Social psychological theories of justice propose two major explanations why procedural fairness is generally much appreciated and related to higher outcome satisfaction. The *instrumental* explanation holds “that people are concerned about justice because it serves their self-interest of maximizing their outcomes in the long run” (Mikula, 2001, p. 8066, e.g., Thibaut and Walker, 1975). This perspective holds that fair procedures are appreciated because they give more (indirect) control over the process and the related outcomes.

Mediation analysis has shown that perceived control accounts for some but not all of the positive effects of procedural justice (Lind *et al.*, 1990). The *relational* or group-value explanation, on the other hand, holds that “people are concerned about their position in groups. They use experiences with their treatment by authorities as a source of information about their social position. The evidence that they are treated justly indicates that they are worthy members of the group” (Mikula, 2001, p. 8066, e.g., Tyler and Lind, 1992). Being treated with an unfair procedure is an indication that one has low status and is given little respect. Hence, unfair procedures and their outcomes are negatively evaluated and may give rise to psychological stress.

The effect of procedural fairness on evaluations of noise has been investigated (Maris *et al.*, 2007, 2004). In a laboratory experiment, participants are exposed to aircraft sound (50 vs 70 dB A) while they work at a reading task. The preceding sound management procedure is either fair (participants are given “voice” before they are exposed to the sound) or neutral (participants are simply told they will be exposed to the sound). “Within the 70 dB SPL condition, the fair procedure reduces the mean annoyance level to approximately the level in the 50 dB SPL condition. Up to 9% of variance in annoyance scores can be explained by the procedure manipulation” (Maris *et al.*, 2007). The procedure effect is not found in the 50 dB SPL conditions. The study demonstrates that under laboratory conditions and with rather loud sound, a fair sound management procedure ameliorates noise annoyance. It does not answer the question whether an *unfair* procedure can cause a collective increase in annoyance.

The present study investigates whether a collective increase of noise annoyance can be due to procedural unfairness. In other words: Can a procedurally unfair interaction with the operators of the sound source cause excess noise annoyance? The present experiment largely replicates the earlier experiment by Maris *et al.* (2007). Participants go through a noise management procedure before they are to perform a reading task while being exposed to annoying sound (played at 50 or 70 dB A). In the neutral procedure conditions participants are simply told that they will be exposed to aircraft sound. The unfair procedure is nontransparent as well as inconsistent: Participants are promised that they will be exposed to the sound type of their choice (nature, radio, or aircraft sound), which they, on the experimenter’s request, have clearly indicated on a form. The experimenter reads the form and gives them aircraft sound without

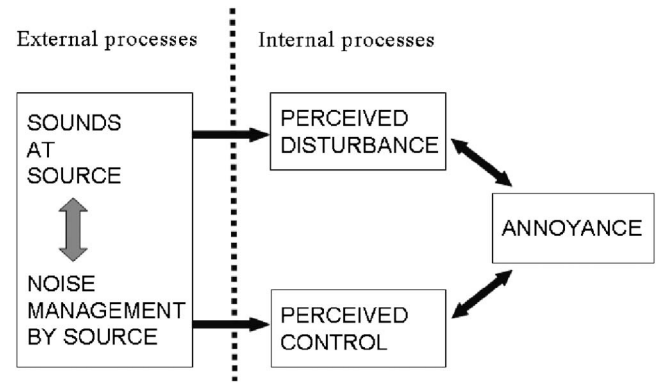


FIG. 1. Social psychological model of noise annoyance. The model, which is a simplification of the model by Stallen (1999), considers both the sound (“sounds at source,” e.g., sounds generated by the source) and its management (“noise management by source,” e.g., activities by the source ranging from keeping the sound volume within limits, to supplying residents with sound insulation or information, to asking the opinion of residents) as determinants of noise annoyance. The perception of these external processes results in perceptions of disturbance and/or control (internal processes). A perceived misbalance between disturbance and control results in annoyance. The model predicts that an increase in noise annoyance results from a deterioration of the acoustics, or from unsound management.

explaining why they do not follow the previously described procedure. No word of regret is given. (Giving participants a sincere apology or plausible explanation for the broken promise will give the unfair procedure a fair dimension (Bies and Shapiro, 1988). This is undesirable, as this fairness may exceed, or interact with, the unfairness of the broken promise. Alternatively, ambiguity is avoided because then the fairness of the interaction will lie entirely in the eye of the beholder rather than in the manipulation. Therefore, it is needed that the experimenter interacts in this slightly disrespectful manner.) Sound exposure levels and noise management procedures are varied systematically, and the dependent variables are assessed with a questionnaire. Data from participants who chose aircraft sound is excluded from the analyses.

The model used for the design of the study is a simplified version of the social psychological model of noise annoyance [Stallen, 1999; see Fig. 1; for a more detailed description of the experimental model see Maris *et al.* (2007)]. In the model, noise annoyance is represented as an expression of psychological stress, arising when the perceived level of disturbance due to the sound outgrows the perceived level of control over the sound (“internal processes”). People are expected to pay attention to the sound *and* to the sound management procedure; therefore both are included on the stimulus side of the model (“external processes”). The model further holds that the perceived disturbance is a function of the perceived sound, and that the perceived control is a function of the perceived sound management. Whether or not noise annoyance arises depends on the perception and evaluation of the sound in combination with the sound management.

It is expected that acoustical as well as procedural aspects of the noise exposure situation will influence the level of noise annoyance in the experiment. Obviously, it is hypothesized that in the high sound pressure level (SPL) conditions the annoyance levels will be higher than in the low

SPL conditions (Hypothesis 1). In addition, it is expected that systematic differences in the sound management procedure will yield systematic and collective differences in noise annoyance. Specifically, it is hypothesized that noise annoyance will be significantly higher in the unfair procedure conditions than in the neutral procedure conditions; that is: Excess annoyance will arise in the unfair procedure conditions (Hypothesis 2). The effects of procedural unfairness can be enhanced by negative outcomes or stress (Van den Bos *et al.*, 1998; Vermunt and Steensma, 2003). In the Maris *et al.* study (2007), a fair process effect was found only when the sound was loud (that is, when the outcome is negative). In the present experiment, it is expected that, to the participants in the unfair procedure conditions, the sound will be a negative outcome whether it is loud or not. Because participants have explicitly requested either nature or radio sound, receiving aircraft sound (of equal loudness) will be considered a negative outcome.

Therefore, an effect of procedural unfairness on noise annoyance is expected in the low SPL conditions *and* in the high SPL conditions (Hypothesis 3). No expectations are formulated regarding the relative strength of the procedure effects in the low and high SPL conditions. It is possible that the sound triggers a stronger procedure effect when it is both unwanted and loud (unfair, high SPL condition) than when it is unwanted but not loud (unfair, low SPL condition). It is also possible that the procedure effect is either triggered or not, and is always of the same strength.

Social psychological theories of justice offer instrumental and relational explanations for an effect of procedural unfairness on evaluations of sound (e.g., Lind *et al.*, 1990). It is explored whether instrumental or relational concerns mediate the procedure effect in the current study. It is hypothesized that if the effect of procedural unfairness on noise annoyance is mediated by instrumental concerns, the perceived control over the sound will mediate the procedure effect on noise annoyance (Hypothesis 4). Alternatively, if relational concerns mediate the procedure effect, the perceived regard of the experimenter to the participant will mediate the procedure effect (Hypothesis 5).

II. METHOD

A. Participants

One hundred and ten students, recruited from all departments of Universiteit Leiden, The Netherlands, are paid 5 Euro each to participate in the experiment, which lasts approximately 50 min. The participants are randomly assigned to each cell of the experimental design. Data from eight cases are excluded because these cases have indicated to have hearing problems. Data from another twelve cases are excluded from the analyses because these cases have chosen aircraft sound (see Sec. II D). Data from 90 students (74% female, mean age 21 years (s.d.=2.5)) are analyzed.

B. Experimental design

The experimental design is a 2 [(procedure: Neutral versus unfair) \times 2 (sound pressure level (SPL): Low (50 dB) versus high (70 dB)] complete factorial design.

C. Laboratory layout and stimulus material

The laboratory consists of four separate cubicles, each of which contains a desk and chair, and a complete PC set with two loudspeakers plus one subwoofer. In each session all participants are exposed to the same sound pressure level and procedure.

The three sound samples used in the introduction are (1) recordings of bird song taken from a CD (nature sound; Korrenromp, 2000), (2) recordings of a radio broadcasting show including both music and speech (radio sound), and (3) the sound of an aircraft passage (an excerpt from the experimental sample; aircraft sound). All are played at approximately 60 dB A (1 min Leq).

The experimental sample is composed of self-recorded audio material of aircraft passages of various loudness and duration.² The experimental sample is played at either 50 dB A (15 min Leq) (low SPL condition) or 70 dB A (15 min Leq) (high SPL condition), which implies a sound level of quiet background noise in the low condition or of speech interfering loudness in the high condition. The maximal sound pressure level is 68 or 88 dB A L_{max}, respectively. All sound pressure levels are measured in the cubicle at the position of the listener's ears.

The reading task (an English text with multiple choice questions, taken from a Dutch exam from pre-university education) is selected to match the cover story (see Sec. II D) and because it assures participants' motivation to perform well and closely matches their capacities. With too easy a task the experimental noise may not cause any disturbance and hence not induce annoyance. Too difficult a task may give rise to performance effects and task related frustration, which may cloud the effects of the procedure and/or the SPL manipulation (Smith, 1989).

D. Experimental procedure and manipulations

Upon their arrival at the laboratory, the experimenter leads the participants to their cubicle. After being seated, participants are left to themselves. Before starting to interact with the computer, participants read and sign an informed consent form, which informs them about their rights, among them the right to terminate their participation in the experiment at any moment. The computer is used for the presentation of the stimulus information and the recording of the dependent variables. (See Fig. 2 for a visual representation of the flow of the experiment.)

As a cover story, participants are told that they are engaged in a study on potential performance effects of disturbing sound during high school exams. As an introduction to the interaction with the computer, a series of three short sound samples dispersed by short questions is presented to the participants. The samples are of birds singing, a radio show, and of an aircraft passage. In fact, this series of samples gives all participants the same frame of reference with regard to the loudness and pleasantness of sounds in the experimental situation, as well as sets the stage for the procedure manipulation.

In the *neutral procedure* conditions, after participants are informed that in the main experiment they will be listen-

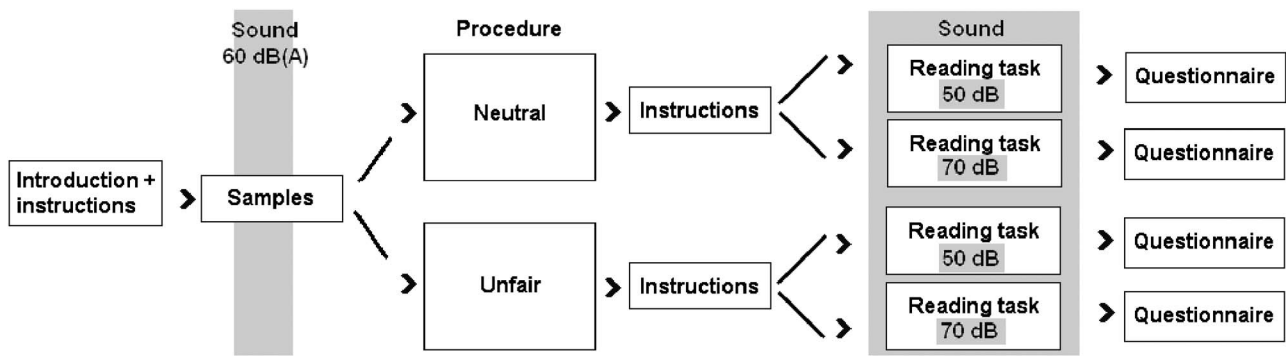


FIG. 2. Visual tree representation of the flow of the experiment.

ing to a 15-min sample of aircraft sound, they are asked to evaluate the three introductory samples before starting the main experiment. “Imagine yourself reading a difficult text hearing one of the three sound types (nature, radio, aircraft) as a background sound. Which sound type would seem the least taxing to you? Please write down your motivated answer.” They write down their, often elaborate, answer on a paper “answering form.”

The computer interface is designed in such a way that the experiment can only continue after the experimenter has entered a password to select either natural, radio, or aircraft sound. The instructions tell the participant to open the door of their cubicle in order to signal to the experimenter that they are ready to proceed. The experimenter enters the cubicle, collects the paperwork (informed consent and answering form), enters the password for aircraft sound, and says: “I have set the computer to aircraft sound.” The experimenter then leaves the participant to continue with the reading task.

In the *unfair procedure* conditions, participants are informed that in the main experiment they will be listening to a 15-min sample of their choice: nature, radio, or aircraft sound. Their choice is given significance by a remark that earlier research has indicated that participants who have been given the sound of their choice are far less tired after the experiment. Then, they are given the opportunity to indicate their preference: “Imagine yourself reading a difficult text hearing one of the three sound types (nature, radio, aircraft) as a background sound. Which sound type would seem the least taxing to you? Please write down your motivated choice.” They write down their, often elaborate, answer on a paper “choice form”, on which they also have to clearly visible circle the sample type of their choice.

The computer interface is designed in such a way that the experiment can only continue after the experimenter has entered a password to select either natural, radio, or aircraft sound. The instructions tell the participant to open the door of their cubicle in order to signal to the experimenter that they are ready to proceed. The experimenter enters the cubicle, collects the paperwork (informed consent and choice form), has a look at the choice form, notices the indicated preference, and says “Oh yeah, your preference, well...,” enters the password for aircraft sound, and says “I have set the computer to aircraft sound.” The experimenter does not

give any explanation for not following the procedure, and leaves the (often protesting) participant to continue with the reading task.

In *all conditions*, participants then start working on the reading task and related questions while being exposed to the aircraft sound sample (played at either 50 or 70 dB A). The reading task and the sound are automatically terminated after 15 min (none of the participants have by then finished the task). The computer then presents the questionnaire, which assesses the dependent variables and the manipulation checks. After the participants have completed the questionnaire, they check with the experimenter who thoroughly debriefs them and pays them.

E. Measures

1. Manipulation checks

One question checks the perceived loudness of the experimental sound (“perceived loudness”): “If you were to give a grade for the average loudness of the aircraft sound, what grade would you give?” Participants respond by clicking on one out of ten virtual buttons, shaded from white to black (see Fig. 3). The intensity of the grayscale indicates the intensity of the sound, and the first and the last button are labeled verbally (white=“the softest sound I have ever heard,” black=“the loudest sound I have ever heard”). The scores are coded into numbers ranging from 1 (white) to 10 (black).³ The mean score on this measure [$M(s.d.) = 6.92(1.77)$] is significantly higher than the scale’s midpoint 5.5 [$t(89) = 7.63, p < 0.001$], which indicates that on average participants considered the sound to be loud.

One question checks the perceived procedural fairness (“perceived procedural fairness”): “In my opinion, the procedure used by the researcher to select the sound I got to

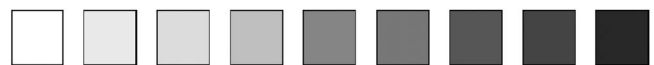


FIG. 3. The grayscale of the manipulation check for perceived loudness of the sound. Participants respond by clicking on one out of ten virtual buttons, shaded from white to black. The intensity of the grayscale indicates the intensity of the sound; the first and the last button are labeled verbally (white=“the softest sound I have ever heard,” black=“the loudest sound I have ever heard”).

listen to, is... . Verbal labels are given for the end points of the scale (1 = “very unfair” to 7 = “very fair.”). The mean score on the scale is $M(s.d.)=4.04(1.56)$.

Three explorative measures of task performance, to be used as a check for unintended performance effects, are automatically registered by the computer [“Time:” Time in seconds taken to read the first text and answer the first question, $M(s.d.)=88.26(40.04)$, 5% trimmed mean = 86.03; “Correct:” Total number of correct answers, $M(s.d.)=10.20(3.64)$; “False:” Total number of false answers, $M(s.d.)=4.19(2.86)$].

2. Dependent variables

The questionnaire includes three items that assess the participant’s noise annoyance (“noise annoyance”) with the experimental sound: (i) “To what extent did the sound annoy you while you were working at the task?,” (ii) “How did you experience the aircraft sound while answering the exam questions?,” (iii) “How pleasant did you feel the aircraft sound was while working on the exam?” Answers are given on a seven-point numerical rating scale, with verbal labels at the end points: (i) 1 = “not at all annoying,” 7 = “highly annoying,” (ii) 1 = “very positive,” 7 = “very negative,” (iii) 1 = “very pleasant,” 7 = “very unpleasant.” A noise annoyance scale is constructed by averaging the scores on the three items (Cronbach’s $\alpha=0.73$). The mean noise annoyance score is 5.65 ($s.d.=0.97$). The Pearson’s correlation between noise annoyance and the manipulation check perceived loudness is $r=0.38$ ($p<0.001$).

Instrumental concerns. One item assesses the participant’s perceived control over the experimental sound: “During the task, to what extent did you feel to have (had) control over the sound you were being exposed to?” Answers to this “perceived control” item are given on a seven-point numerical rating scale, with verbal labels at the end points: 1 = “not at all” to 7 = “to a great extent.” The average score on the perceived control item is $M=2.56$ ($s.d.=1.74$).

Relational concerns. Two items assess the perceived regard of the experimenter to the participant: (i) “The experimenter made an effort not to tax me unnecessarily with the sound,” and (ii) “In your opinion, how respectful have I, the researcher, treated you?” Answers are given on a seven-point numerical rating scale, with verbal labels at the end points: (i) 1 = “totally disagree” to 7 = “completely agree,” and (ii) 1 = “very disrespectful,” 7 = “very respectful.” No relational concern scale is constructed because the two items share relatively little variance ($r=0.35$; $p<0.001$). The two items, (i) effort [$M(s.d.)=3.87(1.77)$] and (ii) respect [$M(s.d.)=5.10(1.53)$], are analyzed separately.

Finally, some general questions (e.g. gender, self-reported hearing impairments (“Do you have any hearing impairment?,” response categories (i) “yes,” (ii) “somewhat,” (iii) “no”) are included. Participants who indicate to have (slight) hearing impairments (categories “yes:” $N=0$, and “somewhat:” $N=8$) are identified as having hearing impairments.

III. RESULTS

All analyses have been performed with and without the self-reported hearing impaired cases. The reported results are exclusive of the hearing impaired cases. Unless noted otherwise, the results are not notably different when the hearing impaired cases are included. By default, all reported significance levels are given for two-tailed tests with an alpha value of $\alpha=0.05$. A statistical test for homogeneity of error variances is reported for all analyses of variance, because the experimental design is slightly unbalanced (data from participants who chose aircraft sound is not included in the analyses). When error variances are nonhomogeneous, a more conservative statistic is reported in addition to the common F -statistic.

A. Manipulation checks

1. Perceived loudness

Analysis of variance (ANOVA) with perceived loudness as the dependent variable and sound pressure level (SPL) and procedure as the independent variables shows a main effect of SPL [$F(1, 86)=55.05$, $p<0.001$, $\eta^2=0.39$]. Levene’s test for equality of error variances [$F(3, 86)=5.03$, $p<0.005$] indicates that equal variances between groups cannot be assumed. The more conservative Welch’ variance-weighted ANOVA (one-way) indicates that significant differences between groups exist [$F(3, 41.94)=20.40$, $p<0.001$], and a post-hoc contrast test (equal variances not assumed) indicates that the mean loudness scores of the high SPL and low SPL conditions are significantly different [$t(51.96)=-6.92$, $p<0.001$]. The aircraft sound in the high sound conditions is perceived to be significantly louder [$M(s.d.)=8.00(1.17)$ than in the low sound conditions $M(s.d.)=5.80(1.58)$], indicating that the SPL manipulation was successful. No other significant effects are found [procedure: $F(1, 86)=0.26$, n.s.; SPL * procedure: $F(1, 86)=0.05$, n.s.], which indicates that the procedure manipulation did not influence perceived loudness.

2. Perceived procedural fairness

ANOVA with perceived procedural fairness as the dependent variable and SPL and procedure as the independent variables indicates that the unfair procedure [$M(s.d.)=3.56(1.64)$] is perceived to be significantly less fair than the neutral procedure ($M(s.d.)=4.71(1.18)$; $F(86)=13.41$, $p<0.001$, $\eta^2=0.14$); equal error variances can be assumed: Levene’s $F(3, 86)=1.76$, $p=0.16$, n.s.]. The deviation from the neutral score (4) is significant for the neutral condition [$t(37)=3.70$, $p<0.01$] and marginally significant for the unfair condition [$t(51)=-1.95$, $p=0.06$]. Strictly, the unfair procedure is not unfair in an absolute sense, but relative to the neutral condition it is significantly less fair. It is concluded that the procedure manipulation has been successful.

The main effect of SPL on perceived procedural fairness [$F(1.86)=4.06$, $p<0.05$, $\eta^2=0.05$] indicates that the perceived fairness of the procedure is influenced by the sound pressure level. In the high sound condition the procedure is evaluated as more fair [$M(s.d.)=4.35(1.70)$] than in the low

TABLE I. Noise annoyance scores (1 = “not annoyed at all,” 7 = “highly annoyed”) arranged by conditions of sound pressure level (low or high) and procedure (neutral or unfair). Cell means and marginal means (M), standard deviations (S.D.), and number of cases per cell (N) are given.

Sound pressure level	Procedure	M	s.d.	N
Low 50 dB	Neutral	5.07	1.12	18
	Unfair	5.55	0.91	26
	Total low	5.34	1.02	44
High 70 dB	Neutral	5.72	0.83	20
	Unfair	6.14	0.83	26
Total	Total high	5.96	0.85	46
	Neutral	5.41	1.01	38
	Unfair	5.85	0.91	52
Total		5.66	0.97	90

sound condition [M (s.d.)=3.73(1.35)]. No interaction of SPL and procedure on perceived procedural fairness is found [SPL * procedure: $F(1, 86)=0.87$, n.s.].

3. Performance measures

Multivariate analysis of variance with the performance measures time, correct, and false as the dependent variables and SPL and procedure as the independent variables is performed to check for unintended performance effects, and shows no significant multivariate effects [SPL: $F(3, 84)=0.20$, n.s.; procedure: $F(3, 84)=0.46$, n.s.; SPL * procedure: $F(3, 84)=0.46$, n.s.; equal error variances can be assumed: Box’s $M=14.24$, $F(18, 21118.37)=0.74$, $p=0.77$, n.s.]. The manipulations have not induced differences in performance.

B. Dependent variables

ANOVA with noise annoyance as the dependent variable and SPL and procedure as the independent variables is performed to test Hypotheses 1–3. Marginal means and cell means are summarized in Table I. The effects are described in the remainder of this section. Correlations between the dependent and independent variables are given in Table II.

It has been hypothesized that higher sound levels will result in higher noise annoyance levels (Hypothesis 1). The ANOVA shows a significant main effect of SPL on noise

TABLE II. Pearson’s correlations and point-biserial correlations between the dependent variables and the independent variables. Procedure and sound pressure level (SPL) are dichotomous variables (procedure: 1=neutral, 2=unfair; SPL: 1=50 dB, 2=70 dB).

$N=90$	SPL	Noise annoyance	Perceived control	Effort	Respect
Procedure	-0.03	0.22 ^a	-0.09	-0.32 ^b	-0.52 ^c
SPL		0.31 ^b	-0.06	-0.31 ^b	0.12
Annoyance			-0.10	-0.25 ^a	-0.16
P. control				0.23 ^a	0.09
Effort					0.35 ^c

^aCorrelation is significant at the 0.05 level (two tailed).

^bCorrelation is significant at the 0.01 level (two tailed).

^cCorrelation is significant at the 0.001 level (two tailed).

annoyance, indicating that participants who have been exposed to high sound are more annoyed than those receiving low sound [$F(1, 86)=9.94$, $p<0.01$, $\eta^2=0.10$; equal error variances can be assumed: Levene’s $F(3, 86)=1.93$, $p=0.13$, n.s.; see Table I for marginal means]. This finding confirms Hypothesis 1: The high sound level induces higher noise annoyance levels than the low sound level.

In the ANOVA, the main effect of procedure on noise annoyance [$F(1, 86)=5.32$, $p<0.05$, $\eta^2=0.06$] shows that participants who have been confronted with a broken promise with regard to their noise exposure (unfair procedure), report more noise annoyance [M (s.d.)=5.85(0.91)] than those who have simply been told they will be hearing aircraft sound (neutral procedure) [M (s.d.)=5.41(1.01)]. This finding confirms Hypothesis 2: The unfair procedure yields higher noise annoyance levels than the neutral procedure.

No interaction effect of procedure by SPL on noise annoyance is found [$F(1, 86)=0.02$, $p=0.89$, n.s.]. The effect of procedure on noise annoyance is found for the participants who have been exposed to 50 dB as well as for those who have been exposed to 70 dB, which confirms Hypothesis 3. The absence of an interaction effect also indicates that the procedure effect has the same strength in both SPL conditions. The effects of procedure and SPL on noise annoyance are additive and independent. The effect sizes of the two effects indicate that the effect of SPL is somewhat stronger than that of procedure. The proportion of variance in annoyance scores uniquely explained by SPL is 10%, for procedure this proportion is 6%.

The effect of the SPL and procedure manipulations on the variables perceived control, effort, and respect, the proposed mediators of the procedure effect, is investigated with three separate ANOVAs (the results are reported in Table III). For the actual mediation analyses, the indirect effects are estimated with a bootstrapping method (Shrout and Bolger, 2002) using the SPSS-macro provided by Preacher and Hayes (2004). This approach to mediation analysis is stated to be more accurate than traditional mediation analysis approaches (e.g., MacKinnon *et al.*, 2007). The SPSS-macro provides an estimate of the true indirect effect and its bias-corrected 95% confidence interval. In addition, the SPSS-macro generates the necessary output to assess the mediation using the traditional Baron and Kenny (1986) criteria, as well as a Sobel test of the observed indirect effect (Sobel, 1982). (The output of the mediation analyses is given in Table IV.)

ANOVA with perceived control as the dependent variable and SPL and procedure as the independent variables indicates that the manipulations have not induced group differences in perceived control [SPL: $F(1, 86)=0.23$, $p=0.64$, n.s.; procedure: $F(1, 86)=0.70$, $p=0.41$, n.s.; SPL * procedure: $F(1, 86)=0.30$, $p=0.58$, n.s.; equal error variances can be assumed: Levene’s $F(3, 86)=1.33$, $p=0.27$, n.s.; for cell means, see Table III]. The Baron and Kenny (1986) criteria indicate no indirect effect of procedure on noise annoyance through perceived control. The total effect of the procedure on noise annoyance [indicated as $b(YX)$ in Table IV] is statistically significant ($p<0.05$). However, neither the effect of the procedure on perceived control [indicated as $b(MX)$ in

TABLE III. Perceived control, effort, and respect scores (higher scores indicate higher perceived control, effort, respect) arranged by conditions of sound pressure level (low or high) and procedure (neutral or unfair). Cell means and marginal means (M), standard deviations (s.d.), and number of cases per cell (N) are given.

Potential mediator variable	Sound pressure level	Procedure	M	s.d.	N
Perceived control	Low 50 dB	Neutral	2.72	1.90	18
		Unfair	2.62	1.65	26
		Total low	2.66	1.74	44
	High 70 dB	Neutral	2.75	2.02	20
		Unfair	2.23	1.51	26
		Total high	2.46	1.75	46
	Total	Neutral	2.74	1.94	38
		Unfair	2.42	1.58	52
		Total	2.56	1.74	90
Effort	Low 50 dB	Neutral	5.17	1.15	18
		Unfair	3.92	1.74	26
		Total low	4.43	1.63	44
	High 70 dB	Neutral	3.95	1.50	20
		Unfair	2.85	1.78	26
		Total high	3.33	1.74	46
	Total	Neutral	4.53	1.47	38
		Unfair	3.39	1.83	52
		Total	3.87	1.77	90
Respect	Low 50 dB	Neutral	5.72	1.02	18
		Unfair	4.35	1.41	26
		Total low	4.91	1.43	44
	High 70 dB	Neutral	6.30	0.80	20
		Unfair	4.50	1.66	26
		Total high	5.28	1.62	46
	Total	Neutral	6.03	0.94	38
		Unfair	4.42	1.53	52
		Total	5.10	1.53	90

Table IV, $p=0.40$, n.s.] nor the effect of perceived control on noise annoyance, controlling for the effect of procedure [indicated as $b(YM.X)$ in Table IV, $p=0.45$, n.s.] are significant. The Sobel test of the indirect effect is highly insignificant ($p=0.67$, n.s.). The bootstrapped estimation of the true indirect effect of procedure on noise annoyance through perceived control is 0.01 [95% confidence interval (CI): -0.04 to 0.10 ; n.s.]. Hypothesis 4 that perceived control mediates the effect of procedure on noise annoyance is rejected. There is no indication that instrumental concerns explain the effect of procedural unfairness on noise annoyance.

ANOVA with effort as the dependent variable and SPL and procedure as the independent variables indicates that group differences in effort are induced by SPL as well as by procedure [SPL: $F(1, 86)=11.23$, $p<0.002$, $\eta^2=0.12$; procedure: $F(1, 86)=11.77$, $p<0.002$, $\eta^2=0.12$; SPL * procedure: $F(1, 86)=0.04$, $p=0.84$, n.s.; equal error variances can be assumed: Levene's $F(3, 86)=1.30$, $p=0.28$, n.s.]. The perceived effort made by the experimenter not to tax the participants unnecessarily is lower in the high SPL conditions than in the low SPL conditions [$M(s.d.)_{high}=3.33(1.74)$ vs $M(s.d.)_{low}=4.43(1.63)$], and lower in the unfair procedure conditions than in the neutral procedure conditions

[$M(s.d.)_{unfair}=3.39(1.83)$ vs $M(s.d.)_{neutral}=4.53(1.47)$; for cell means, see Table III]. The Baron and Kenny (1986) criteria indicate an indirect effect of procedure on noise annoyance through effort. In addition to the significant effect $b(YX)$, both the effect of the procedure on effort [$b(MX)$, $p<0.01$], and the effect of effort on noise annoyance, controlling for the effect of procedure [$b(YM.X)$, $p=0.07$, one-tailed test: $p<0.05$], are significant. Finally, the direct effect of procedure on noise annoyance, controlling for effort [indicated as $b(YX.M)$ in Table IV; $p=0.15$] is reduced relative to the direct effect $b(YX)$. However, the Sobel test of the indirect effect of procedure on noise annoyance through effort is not significant ($p=0.12$). The bootstrapped estimation of the true indirect effect is 0.13, but it is only marginally significant from zero (95% CI: -0.01 to 0.32 , $p>0.05$; 90% CI: 0.004 to 0.29 , $p<0.10$). It is concluded that the indirect effect of procedure on noise annoyance through effort is not significant at $p<0.05$.

ANOVA with respect as the dependent variable and SPL and procedure as the independent variables indicates that group differences in respect are induced only by procedure [SPL: $F(1, 86)=1.70$, $p=0.20$, n.s.; procedure: $F(1, 86)=32.08$, $p<0.001$, $\eta^2=0.27$; SPL * procedure: $F(1, 86)=0.57$, $p=0.45$, n.s.; due to unequal error variances with the higher error variance in the groups with a larger N , the F statistic is more conservative [Levene's $F(3, 86)=4.01$, $p<0.02$]. The perceived respectfulness of the treatment by the experimenter is lower in the unfair procedure conditions than in the neutral procedure conditions [$M(s.d.)_{unfair}=4.42(1.53)$ vs $M(s.d.)_{neutral}=6.03(0.94)$; for cell means, see Table III]. In none of the conditions, the treatment by the experimenter is perceived to be disrespectful. The Baron and Kenny (1986) criteria indicate no indirect effect of procedure on noise annoyance through respect. The effect $b(YX)$, and the effect of the procedure on respect [$b(MX)$, $p<0.01$, see Table IV] are significant, but the effect of respect on noise annoyance, controlling for the effect of procedure [$b(YM.X)$, $p=0.63$, n.s.] is not significant. The Sobel test of the indirect effect of procedure on noise annoyance through respect is insignificant, too ($p=0.64$, n.s.). The bootstrapped estimation of the true indirect effect is 0.06, but it is highly insignificant (95% CI: -0.20 to 0.32 ; n.s.).

Hypothesis 5, which holds that relational concerns mediate the effect of procedure on noise annoyance, has not been tested with the intended scale for relational concerns. Instead, two separate analyses have been used to test the indirect effect of procedure on noise annoyance through the two individual items (effort and respect). Both analyses have failed to find a significant indirect effect. Considering the results of these two mediation analyses, hypothesis 5 is rejected.

IV. CONCLUSIONS

The present paper, a social psychological approach to noise annoyance, rooted in theory, is proposed and experimentally tested. The core idea is that any sound is indissolubly associated with its source and that therefore being exposed to man-made sound is a social experience. A person's

TABLE IV. Output of the SPSS procedure (Preacher and Hayes, 2004) for estimating the indirect effect of procedure on noise annoyance through the respective proposed mediators perceived control, effort, and respect. The results are organized by proposed mediator variable. For the observed direct and total effects among the independent (X), the dependent (Y), and the mediator variable (M), the B coefficients (B coeff) and standard error (s.e.), t -statistic, and two-tailed p values are given. For the indirect effect of the independent on the dependent through the mediator, a Sobel significance test for the observed indirect effect, and a bootstrap estimation and confidence intervals of the true mean are given. The sample consists of 90 records. For each bootstrap estimation 3000 resamples are taken.

Proposed mediator variable	Statistics for direct, total, and indirect effects between procedure (X), noise annoyance (Y), and proposed mediator variable (M)					
Perceived control	Direct and total effects					
	Effect	B coeff.	s.e.	t	p	
	$b(YX)$	0.43	0.20	2.13	0.04	
	$b(MX)$	-0.31	0.37	-0.85	0.40	
	$b(YM.X)$	-0.04	0.06	-0.76	0.45	
	$b(YX.M)$	0.42	0.21	2.05	0.04	
	Indirect effect and significance using normal distribution					
		Value	s.e.	z	p	
	Sobel	0.01	0.03	0.42	0.67	
	Bootstrap results for indirect effect					
		Mean	s.e.	LL95%CI	UL95%CI	
	Effect	0.01	0.03	-0.04	0.10	
	Effort	Direct and total effects				
		Effect	B coeff.	s.e.	t	p
$b(YX)$		0.43	0.20	2.13	0.04	
$b(MX)$		-1.14	0.36	-3.18	0.00	
$b(YM.X)$		-0.11	0.06	-1.85	0.07	
$b(YX.M)$		0.31	0.21	1.45	0.15	
Indirect effect and significance using normal distribution						
		Value	s.e.	z	p	
Sobel		0.13	0.08	1.54	0.12	
Bootstrap results for indirect effect						
		Mean	s.e.	LL95%CI	UL95%CI	
Effect		0.13	0.09	-0.01	0.32	
Respect		Direct and total effects				
		Effect	B coeff.	s.e.	t	p
	$b(YX)$	0.43	0.20	2.13	0.04	
	$b(MX)$	-1.60	0.28	-5.72	0.00	
	$b(YM.X)$	-0.04	0.08	-0.48	0.63	
	$b(YX.M)$	0.37	0.24	1.56	0.12	
	Indirect effect and significance using normal distribution					
		Value	s.e.	z	p	
	Sobel	0.06	0.13	0.47	0.64	
	Bootstrap results for indirect effect					
		Mean	s.e.	LL95%CI	UL95%CI	
	Effect	0.06	0.13	-0.20	0.32	

evaluation of the sound is affected by the social process between themselves(s) and the operator(s) of the source. The results from the laboratory experiment confirm that the unfairness of the sound management procedure influences the evaluation of the sound. Relative to a neutral sound management procedure, an unfair procedure is found to yield collective excess annoyance. Defining exposure to man-made sound as a social experience, furthers the theoretical understanding of noise annoyance, which may well inspire new approaches to the abatement and prevention of noise annoyance.

The current results indicate that both the sound pressure level and the unfairness of the sound management procedure affect noise annoyance. The two effects are independent and

additive. In the earlier studies on procedural fairness and noise annoyance (Maris *et al.*, 2007; 2004) the procedure effect interacted with the sound pressure level. For the current experiment, no expectation has been formulated regarding an interaction effect of SPL and procedure. It is found that the procedure effect has the same strength in both SPL conditions. The sound has not triggered a stronger procedure effect when it is both unwanted and loud (unfair, high SPL condition) than when it is unwanted but not loud (unfair, low SPL condition). What explains the absence of an interaction effect in the present study? One explanation is that outcome negativity is a dichotomous rather than a continuous variable: An outcome is then perceived to be either negative or not, and hence no intensities of negativity are discerned. An-

other explanation is that the increase (or decrease) of noise annoyance due to a procedure effect is limited to about half a scale point. It is most likely, however, that due to a ceiling effect in the unfair-high SPL conditions no increased procedure effect has been found. The average annoyance score in the unfair-high SPL condition [$M(S.D.)=6.14(0.83)$] indicates that most participants already scored either 6 or 7 on the seven-point scale.

Some issues of validity need to be addressed. With regard to the fairness manipulation, it appears to be difficult to create a truly unfair procedure in a lab situation. Although the unfair procedure is significantly less fair than the neutral procedure, in an absolute sense the unfair procedure is perceived to be only marginally unfair. Possibly, in an experimental setting, the participants expect not to receive voice, causing its absence not to be felt as a salient violation of an (implicit) fairness norm. Given the significant effect of the procedure on noise annoyance it can be argued that it is the level of (un)fairness relative to a collective norm, and not so much the exact point on the fair—unfair continuum, that matters. Further research is needed to study the effect of strongly unfair procedures, as well as the effects of other procedural fairness criteria on noise annoyance.

The manipulation check for the perceived fairness of the procedure indicates that the sound pressure level of the aircraft sound has influenced the perceived fairness of the sound management procedure. When participants have listened to the 70 dB sample, they perceive the procedure to be fairer than the participants who have listened to the 50 dB sample. Since higher procedural fairness is associated with lower noise annoyance levels, this effect can have reduced the strength of the annoying effect of the SPL manipulation. The effect of SPL on the perceived fairness of the procedure is no alternative explanation for the effect of the procedure manipulation on noise annoyance.

Third, in the laboratory the interpersonal distance between the exposed and the operators of the source is relatively small. One may wonder whether effects of social processes on evaluations of noise can be replicated when the interpersonal distance between the exposed and the operators of the source is big (as is often the case in field settings), or when a real persona to interact with is lacking altogether (e.g., when the noise source is an institution). There is, however, evidence that people have a strong tendency to attribute social meaning to situations. For instance, research has shown that most people spontaneously and effortlessly ascribe motivations, intentions, and interactive behaviors to geometrical shapes moving about in a silent cartoon animation (e.g., the shapes are said to chase each other, or play, and to get frightened or elated) (e.g., Heider and Simmel, 1944; Klin, 2000). Other studies have shown that it is common for users of mass media to form so-called parasocial relationships with media figures (like celebrities, but also cartoon characters, or even magazines), in which the user responds behaviorally and cognitively to the media figure as though in a typical social relationship (e.g., Giles, 2002; Horton and Wohl, 1956; Cohen, 2004). Open interviews with people annoyed by the sound of wind turbines in Sweden illustrate that people perceive some kind of social relationship with the

owner of the wind turbine, and perceive its annoying sound as a violation of social norms (Pedersen, *et al.*, 2004). In sum, it is important to consider the difference between laboratory and field setting but it seems warranted to make careful generalizations from the current results to field settings.

With regard to the quality of the sound manipulation, some remarks need to be made. The recording and play back of the sound will not have created an optimal soundscape. Still, it is not likely that sound quality issues endanger the conclusions drawn from the data. The sound quality has been identical for all participants, ruling out the possibility that the effects found are due to artifacts of sound quality differences. In addition, research has indicated that the cognitive responses to “source events” (as opposed to “background sound” where the source is not easily identifiable) are rather robust to changes in sound reproduction method (Guastavino *et al.*, 2005). The current authors have no reason to expect that the effect of procedural fairness on noise annoyance found in the experiment will be an artifact due to the quality of the sound reproduction.

A surprising number of participants have indicated not to dislike aircraft sound. (In total 10.9% of the participants choose aircraft sound, against 13.6% radio sound and 75.5% nature sound.) A preliminary exploration of the arguments participants give for choosing aircraft sound indicates that they are used to hearing aircraft sound in their home situation, or they expect that its presumed monotony will not distract them as much as the spoken words audible in the radio sample or the twittering bird song in the nature sample.

Based on the results of the explorative mediation analyses, it cannot be concluded whether instrumental or relational concerns mediate the effect of procedural fairness on noise annoyance. Some remarks can be made. The proposed mediation by instrumental concerns is not confirmed by the data. The perceived control scores indicate that participants in general experienced very little control over the sound, and that neither of the manipulations has induced differences in perceived control. Given the fact that many studies have demonstrated an influence of perceived control on evaluations of noise (e.g., Glass and Singer, 1972), it seems advisable not to discard perceived control as a mediator of procedure effects too aptly. To investigate whether procedure effects on noise annoyance can be mediated by instrumental concerns, future studies need to apply manipulations of procedural fairness designed to induce differences in perceived control, and assess perceived control in a more advanced way.

The explorative analysis of mediation of the procedure effect on noise annoyance by relational concerns has yielded mixed results. First, the two items intended to assess the regard of the experimenter (effort and respect) appear not to tap one and the same concept. The procedure manipulation has induced strong differences in perceived respect, but those differences do not translate into differences in noise annoyance. The procedure-induced differences in effort, on the other hand, are considered a mediator of the procedure effect on noise annoyance according to the Baron and Kenny (1986) criteria. Even though the indirect effect of procedure on noise annoyance through effort is only marginally signifi-

cant, it seems premature to discard relational concerns as a mediator of procedure effects on noise annoyance. Future research is needed to construct a reliable and valid relational concern scale.

Noise annoyance is not solely a function of the characteristics of the acoustic stimulus. In the experiment, the fairness of the sound management procedure is an important determinant of noise annoyance besides the sound pressure level. Therefore, instead of marginalizing deflections from the dosage-response curve as “response bias,” such deflections need to be recognized as a key to a better understanding of the psychology of noise annoyance that can be a guide toward innovative abatement strategies. As a practical consequence, inventories and action plans aiming at the prevention or abatement of noise annoyance need to search for and address the social processes that influence the sound evaluation [for practical suggestions, see [Fields et al. \(2000\)](#)]. It is important to know whom people perceive to be the operators of the source, how the sound management procedures are evaluated, and according to which criteria. Social psychological theories of justice offer a range of procedural fairness criteria that may be of practical use. There will be cultural differences with regard to which type of procedure people regard as just, but the wish to be treated justly seems to be universal ([Montada, 2001](#)). Unsound management is best avoided.

ACKNOWLEDGMENTS

This research was financed by the Platform Nederlandse Luchtvaart (PNL, Platform Dutch Aviation), The Netherlands. The authors would like to thank Tim Weeber, who took part in the preparation and the data collection of the experiment as part of his Master’s thesis.

¹The audio material is recorded outdoors, with clear weather conditions, on one location in the vicinity of a runway in use for landings only. A professional company has removed ambient sounds from the recordings by a professional company. The 15-min experimental sample is made up of 11 noise events of aircraft passages of various loudness, duration, and aircraft type. The quiet time dispersing two passages (1 min, on average) is shorter than in real life and of variable duration.

²In social psychological theories of justice, the concepts fairness and justice are used interchangeably. The semantics of the fairness (or justice) concept are context dependent. Philosophers have studied the concept for centuries, if not millennia, and in a whole range of scientific disciplines (e.g., law, political sciences, anthropology, sociology) research on fairness (or justice) is conducted. “In contrast to other disciplines, social psychology does not take a normative approach [to justice]. It deals with justice in a descriptive rather than a prescriptive way. The aim is not to define what is just and unjust, and how justice can be achieved. The focus on the contrary is on the subjective sense of justice and injustice and its impact on human action and judgment. Social psychologists study what people regard as just and unjust under given circumstances, how people deal with the concept of justice, how they react to situations that they regard as unjust, and under which circumstances, and why, people care about justice” ([Mikula, 2001](#), pp. 8063–8064). Although it is very likely that substantial cultural differences exist with regard to which procedures people regard as just, the wish to be treated in a just way appears to be an anthropological universal ([Montada, 2001](#)).

³For the perceived loudness measure, a ten-point scale is used (deviant from the annoyance measure, which uses a seven-point scale) to prevent participants from ticking the exact same number on the annoyance measure and the perceived loudness measure, aiming to give a consistent (socially desirable) rather than a faithful answer. In earlier experiments the exact same manipulation of sound pressure levels was applied ([Maris et](#)

[al., 2007, 2004](#)). In these experiments, a numerical scale (ranging from 1 to 10) has been used. This numerical scale has indicated that the SPL manipulation induces very stable and strong differences in perceived loudness. In the current experiment, instead of a numerical scale, a discrete grayscale is used. It is expected that compared to numbers, visual intensity may be a more natural representation of auditory intensity.

- Adams, J. S. (1965). “Inequity in social exchange,” in *Advances in Experimental Social Psychology*, edited by L. Berkowitz (Academic, New York), Vol. 2, pp. 267–299.
- Baron, R. M., and Kenny, D. A. (1986). “The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations,” *J. Pers. Soc. Psychol.* 51, 1173–1182.
- Bies, R. J., and Moag, J. S. (1986). “Interactional justice: Communications criteria of fairness,” in *Research on Negotiation in Organizations*, edited by R. Lewicki, M. Bazerman, and B. Sheppard (JAI Press, Greenwich, CT), Vol. 1, pp. 43–55.
- Bies, R. J., and Shapiro, D. L. (1988). “Voice and justification: Their influence on procedural fairness judgments,” *Acad. Manage J.* 31, 676–685.
- Cederlöf, R., Jonsson, E., and Sörenson, S. (1967). “On the influence of attitudes to the source on annoyance reactions to noise. A field experiment,” *Scand. J. Work Environ. Health* 48, 46–59.
- Cohen, J. (2004). “Parasocial break-up from favorite characters: The role of attachment styles and relationship intensity,” *J. Soc. Pers. Relat.* 21, 187–202.
- Colquitt, J. A., Conlon, D. E., Wesson, W. J., Porter, C. O. L. H., and Ng, K. Y. (2001). “Justice at the millennium: A meta-analytic review of 25 years of organizational justice research,” *J. Appl. Psychol.* 86, 425–445.
- Deutsch, M. (1975). “Equity, equality, and need: What determines which value will be used for distributive justice?,” *J. Soc. Issues* 31, 137–150.
- Dubois, D., Guastavino, C., and Raimbault, M. (2006). “A cognitive approach to urban soundscapes: Using verbal data to access everyday life auditory categories,” *Acta. Acust. Acust.* 92, 865–874.
- Fidell, S. (2003). “The Schultz curve 25 years later: A research perspective,” *J. Acoust. Soc. Am.* 114, 3007–3015.
- 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., Schultz, T., and Green, D. M. (1988). “A theoretical interpretation of the prevalence of noise-induced annoyance in residential populations,” *J. Acoust. Soc. Am.* 84, 2109–2113.
- Fields, J. M. (1993). “Effect of personal and situational variables on noise annoyance in residential areas,” *J. Acoust. Soc. Am.* 93, 2753–2763.
- 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 CR-2000-210280, NASA Langley Research Center, Hampton, VA. Accessed from the internet, 24 February 2006, http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20010011059_2001008078.pdf (last viewed 10/18/07).
- Folger, R. (1977). “Distributive and procedural justice: Combined impact of ‘voice’ and improvement on experienced inequity,” *J. Pers. Soc. Psychol.* 35, 108–119.
- Giles, D. C. (2002). “Parasocial interaction: A review of the literature and a model for future research,” *Media Psychol.* 4, 279–305.
- Glass, D. C., and Singer, J. E. (1972). *Urban Stress. Experiments on Noise and Social Stressors* (Academic, New York).
- Green, D. M., and Fidell, S. (1991). “Variability in the criterion for reporting annoyance in community noise surveys,” *J. Acoust. Soc. Am.* 89, 234–243.
- Greenberg, J. (1993). “The social side of fairness. Interpersonal and informational classes of organizational justice,” in *Justice in the Workplace. Approaching Fairness in Human Resource Management*, edited by R. Cropanzano (Erlbaum, Hillsdale, NJ), pp. 79–103.
- Guastavino, C. (2006). “The ideal urban soundscape: Investigating the sound quality of French cities,” *Acta. Acust. Acust.* 92, 945–951.
- Guastavino, C. (2007). “Categorizations of environmental sounds,” *Can. J. Psychol.* 61, 54–63.
- Guastavino, C., Katz, B. F. G., Polack, J.-D., Levitin, D. J., and Dubois, D. (2005). “Ecological validity of soundscape reproduction,” *Acta. Acust. Acust.* 91, 333–341.
- Guski, R. (1999). “Personal and social variables as co-determinants of noise annoyance,” *Noise Health* 3, 45–56.
- Heider, F., and Simmel, M. (1944). “An experimental study of apparent behavior,” *Am. J. Psychol.* 57, 243–259.

- Horton, D., and Wohl, R. R. (1956). "Mass communication and para-social interaction," *Psychiatry* **19**, 215–229.
- Job, R. F. S. (1988). "Community response to noise: A review of factors influencing the relationship between noise exposure and reaction," *J. Acoust. Soc. Am.* **83**, 991–1001.
- Klin, A. (2000). "Attributing social meaning to ambiguous visual stimuli in higher-functioning Autism and Asperger syndrome: The social attribution task," *J. Child Psychol. Psychiatry* **41**, 831–846.
- Korenromp, J. (2000). "Als ik de marels zo hoor roepen. (When I hear the black-tailed godwits call)," [CD-Rom]. Privately published recordings of bird song. Available from Jos Korenromp, Schaepmanstraat 26, 7141 TP Groenlo, The Netherlands.
- Leventhal, G. S. (1976). "Fairness in social relationships," in *Contemporary Topics in Social Psychology*, edited by J. W. Thibaut, J. T. Spence, and R. C. Carson (General Learning, Morristown, NJ), pp. 211–240.
- Lind, E. A., Kanfer, R., and Earley, P. C. (1990). "Voice, control, and procedural justice: Instrumental and noninstrumental concerns in fairness judgments," *J. Pers Soc. Psychol.* **59**, 952–959.
- Lind, E. A., and Tyler, T. R. (1988). *The Social Psychology of Procedural Justice* (Plenum, New York).
- MacKinnon, D. P., Fairchild, A. J., and Fritz, M. S. (2007). "Mediation analysis," *Annu. Rev. Psychol.* **58**, 593–614.
- Maris, E., Stallen, P. J., Steensma, H., and Vermunt, R. (2004). "The influence of procedural fairness on evaluations of noise," IAPS, 2004, Vienna, Austria, 7–10 July; Internoise 2004, Prague, Czech Republic, 22–25 August.
- Maris, E., Stallen, P. J., Vermunt, R., and Steensma, H. (2007). "Noise within a social context: Annoyance reduction through fair procedures," *J. Acoust. Soc. Am.* **121**, 2000–2010.
- Maziul, M., and Vogt, J. (2002). "Can a telephone service reduce Annoyance?," 43rd Conference Deutsche Gesellschaft für Psychologie, Berlin, Germany, 22–26 September.
- Miedema, H. M. E., and Vos, H. (1998). "Exposure-response relationships for transportation noise," *J. Acoust. Soc. Am.* **104**, 3432–3445.
- Miedema, H. M. E., and Vos, H. (1999). "Demographic and attitudinal factors that modify annoyance from transportation noise," *J. Acoust. Soc. Am.* **105**, 3336–3344.
- Mikula, G. (2001). "Justice: Social psychological perspectives," in *International Encyclopedia of the Social and Behavioral Sciences*, edited by N. J. Smelser and P. B. Baltes (Elsevier, Amsterdam), pp. 8063–8067.
- Montada, L. (2001). "Justice and its many faces: Cultural concerns," in *International Encyclopedia of the Social and Behavioral Sciences*, edited by N. J. Smelser and P. B. Baltes (Elsevier, Amsterdam), pp. 8037–8042.
- Pedersen, E., Persson Waye, K., and Hallberg, L. R. M. (2004). "Living close to wind turbines—A qualitative approach to a deeper understanding," paper presented at the 33rd International Congress and Exposition on Noise Control Engineering, Inter-noise 2004, August 22–25, Prague, Czech Republic.
- Preacher, K. J., and Hayes, A. F. (2004). "SPSS and SAS procedures for estimating indirect effects in simple mediation models," *Behav. Res. Methods Instrum. Comput.* **36**, 717–731.
- Shrout, P. E., and Bolger, N. (2002). "Mediation in experimental and non-experimental studies: New procedures and recommendations," *Psychol. Methods* **7**, 422–445.
- Smith, A. (1989). "A review of the effects of noise on human performance," *Scand. J. Psychol.* **30**, 185–206.
- Sobel, M. E. (1982). "Asymptotic confidence intervals for indirect effects in structural equation models," in *Sociological Methodology 1982*, edited by S. Leinhardt (Jossey-Bass, San Francisco, CA), pp. 290–312.
- Sörensen, S. (1970). "On the possibilities of changing the annoyance reaction to noise by changing the attitudes to the source of annoyance," *Nordisk Hygienisk Tidskrift Supplementum 1* (Tryckeri Balder AB, Stockholm), pp. 1–76.
- Stallen, P. J. (1999). "A theoretical framework for environmental noise annoyance," *Noise Health* **3**, 69–79.
- Staples, S. L. (1996). "Human response to environmental noise. Psychological research and public policy," *Am. Psychol.* **51**, 143–150.
- Staples, S. L. (1997). "Public policy and environmental noise: Modeling exposure or understanding effects," *Am. J. Public Health* **87**, 2063–2067.
- Tepper, B. J. (2001). "Health consequences of organizational injustice: Tests of main and interactive effects," *Org. Behav. Hum. Decis. Process* **86**, 197–215.
- Thibaut, J., and Walker, L. (1975). *Procedural Justice: A Psychological Analysis* (Erlbaum, Hillsdale, NJ).
- Tyler, T. R., and Lind, E. A. (1992). "A relational model of authority in groups," in *Advances in Experimental Social Psychology*, edited by M. Zanna (Academic, San Diego, CA), Vol. **25**, pp. 115–191.
- Van den Bos, K., Wilke, H. A. M., Lind, E. A., and Vermunt, R. (1998). "Evaluating outcomes by means of the fair process effect: Evidence for different processes in fairness and satisfaction judgments," *J. Pers Soc. Psychol.* **74**, 1493–1503.
- Van Gunsteren, H. R. (1999). "When noise becomes too much noise...," *Noise Health* **3**, 3–5.
- Vermunt, R., and Steensma, H. (2001). "Stress and justice in organizations: An exploration into justice processes with the aim to find mechanisms to reduce stress," in *Justice in the Workplace. From Theory to Practice*, edited by R. Cropanzano (Erlbaum, London), Vol. **2**, pp. 27–48.
- Vermunt, R., and Steensma, H. (2003). "Physiological relaxation: Stress reduction through fair treatment," *Soc. Justice Res.* **16**, 135–150.
- Vermunt, R., and Steensma, H. (2005). "How can justice be used to manage stress in organizations?," in *Handbook of Organizational Justice*, edited by J. G. Greenberg and J. A. Colquitt (Erlbaum, Mahwah NJ), pp. 383–410.
- World Health Organization. (2004). "WHO LARES. Final report. Noise effects and morbidity," H. Niemann and C. Maschke, Berlin Center of Public Health. Retrieved 21 June 2006 from http://www.euro.who.int/Document/NOH/WHO_Lares (last viewed 10/17/07).

Coherent acoustic communication in a tidal estuary with busy shipping traffic

Paul A. van Walree^{a)}

*The Netherlands Organisation for Applied Scientific Research TNO, Underwater Technology Group,
Oude Waalsdorperweg 63, P.O. Box 96864, 2509 JG The Hague, The Netherlands*

Jeffrey A. Neasham

*School of Electrical, Electronic and Computer Engineering, University of Newcastle upon Tyne,
Merz Court, Newcastle upon Tyne NE1 7RU, United Kingdom*

Marco C. Schrijver

Zeeland Department, Rijkswaterstaat, P.O. Box 5014, 4330 KA Middelburg, The Netherlands

(Received 8 February 2007; revised 10 September 2007; accepted 10 September 2007)

High-rate acoustic communication experiments were conducted in a dynamic estuarine environment. Two current profilers deployed in a shipping lane were interfaced with acoustic modems, which modulated and transmitted the sensor readings every 200 s over a period of four days. QPSK modulation was employed at a raw data rate of 8 kbit/s on a 12-kHz carrier. Two 16-element hydrophone arrays, one horizontal and one vertical, were deployed near the shore. A multichannel decision-feedback equalizer was used to demodulate the modem signals received on both arrays. Long-term statistical analysis reveals the effects of the tidal cycle, subsea unit location, attenuation by the wake of passing vessels, and high levels of ship-generated noise on the fidelity of the communication links. The use of receiver arrays enables vast improvement in the overall reliability of data delivery compared with a single-receiver system, with performance depending strongly on array orientation. The vertical array offers the best performance overall, although the horizontal array proves more robust against shipping noise. Spatial coherence estimates, variation of array aperture, and inspection of array angular responses point to adaptive beamforming and coherent combining as the chief mechanisms of array gain. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2793706]

PACS number(s): 43.60.Dh, 43.60.Fg, 43.60.Gk [EJS]

Pages: 3495–3506

I. INTRODUCTION

A. The Scheldt Estuary

The Western Scheldt is part of the Scheldt Estuary, which meanders through the south-west of the Netherlands (Fig. 1). The present shape of the estuary was established around 1800 A.D. when poldering and the construction of dikes limited the freedom of movement of the tidal channels. Van Veen¹ defines the tidal system as mutually evasive ebb and flood channels. In this definition the flood channel is open to the flood current and exhibits a sill at the upstream end; the ebb channel is open to the ebb current and exhibits a sill at the seaward end. The curved ebb channel is deeper than the straight flood channel and is therefore used as the main navigation channel in the estuary.

The tidal wave, which runs up from north to south along the coastline of the Netherlands, is amplified as it travels up the estuary. The mean tidal range increases from 3.8 m at Flushing to 5.2 m at Antwerp (Belgium), which is located 78 km upstream. Large vessels use the tidal wave to pass the sills in the navigation channel. Owing to the difference between the tidal volume and the discharge of the River

Scheldt, about 10^9 and 1.5×10^6 m³ per tidal cycle, respectively, and the morphology, the estuary is well mixed, with a salinity that decreases toward Antwerp.

Apart from being an area with unique flora and fauna, the Western Scheldt also has an economic function as the main navigation channel to the harbors of Flushing, Terneuzen, Ghent and Antwerp. Approximately 25 000 seagoing vessels cross the Western Scheldt en route to or from Antwerp yearly. Moreover, an estimated 45 000 barges and 10 000 recreational ships traverse the estuary each year. Because of international competition, accessibility for large ships is of vital interest to the harbors. The maneuverability of big freighters in the winding navigation channel is limited and depends on, among other things, currents and water depth. The real-time availability of current and keel clearance data to mariners and vessel traffic management would improve the nautical safety on the Western Scheldt. A direct measurement of the keel clearance with the help of an upward looking acoustic Doppler current profiler (ADCP) was described in a previous paper.² The present paper describes the execution and laboratory analysis of sea trials named DACAPO (demonstration of acoustic communication for ADCP read-out), aimed at the wireless acquisition of current profiles in the navigation channel. These sea trials took place near the village of Hansweert in August 2005.

^{a)}Author to whom correspondence should be addressed. Electronic mail: paul.vanwalree@tno.nl

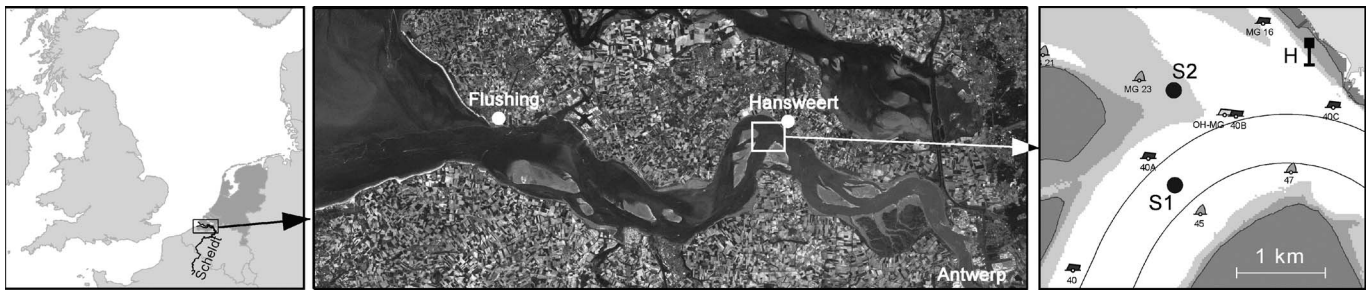


FIG. 1. Area map with the Netherlands, the Scheldt Estuary, and the sea trials location. Subsea modems are located at positions S1 and S2; H indicates the location of measurement pole Hansweert with the receiver barge. The two black curves delineate the main navigation channel.

B. Underwater communication in shallow water

Acoustic communication in shallow waters is an active field of research with significant challenges to overcome.³ Multiple boundary interactions under dynamic conditions cause high levels of multipath with a high degree of temporal and spatial variability. The underwater channel is usually doubly spread, with a considerable dispersion in the time delay and Doppler dimensions. Reports of time dispersion spanning tens or hundreds of milliseconds are not unusual. High-rate, phase-coherent communication systems thus require complex receiver structures. Often such acoustic telemetry channels are not limited by noise, but by reverberation beyond the capability of equalizer structures.³ Temporal variability further demands adaptive filter update algorithms in order to track the impulse response. Explicit carrier/symbol synchronization loops may be required to mitigate Doppler shifts in horizontal channels, even when the transmitter or receiver platform is moored and subject to mainly vertical movement on the waves.⁴

Long-term environmental influences due to, for example, tidal, diurnal, or seasonal cycles are not yet well understood. Investigation of these effects requires modeling or measurements of propagation parameters or acoustic communications directly, stretching over extended periods.⁵⁻⁷ The present study reveals a strong tidal influence on the performance of a shallow-water communication link, for which the tidal range compares to the water depth. Further, the busy shipping traffic in the Western Scheldt gives rise to numerous

instances of acoustic communication limited by ship-generated noise or extinction by ship wakes, rather than reverberation.

This paper is organized as follows. Section II describes the sea trials geometry and experimental setup. The interfacing between current profilers and the acoustic modems is outlined, the modulation is detailed, and a block diagram of the receiver architecture is presented. Section III tabulates the long-term averaged receiver performance, computes correlation coefficients between receiver output signal-to-noise ratio (SNR) and a set of acoustic and environmental parameters, and shows several example receptions of modem signals in the presence of ship noise and other disturbances. In Sec. IV the receiver performance is quantified for different array configurations, array angular responses are determined, and candidate mechanisms of array gain are discussed. Findings are summarized in Sec. V.

II. DACAPO SEA TRIALS

A. Experimental setup

The DACAPO test area is centered on the Hansweert measurement pole, an existing sensor station for environmental monitoring. Figure 2 depicts the experimental setup, which is further illustrated with a backdrop of shipping traffic by the photograph in Fig. 3. Two pyramidal frames were equipped with an ADCP, an acoustic modem, and battery containers for both instruments. The frames were deployed on the seafloor at distances of 1820 and 1440 m from the measurement pole, at water depths of 20 and 8 m, respectively. A barge was moored adjacent to the pole, where the water depth is particularly shallow at 6 m. Figure 4 depicts the bathymetry along the acoustic transects. S1 is located within the principal shipping lane at a position ensuring a

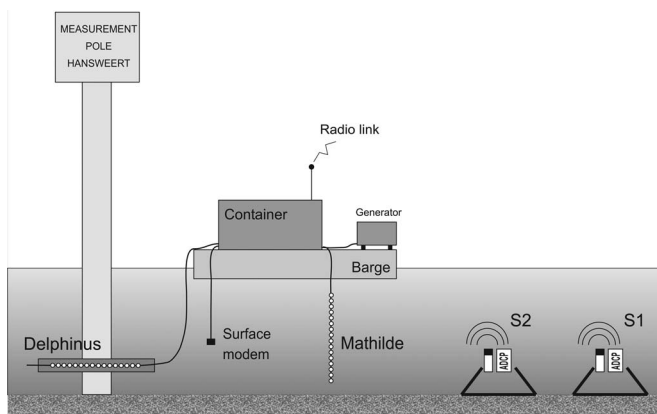


FIG. 2. Schematic setup of the experiment.



FIG. 3. Photograph taken from the Hansweert side of the Scheldt Estuary.

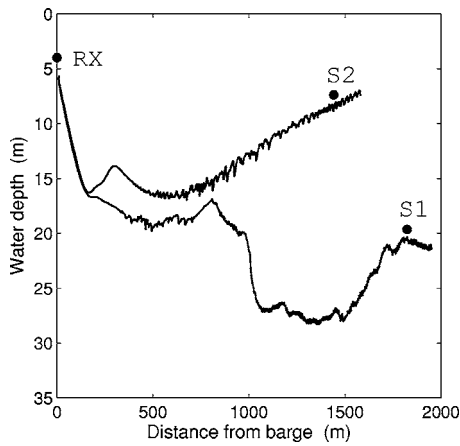


FIG. 4. Measured bathymetry along the communication tracks between the two subsea modems and the barge at RX. The water depth is given relative to NAP (normal Amsterdam level).

line of sight with the receive arrays. S2 is deployed at the boundary of a shallower, secondary navigation channel (see Fig. 1). Passing bulk carriers and container ships invariably follow the bend in the main lane, traversing the line of sight between S1 and H if they pass this subsea unit at the west side. Barges and pleasure boats can alternatively opt for the secondary channel, intersecting the line of sight between S2 and H.

The barge supported a container with equipment for the recording of raw acoustic data, for offline processing, and a radio link to convey the ADCP readings ashore. There was also an AIS (automatic identification system) receiver to register passing vessels equipped with an AIS transponder. A third acoustic modem (surface unit) was provided on the barge to establish a highly reliable command link with the subsea units, using spread-spectrum modulation techniques. Two hydrophone arrays were deployed to investigate the quality of reception as a function of array orientation. A horizontal array “Delphinus,” originally designed for the detection of marine mammals,⁸ was firmly fixed against the measurement pole. An intermediate plate of absorbing cork was inserted to reduce reflections from the pole. A similar array “Mathilde” was suspended vertically from the barge, with a weight to minimize tilting by the current. At a carrier frequency of 12 kHz, both arrays are approximately half-wavelength spaced with 16 elements evenly distributed over an aperture of order 1 m (see Table I).

After deployment of all equipment, the surface modem was briefly lowered from the barge. A trigger command was sent to the subsea units, which subsequently entered a mode of periodic ADCP interrogation. Every 200 s a current profile was measured, modulated, and broadcast by the omnidirectional modem. The two subsea units were triggered with an initial 100-s phase offset to avoid collisions of the acoustic communication packets. The source level of acoustic transmissions was no more than 182 dB re 1 μ Pa at 1 m, corresponding to approximately 36 W of electrical power drawn from the subsea battery supply. On the barge the received signals were digitized with a data acquisition system, storing all hydrophone channels of both arrays at a sampling frequency of 48 kHz. At sea the recordings were processed with

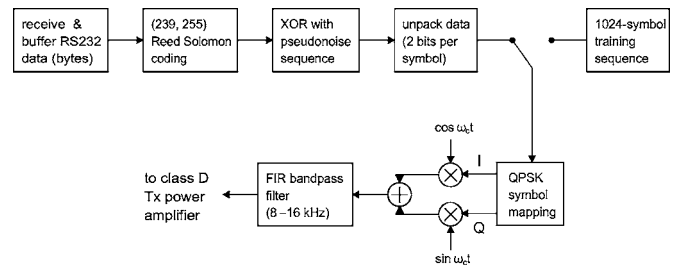


FIG. 5. Signal processing chain for the transmitter modulation.

a communications receiver, whereupon extracted current profiles were transferred to land via radio telemetry. After the sea trials the raw data were used for the laboratory analysis described in this paper.

The total duration of the DACAPO sea trials was three weeks, but the vertical array was available only during the first days. This paper analyzes the time span during which both arrays were operated. A total of 1608 data packets were received from each of the two modems during this four-day period.

B. Subsea modems, modulation and interfacing

The Tritech AM300 acoustic modem⁹ was adapted for the subsea modems in this experiment and interfaced to an RD Instruments ADCP via RS232 communication at 19 200 baud. The micropower standby processor of the AM300 was configured to wake up the modem at a regular interval of 200 s. On waking up, the modem issued commands to the ADCP to wake up, grab a set of current data (approximately 1 kbyte) and then shut down. From there on, all of the coding and modulation is performed by the AM300 floating point DSP engine, implementing the stages described in Fig. 5.

The transmitted signal frame starts with a 20-ms, 9–15 kHz linear frequency-modulated chirp which can be used for Doppler-tolerant frame synchronization or channel measurement. Following that a 1024-symbol pseudorandom training sequence is transmitted for the convergence of the adaptive filters in the receiver. ADCP data are divided up and coded into 5 blocks of a (239,255) Reed Solomon forward error correction code, resulting in a total of 10 200 coded

TABLE I. Array description and choice of receiver parameters: number of filter taps, LMS step-size parameter μ , and the proportional and integral phase tracking constants K_1 and K_2 .

	Horizontal array “Delphinus”		Vertical array “Mathilde”	
	S1	S2	S1	S2
Elements	16		16	
Spacing (cm)	6		7	
Aperture (cm)	90		105	
Feedforward taps	90	50	90	50
Feedback taps	110	70	110	70
μ	1/250	1/150	1/250	1/150
K_1	0	0	0.001	0.001
K_2	0	0	0.0001	0.0001

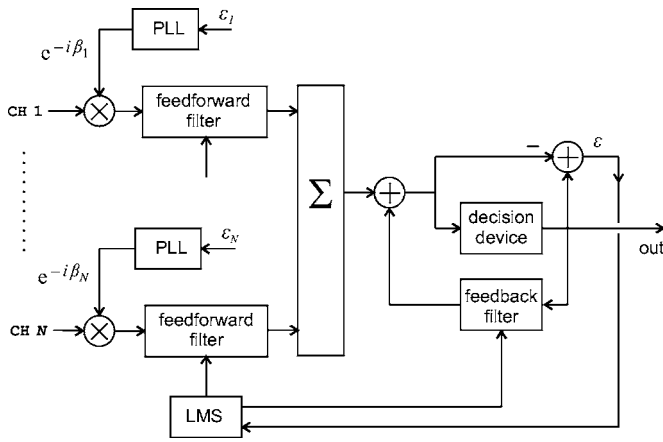


FIG. 6. Receiver block diagram with multiple feedforward filters and a common feedback section. All filter taps are jointly updated with the LMS algorithm to minimize the error ϵ .

bits. The bit stream is subsequently xor-ed with a known pseudonoise sequence to ensure good random properties, essential for the tracking of an adaptive receiver in decision-directed mode. The randomized bits are then divided into 2-bit symbols which are mapped onto a Gray coded quadrature phase-shift keyed (QPSK) constellation and modulated onto a 12-kHz carrier. Sidelobes are reduced by an 8–16 kHz bandpass filter. At a symbol rate of 4000 sps, the total duration of the transmitted data frame (packet) is 1.55 s at an effective data rate of 6.2 kbit/s.

C. Receiver description

Because of the stationary DACAPO geometry the Doppler tolerant chirp is skipped for detection. Instead, frame synchronization is achieved by matched filtering one channel for the pseudorandom training sequence at the start of the frame. This 256-ms sequence offers a higher gain than the 20-ms chirp and allows detection at low SNR. The strongest peak in the filter output is taken to represent the start of the frame. From here the basebanded hydrophone signals enter the adaptive receiver shown in Fig. 6, followed by a descrambler and Reed Solomon decoder (cf. Fig. 5). The receiver is a multichannel decision-feedback equalizer (DFE) with 16 $T/2$ -spaced feedforward filters and a single, T -spaced feedback filter. The theoretical background and merits of this structure have been described well by Stojanovic *et al.*¹⁰ The filter updates its taps at the symbol rate under the minimum mean-square error criterion. In the absence of any spatial coherence the adaptive filter acts as a maximal ratio combiner, exploiting the channel diversity to combat spatial or frequency-selective fading on individual elements. On the other hand, when there is good coherence the receiver can act as a spatial filter.¹¹ It can steer a beam in the signal direction, or steer nulls in the direction of interferers, to increase the output SNR. When multipath components arrive at different angles it can coherently combine or cancel them to reduce intersymbol interference. For arrays of the dimensions used in this experiment, we would expect the

receiver to perform a mixture of diversity combining and adaptive beamforming in its attempt to minimize the mean square error.

Table I shows the filter parameters used to demodulate the recorded signals. A large number of filter taps is required to combat the high level of multipath encountered in the estuary channels, which is worse for S1 than for S2 (see Fig. 7, top panels). 90 feedforward taps and 110 feedback taps are chosen for S1, compared with 50 feedforward taps and 70 feedback taps for S2. The large filter sizes dictate the choice of the least mean squares (LMS) update algorithm, rather than recursive least squares (RLS) or similar, for its low computational complexity. The choice of the phase tracking constants is not related to the different orientations of the arrays, but to the manners of deployment. Mathilde is suspended into the water column from the barge, which moves up and down on the waves. This motion introduces a time-varying Doppler shift with zero mean, which is effectively compensated by a digital phase locked loop (PLL). For fierce waves the vertical movements are accompanied by some array tilting, which introduces a depth dependence of the relative velocity between a hydrophone and the transmitting sub-sea unit. Therefore each hydrophone is equipped with a dedicated PLL. In order to update the PLLs, the error ϵ_i is computed by subtracting the equalized symbol for channel i from the overall hard decision after combining. Conversely, Delphinus is rigidly mounted against the measurement pole and does not require explicit phase tracking. It is remarked that the parameters in Table I were roughly selected so as to obtain the best overall performance for each modem/array combination. A single set of parameters could be used to process all received signals at suboptimal performance, but without detracting from the main observations and conclusions of this paper.

III. EFFECTS OF TIDE AND SHIPPING ON THE COMMUNICATION

A number of acoustic and environmental parameters are analyzed before the receiver performance is discussed. Acoustic parameters extracted from the raw data are the root mean square (rms) background noise level, peak noise level, signal level, channel time spread, and spatial coherence—all to be measured in the frequency band of the communication signal. Environmental parameters are the water level, water temperature, and wind speed. Some of these parameters are strongly correlated with the fidelity of the communication link and necessary ingredients in the discussion.

A. Acoustic and environmental parameters

When frame synchronization has been established, a 1-s background noise segment X is selected just before the start of the frame and another 1-s segment Y inside the frame, containing signal and noise. The peak noise level (dB) is given by

$$L_{\text{noise,peak}} = 20 \log_{10}(\max\{|X_1|, \dots, |X_{N_x}|\}) + C, \quad (1)$$

where N_x is the number of samples of X and C is a known calibration constant. The rms noise level is

$$L_{\text{noise,rms}} = 10 \log_{10} \left(\frac{1}{N_x} \sum_{k=1}^{N_x} |X_k|^2 \right) + C, \quad (2)$$

and the signal level

$$L_{\text{signal}} = 10 \log_{10} \left(\frac{1}{N_y} \sum_{k=1}^{N_y} |Y_k|^2 - \frac{1}{N_x} \sum_{k=1}^{N_x} |X_k|^2 \right) + C, \quad (3)$$

with N_y denoting the number of samples of Y . The signal-to-noise ratio at the input of the receiver follows as

$$\text{SNR}_{\text{in}} = 10 \log_{10} \left[\frac{1}{N_y} \sum_{k=1}^{N_y} |Y_k|^2 \left(\frac{1}{N_x} \sum_{k=1}^{N_x} |X_k|^2 \right)^{-1} - 1 \right]. \quad (4)$$

As the noise characteristics can vary between X and Y , the quantities defined in Eqs. (1)–(4) are estimates rather than precise measurements.

At the other end of the receiver we have the output SNR, defined as

$$\text{SNR}_{\text{out}} = 10 \log_{10} \left(\frac{\sum_n |S_n|^2}{\sum_n |\tilde{S}_n - S_n|^2} \right), \quad (5)$$

where \tilde{S}_n denotes the received symbols and S_n the transmitted symbols. However, as the transmitted symbols are not known for Reed Solomon blocks which fail to decode, the S_n are replaced by the constellation points closest to the received symbols. This leads to an overestimate of the output SNR at very low values, but the difference is negligible for the discussions in this paper. Unlike the input SNR, the output SNR is not a true signal-to-noise ratio but an overall measure of the fidelity of the receiver output that includes the effects of additive noise, intersymbol interference, and other channel distortions.

A measure of the time spread is obtained by considering the accumulated energy output of the filter matched to the pseudorandom training sequence. The delay time interval within which 90% of the acoustic energy arrives is taken as the time spread. Another quantity extracted from the acoustic data is the spatial coherence. The coherence $C_{\text{h1-h16}}$ between the signals Y received on the extreme hydrophones (labeled 1 and 16) is estimated via their power spectral densities $P_{\text{h1-h1}}$ and $P_{\text{h16-h16}}$, and the cross power spectral density $P_{\text{h1-h16}}$, according to

$$C_{\text{h1-h16}} = \frac{100\%}{B} \int_{-B/2}^{B/2} \frac{|P_{\text{h1-h16}}(f)|^2}{P_{\text{h1-h1}}(f)P_{\text{h16-h16}}(f)} df, \quad (6)$$

where $B=8$ kHz specifies the bandwidth of the basebanded communication signal. The coherence thus computed is between 0 and 100% and an indicator of the beamforming potential of the array.

Available environmental parameters include the water level and wind speed, which are routinely logged by instruments on the measurement pole. Although these quantities are measured at the receiver end of the communication links, they are representative of a larger area. The last environmental parameter is the water temperature, which is measured by the ADCPs and included in the transmitted data stream. This

TABLE II. Percentage of correctly received data packets per array and per subsea unit. The mean value of the receiver output SNR and its standard deviation σ are also given. Note that the latter figures are computed after decibel conversion.

	S1	S2
Number of broadcasts	1608	1608
Horizontal array Delphinus		
Checksum correct	1065 (66.2%)	1602 (99.6%)
$\langle \text{SNR}_{\text{out}} \rangle$ (dB)	10.3 ($\sigma=2.7$)	18.3 ($\sigma=2.6$)
Vertical array Mathilde		
Checksum correct	1473 (91.6%)	1597 (99.3%)
$\langle \text{SNR}_{\text{out}} \rangle$ (dB)	13.7 ($\sigma=2.8$)	20.4 ($\sigma=2.3$)
Both arrays jointly equalized		
Checksum correct	1532 (95.3%)	1607 (99.9%)
$\langle \text{SNR}_{\text{out}} \rangle$ (dB)	15.0 ($\sigma=2.7$)	22.4 ($\sigma=1.9$)

quantity is specific for the transmitter end of the links, differs between S1 and S2, and is unlikely to represent the entire communication channel. It is available only for data packets received without errors.

B. Receiver performance

Application of the receiver described in Sec. II C to the recorded data leads to the results in Table II. Receiver performance is considered in terms of the percentage of data packets which are successfully Reed Solomon decoded, and by the mean value of the receiver output SNR. The former figure is of interest to the end user of a communication system, whereas the output SNR allows qualitative and quantitative assessments on a continuous scale. The main observations from Table II are that S1 presents a more difficult channel than S2, and that the vertical array delivers a considerably greater gain with S1 than the horizontal array. S1 success percentages are 91.6% on Mathilde versus 66.2% on Delphinus at a 3.4-dB advantage in mean output SNR. The link from S2 proves to be highly reliable with more than 99% success on both arrays. Oddly, for this modem the percentage of correct checksums is better on Delphinus, while the mean output SNR is 2 dB higher on Mathilde. Inspection of the S2 recordings for which only Delphinus yields a correct checksum reveals that they all coincide with periods of severe shipping noise. This suggests that in such cases Delphinus may profit from beamforming in the horizontal plane to suppress the shipping noise. Overall, however, the vertical array outperforms the horizontal array.

Table II also lists the results for joint equalization of the two arrays. To this end the receptions on Delphinus and Mathilde are first synchronized on the training sequence and subsequently fed to a 32-channel equalizer, still subject to the parameters of Table I. A further improvement is observed with the mean output SNR rising to 15 dB for S1. Its score is now 95.3%, outperforming the and/or operation on two separately equalized arrays (93.1%). For S2 the results were already satisfactory on the individual arrays and joint equaliza-

TABLE III. Mean value and standard deviation of parameters extracted from the raw acoustic data. All levels are measured in the frequency band of the communication signal.

	Horizontal array				Vertical array			
	S1		S2		S1		S2	
	Mean	σ	Mean	σ	Mean	σ	Mean	σ
Time spread (ms)	15.4	4.5	3.9	1.7	16.9	5.6	5.8	3.7
Signal level (dB re 1 μ Pa)	117	3	126	3	119	3	128	2
rms noise level (dB re 1 μ Pa)	102	6	102	6	102	7	102	6
Peak noise level (dB re 1 μ Pa)	122	8	122	8	123	9	123	9
Spatial coherence (%)	55	8	77	7	42	13	59	18

tion increases the mean output SNR by another 2 dB, indicating the feasibility of higher-order constellations to increase bit rate.

Table III shows why S1 is more difficult for the receiver than S2. It lists mean values and standard deviations of the acoustic parameters described in Sec. III A. Specifically, for the horizontal array the time spread is computed for an array steered in the known direction of the modem. This beamforming, which does not noticeably alter the channel impulse response, increases the SNR and allows a more reliable measurement of the time spread. For the vertical array the time spread is measured on each element and simply averaged, and the same applies to signal and noise levels on both arrays. Table III reveals that the time spread is much larger for S1 than for S2. The signal levels are slightly higher on Mathilde, whereas the noise levels do not differ between the two arrays. The difference between the signal level and the rms noise level gives the input SNR [Eq. (4)], which—on average—looks adequate for S1 and luxurious for S2. However, the 9-dB lower input SNR combined with the severe delay spread does explain why S1 is a more difficult modem than S2. Matters are further aggravated by the peak noise level Eq. (1), which averages above the signal level Eq. (3) of S1. No less than 64% of the S1 receptions have a peak noise level exceeding the signal level. The likely cause of these high peak levels is cavitation of ship propellers, whose sound is characterized by a high peak-to-average power ratio.¹² Values up to 160 dB re 1 μ Pa have been registered for extremely noisy or nearby ships. The last parameter in Table III, the spatial coherence between the extreme hydrophones, is higher along the horizontal array and also favors S2 to S1.

C. Correlations

So far we have considered mean values of the receiver performance and acoustic parameters. Figure 7 reveals that this averaging removes interesting temporal features. It displays the receiver output SNR for both subsea units on both arrays, together with the water level, water temperature, impulse response, and received signal level. The latter two quantities do not differ much between the arrays and are only shown for the horizontal Delphinus array. Figure 7 covers a period of 64 h, equivalent to five tidal cycles. The tidal influence is clearly revealed on the channel impulse response. There is a complex periodic structure of the dominant mul-

tipath components for S1, frequently complemented with dense reverberation tails with time delays of 25 ms and beyond. The main multipath components are more closely spaced for S2 and long reverberation tails are absent. For S2 the tide manifests itself not only in the impulse response, but also in the signal level and the output SNR. For reference the water level is included in the graphs of the S2 output SNR. Notice that it is plotted with reversed phase and that the highest SNRs occur at low tide. The increased signal level and decreased time spread operate in unison to facilitate S2 demodulation at low tide. Conversely, a reduction in signal level and increased time spreading render demodulation more difficult at high tide. For S1 there is no apparent effect of the tide on the receiver output SNR, despite its influence on the main propagation paths below 10 ms. The reasons for this appear to be the uncorrelated presence of multipath at delay times beyond 10 ms, and the absence of a water-level effect on the signal level.

The distinct tidal influence on the acoustic parameters of S2 is explained by the large range/depth ratio. The tidal range is a substantial fraction of the nominal water depth, more than 50% at the position of the receive barge. Patterns such as those observed in the time spread graphs of Fig. 7 have been previously described by Weston *et al.* for one-way transmission¹³ and echo-ranging¹⁴ experiments in shallow coastal waters. In their case the cause is identified as being due to interference between the normal modes of propagation in a shallow-water waveguide. Tidal changes in water depth effectively sweep a modal interference pattern past a fixed receiver. The water structure and the depth dependence of the streaming velocity can also vary through the tide, often spoiling perfect symmetry.¹³ That last observation could account for the small mismatch between the water level curve and receiver output SNR in Fig. 7. Especially for Mathilde there exists a phase offset, or at least an asymmetry, implying that the modulation is not due to the water level alone.

Presently it is not known why the behavior of S1 strongly differs from S2 in terms of signal level and time spread. The average signal levels differ by 9 dB, far more than the expected propagation loss difference due to cylindrical or spherical spreading alone. Sound speed profiles and seafloor properties are not available for propagation modeling. The water temperature near the bottom, measured by the

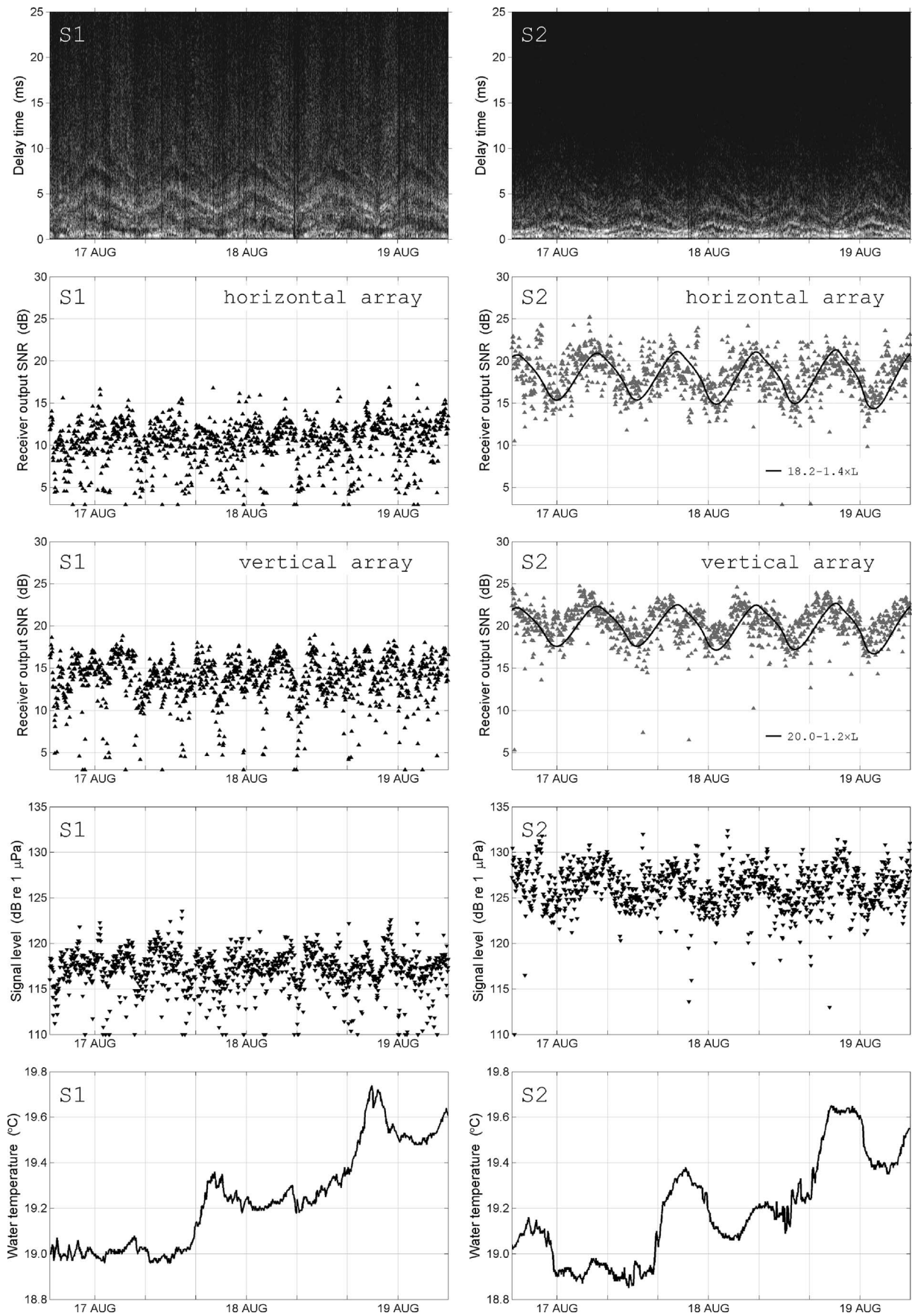


FIG. 7. Channel impulse response vs time of day (year 2005), alongside receiver output SNR, signal level, and water temperature. The water level L (in meters re NAP) is included in two graphs as a guide to the eye. The signal level and impulse response are both extracted for the horizontal array. An accumulated energy threshold of 2% is used to align successive impulse responses at zero delay.

TABLE IV. Correlation coefficients between the receiver output SNR and a set of acoustic and environmental parameters.

	Horizontal array		Vertical array	
	S1	S2	S1	S2
Signal level	0.56	0.75	0.68	0.66
rms noise level	-0.25	-0.12	-0.46	-0.41
Peak noise level	-0.23	-0.10	-0.44	-0.40
Spatial coherence	0.73	0.74	0.63	0.53
Time spread	-0.36	-0.69	-0.63	-0.59
Water level	-0.03	-0.46	-0.02	-0.36
Water temperature	0.06	0.04	0.06	0.13
Wind speed	-0.27	-0.24	-0.32	-0.18

ADCPs (lowermost panel in Fig. 7), essentially beats at twice the tidal cycle and does not correlate well with any other parameter.

Quantification of relationships between the output SNR and the parameters defined in Sec. III A is achieved by computing correlation coefficients. The correlation coefficient between two observables, defined as their covariance divided by the product of their standard deviations, gives the degree to which there exists a linear relationship. Care is required in interpretation, because existing relationships may be nonlinear, and because mere covariation does not necessarily imply a causal dependence. However, we can at least say that relative comparisons of coefficients expose valid behavioral differences between the two subsea units or between the two arrays.

Table IV shows the correlation coefficients. Each correlation with an absolute value greater than 0.1 is significant with a so-called P value (much) smaller than 10^{-3} . P gives the probability of obtaining a correlation as large as the observed value by random chance, when the true correlation is zero. Unsurprisingly the output SNR is negatively correlated with additive noise and time spreading. The SNR is strongly positively correlated with the signal level and coherence. Table IV also confirms the influence of the tide previously noticed in Fig. 7. S2 is negatively correlated, but S1 totally uncorrelated with the water level. Correlations with the water temperature are weak to absent for any transmitter or receiver. Correlations involving the signal level, coherence, time spread, and water level cannot be seen separately, because these quantities correlate strongly with one another—certainly for S2. In contrast, the rms and peak noise levels do not correlate with any other parameter except each other. Table IV reveals that the horizontal array is more noise resistant than the vertical array, with S2 being particularly insensitive. Finally, the negative correlation with the wind speed is significant but no proof of an actual dependence. The correlation with the wind is particularly strong during a thunderstorm on August 19, during which there was a marked reduction in the output SNR for both subsea units on both arrays. This reduction is due to an increased noise level and an even greater decrease in received signal level. A candidate physical cause is wind-generated waves, which produce noise and may adversely affect propagation through an enhanced surface roughness. However, another possible

cause is the noise produced by the concurrent rainfall (no data available for correlation) and signal attenuation due to air bubbles injected by raindrops. It is difficult to separate the effects of wind and rain, also because rain noise is influenced by the wind.¹⁵

D. Example receptions

Figure 8 shows example spectrograms of received signals which illustrate several challenges of acoustic communication in the dynamic waterway. Graph A shows a signal captured at a quiet moment with a reception that stands out clearly against a mild noise background. Graph B reveals impulsive wideband noise attributed to ship propeller cavitation,^{12,16,17} characterized by high peak noise levels. Graph C: Raindrops impinging upon the water surface only a few meters above the receive arrays during the August 19 thunderstorm. The downpour, together with strong winds, introduced broadband noise and a drop in signal level. Several S1 broadcasts were lost. Graph D: A speedboat maneuvering in the vicinity of the measurement pole also caused considerable problems. High noise levels overpowered the modem signals, with S1 reception beyond rescue and S2 barely coping. Graph E illustrates the reception of S1 shortly after the Maersk Duisburg, a 292-m container ship with a draft of 13.5 m, sailed over the subsea unit. There is no trace of the modem signal and this is most likely due to the creation of bubble clouds, which seem to be a common feature of all ship wakes.¹⁸ Air bubbles are strong scatterers and absorbers of sound, and, when present in large numbers, can act as an impervious barrier. Two consecutive broadcasts of S1 were completely lost after the passage of the ship, with the modem signal reemerging after ten minutes. Such a channel recovery time is in agreement with characteristic wake lifetimes of 7–15 min reported in literature.^{18,19} Ship wakes manifest themselves in Fig. 7 as the low-energy receptions causing the black stripes in the S1 impulse response graph. Log files of the AIS recorder show that all vanished modem signals coincide with freighters passing, or just having passed, the line of sight between S1 and the receiver station.

Finally, two examples of the vertical array reception in the presence of ship noise are shown in Fig. 9. The illustration for S1 shows a signal with a poor input SNR and an intricate impulse response. Joint equalization of all 16 channels nonetheless results in adequately resolved QPSK constellation clouds at an output SNR of 10.9 dB. The example for S2 reveals a higher input SNR, a benign impulse response, and a high output SNR at 18.6 dB. However, even in this case there are symbol errors that need correction by the Reed Solomon decoder. The direct cause is the impulsive propeller noise visible in the time series, which causes transient excursions of the mean square error. The short-lived distortions are gone before the equalizer can react, which results in scatter of a few symbols only. However, such error bursts might lead to detrimental error propagation phenomena in the DFE structure which can only be offset by powerful burst error correction codes.

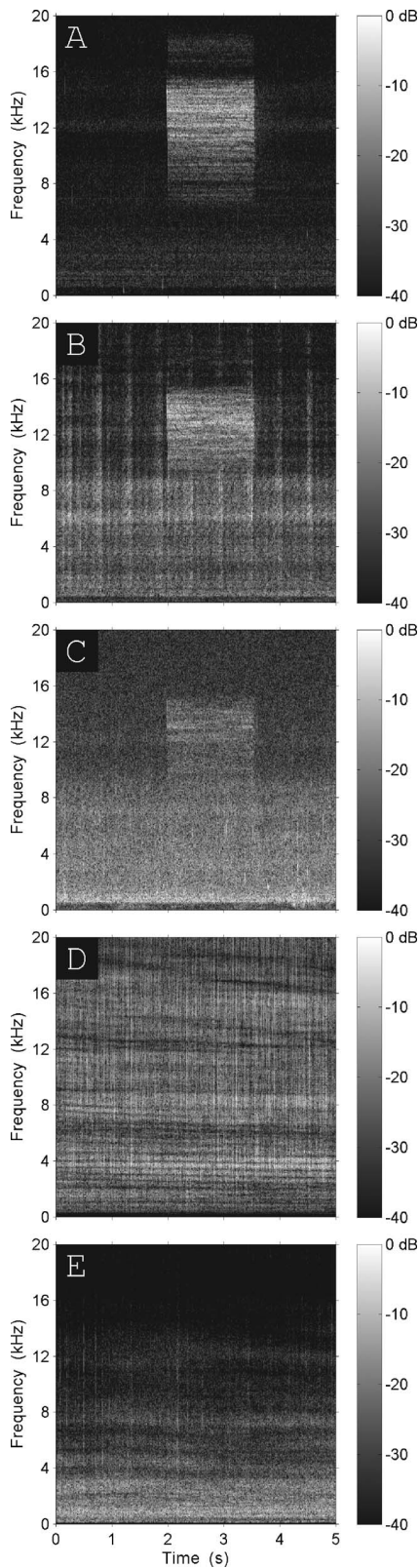


FIG. 8. Spectrograms illustrating the reception of S1 on a Delphinus hydrophone. (A) Quiet moment, (B), propeller clicks, (C), heavy rainfall, (D) nearby speedboat, and (E) wake of a container ship.

IV. MECHANISMS OF ARRAY GAIN

The purpose of this section is to investigate in which manner the adaptive structure of Fig. 6 achieves its gain. To shed light on this matter, the receiver performance is first

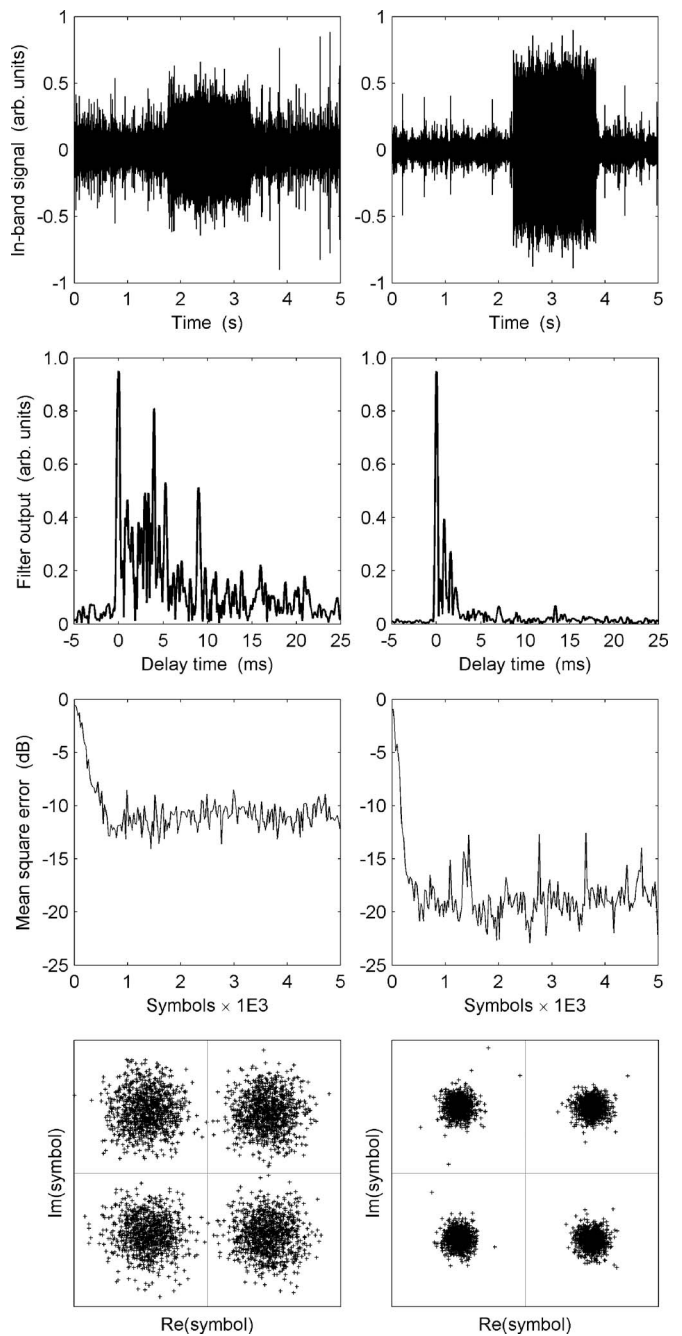


FIG. 9. Examples of the reception of S1 and S2 on the vertical array. Notice that the time series and the impulse response are shown for a single channel, whereas the mean square error and symbol constellation result from joint equalization of sixteen channels.

considered for different combinations of hydrophones. Further insight is provided by examination of array angular responses.

A. Variable number of hydrophones

The recorded data have been processed in batch jobs for different hydrophone configurations. In each batch job the receiver demodulates all received signals using a specified combination of hydrophones. Subsequently the percentage of successfully decoded packets and mean output SNR are determined, as in Table II, which was obtained for demodulation of all 16 hydrophones. Two strategies are followed to

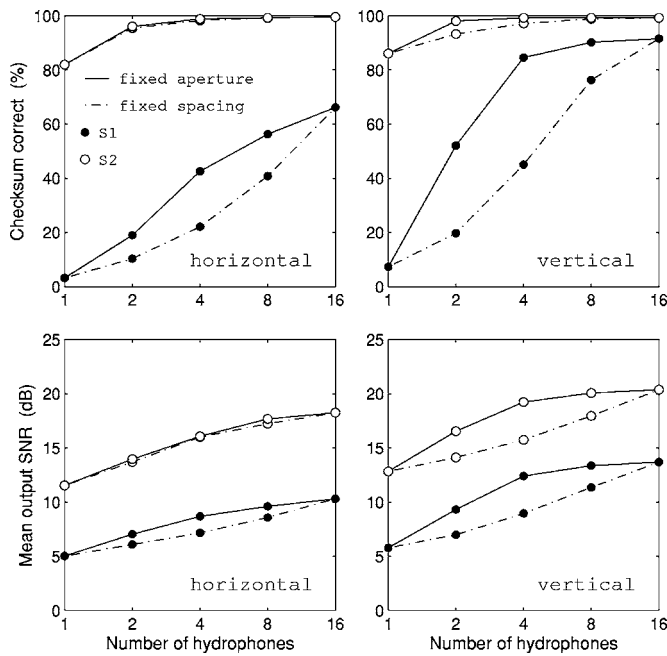


FIG. 10. Performance charts vs. number of hydrophones. Each marker represents 1608 received signals. Hydrophone selection for the fixed aperture: 1; 1/16; 1/6/11/16; 1/3/5/7/10/12/14/16; 1–16. Fixed spacing: 1; 1/2, 1–4, 1–8, 1–16.

span the range between a single hydrophone and the full array. The method labeled “fixed aperture” in Fig. 10 keeps a constant aperture, evenly filling the space between hydrophones 1 and 16. In contrast, the method labeled “fixed spacing” uses a constant distance between successive hydrophones. Here the aperture grows with an increasing number of elements.

Inspection of Fig. 10 shows that the single hydrophone of the vertical array performs a little better than a single hydrophone of the horizontal array. This is because the vertical array moves up and down with the tide and does not experience precisely the same conditions as the horizontal array, which is at a fixed distance from the bottom. More interesting are the transitions between a single hydrophone and the complete array. On the vertical array the fixed-aperture strategy outperforms the fixed-spacing strategy by a considerable margin. The leaps in success percentage or output SNR from one to two, and from two to four elements are much larger with a constant aperture. The difference amounts to a sizable 3.5 dB for the comparison with four hydrophones. On the horizontal array, however, the advantage of the constant-aperture approach is much smaller. The SNR gain for S1 amounts to only 1.5 dB at four hydrophones, whereas the reception of S2 is virtually independent of the hydrophone selection strategy.

B. Array angular response

The adaptive structure of Fig. 6 is a versatile tool capable of both beamforming and diversity combining. An indication of its operation mode in specific cases is obtained by deriving its angular response. To this end the feedforward filter coefficients are frozen at the end of the training sequence. The static filter is subsequently exposed to plane

waves incident at angles between -90° and $+90^\circ$, where 0° is broadside. The waves are phase-shift keyed waveforms composed of random symbols. That is to say, a waveform is used which is otherwise identical to the AM300 modem signals. At each angle the $T/2$ -spaced waveform is properly delayed on the successive array elements, and the power of the summed feedforward filters is measured. This procedure yields an indication of the angular response of the array. Representative examples, one for each modem/array combination, are shown in Fig. 11.

As the known locations of the modems correspond to steering angles of $+6^\circ$ (S1) and $+32^\circ$ (S2), the horizontal array responses reveal that the feedforward filter section has pin-pointed the directions of the modems in the horizontal plane. A distinct main lobe indicates that the filter coherently adds the hydrophone arrivals originating from the modem direction, effectively suppressing noise from other directions. The angular responses derived for the vertical array show a different picture. They are characterized by a broadened response around broadside, composed of two or three separate maxima. These multiple maxima suggest that the filter coherently combines signal components arriving at different angles in the vertical plane, potentially also using the minima in between to null out interference in the form of inconvenient multipath arrivals or ship noise. The shown examples are representative in that there is always a single main lobe on the horizontal array, and nearly always a broader response composed of multiple maxima on the vertical array.

C. Discussion

Fig. 10 revealed that, within limits, the aperture of the vertical array is a more important parameter than the number of elements. Two hydrophones covering the full aperture outperform four hydrophones covering a quarter of the aperture, and four hydrophones covering the full aperture outperform eight hydrophones covering half the aperture. Such behavior can be interpreted as support for the view that the multichannel equalizer acts as a diversity combiner.²⁰ When hydrophones are too closely spaced neighboring elements receive essentially the same information and there is little diversity to be exploited. To profit optimally from diversity gain the spacing should be at least of the order of the vertical coherence length, which is the distance over which the coherence drops by, say, 50%. However, the information in Table III shows that the mean coherence between the top and bottom hydrophone is 42% for S1 and 59% for S2. As the coherence fall-off between neighboring hydrophones is much smaller than these values, the element spacing is much shorter than the coherence length. This observation prompts an alternative explanation that can account for the importance of the aperture. If the interpretation of the vertical array beam patterns of Fig. 11 is correct, i.e., that the filter coherently combines signal arrivals in a relatively narrow angular interval, a large aperture is required to deliver the required angular resolution. In other words, we have a beamforming application for which a high angular resolution is a more important requirement than low sidelobe levels. Spatial aliasing, which arises when elements are spaced further apart than half the

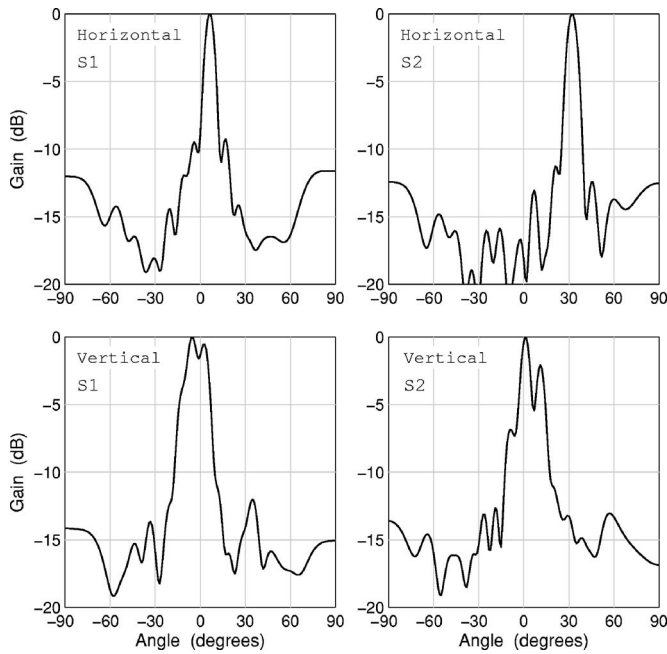


FIG. 11. Broadband (8–16 kHz) angular responses derived from the feed-forward filter coefficients. Negative angles point downward on the vertical array and roughly eastward on the horizontal array.

wavelength, is tolerable provided the grating lobes do not point toward interferers and indeed the adaptive algorithm attempts to avoid this condition.

At broadside a horizontal array has no angular resolution in the vertical plane. Steered away from broadside it does offer some discrimination, but much less than the same array deployed vertically. Coherent combining or cancellation of individual multipath components is hardly an option. Beamforming in the horizontal plane does not specifically require a narrow main lobe because possible interference is mostly restricted to shipping noise, which can come from any direction. From a beamforming point of view it is no surprise that the difference between the fixed-aperture and fixed-spacing strategies in Fig. 10 is small on the horizontal array, as array gain is generally proportional to the number of elements for a half-wavelength spacing or larger.

The angular responses of Fig. 11 suggest that the adaptive receiver acts as a coherent combiner and that beamforming is the primary mechanism of operation. Angular responses derived from a pure diversity combiner tend not to show distinct lobes in any particular direction. Uncertainty related to the presence of the feedback section is removed by noting that its omission from the equalization process, although reducing the output SNR, does not alter the angular responses other than by a few decibels in the sidelobes. The finding remains that beamforming and coherent combining at least both play an important and complementary role. Table IV corroborates this finding through strong positive correlations between the receiver output SNR and the coherence—a requirement for beamforming—for each modem/array combination. Perhaps the strongest evidence is furnished by the beam patterns in Fig. 12. These patterns are obtained by feeding the signals used for Fig. 11 to a classical beamformer. At an input SNR of more than 18 dB for all ex-

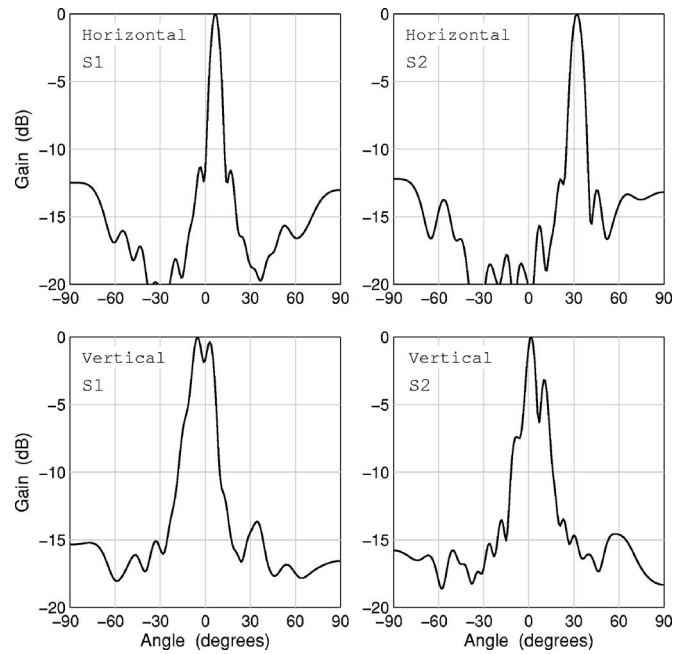


FIG. 12. Beam patterns obtained with a conventional delay-and-sum beamformer for an unshaded array.

amples, these beam patterns reveal the angular distribution of signal energy. The similarities between Figs. 11 and 12 are simply striking. In this light it is also interesting to note that a delay-and-sum beamformer in conjunction with a single-channel equalizer makes a suitable receiver for the horizontal array. A previous paper² even reported a somewhat better performance compared with the multichannel DFE of the present paper. This advantage, however, must be ascribed to the use of the capable RLS tap update algorithm, whose computational load is acceptable for a single channel. In contrast, a classical beamformer utilizing a single beam leads to a poor receiver performance on the vertical array, regardless of the precise steering angle and tracking algorithm.

As the sound is not fully coherent along the arrays, there is some diversity to be exploited. For modem S2 received on the horizontal array the element spacing is apparently unimportant, from which it is concluded that the contribution of diversity gain to the total array gain is negligible. For the other modem/array combinations it is difficult to quantify the contributions of diversity gain and beamforming gain separately. A loose estimate is obtained by noting that the coherence lengths are of the same order as the array aperture, which could in principle deliver a ~ 3 dB diversity gain across the aperture. The remainder of the array gain would then need to originate from spatial filtering.

V. CONCLUSIONS

Four point-to-point acoustic communication links were tested in estuarine channels characterized by busy vessel traffic and a tidal range that is comparable to the water depth. The four links were realized between two seafloor modems placed in a navigation channel, and one horizontal and one vertical array deployed near the shore. A multichannel DFE was employed to demodulate modem signals received on both arrays during a four-day time span. The horizontal array

proves more successful in scenarios limited by shipping noise, but the vertical array offers the best performance overall. For one subsea unit the receiver output SNR is modulated by the tidal cycle through the channel impulse response and received signal level. The other subsea unit is susceptible to the wake of passing bulk carriers, which strongly attenuates the communication signals. It is argued that the primary mechanism of gain from the horizontal array is beamforming, giving an overall gain in SNR particularly in the case of dominant noise from passing ships. Coherent addition and cancellation of angularly separated multipath arrivals are considered to account for the gain of the vertical array, which also exploits diversity across its aperture.

From this study it is concluded that the success of an environmental monitoring system, relying on high-rate underwater acoustic communication, depends critically on the position of transmitters and receivers, the presence of a hydrophone array, its orientation, and the element spacing. On the vertical array described above, at least 90% of 1-kbyte data packets, transmitted at a raw rate of 8 kbit/s, could be received without error throughout a long-term deployment. Incremental improvements to the present system can be achieved, for example, by introducing more coding or by resorting to the computationally demanding RLS filter update algorithm. However, detrimental events like the attenuation of modem signals by ship wakes still preclude 100% availability of communication links in the Western Scheldt environment. If real-time requirements are relaxed, the availability can be further enhanced by automatic repeat request protocols, or by building in intelligence to avoid transmission as large vessels are passing. Alternatively, multiple redundant receiver stations could be deployed around the estuary.

ACKNOWLEDGMENTS

Tritech International Ltd. are acknowledged for supplying and adapting acoustic modems, and Hoondert & Zoon B.V. for invaluable support at sea. Further, the DACAPO sea trials would not have succeeded without the commitment of many colleagues from Rijkswaterstaat and TNO, with honorable mention to Jan van 't Westende and Jeroen Janmaat. The authors are finally obliged to Mathilda L. Driesenaar for loaning the aptly named vertical array just one week before her own sea trials.

- ¹J. van Veen, "Ebb and flood channel systems in the Netherlands tidal waters," *J. Royal Dutch Geogr. Soc.* **67**, 303–325 (1950).
- ²P. A. van Walree, M. C. Schrijver, and J. A. Neasham, "Acoustic communication for the monitoring of current and vessel draft," *Proceedings of the Eighth European Conference on Underwater Acoustics ECUA 2006*, Carvoeiro, Portugal, pp. 843–848.
- ³D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," *IEEE J. Ocean. Eng.* **25**, 4–27 (2000).
- ⁴B. S. Sharif, J. Neasham, O. R. Hinton, and A. E. Adams, "A computationally efficient Doppler compensation system for underwater acoustic communications," *IEEE J. Ocean. Eng.* **25**, 52–61 (2000).
- ⁵N. M. Carbone and W. S. Hodgkiss, "Effects of tidally driven temperature fluctuations on shallow-water acoustic communications at 18 kHz," *IEEE J. Ocean. Eng.* **25**, 84–94 (2000).
- ⁶R. J. McDonald, K. W. Commander, J. S. Stroud, J. E. Wilbur, and G. B. Deane, "Channel characterization for communications in very shallow water," *J. Acoust. Soc. Am.* **111**, 2407 (2002).
- ⁷M. Siderius, M. B. Porter, P. Hursky, V. McDonald, and the KauaiEx Group, "Effects of ocean thermocline variability on noncoherent underwater acoustic communications," *J. Acoust. Soc. Am.* **121**, 1895–1908 (2007).
- ⁸S. P. Beerens, J. Kromjongh, M. van Spellen, and S. P. van IJsselmuide, "Sea trial of Delphinus; a towed array for marine mammal detection," *Proceedings of the European Conference on Undersea Defence Technology, UDT Europe 2005*, Amsterdam, the Netherlands.
- ⁹<http://www.tritech.co.uk/products/datasheets/acoustic-modem-300.pdf>, last accessed on February 7, 2007.
- ¹⁰M. Stojanovic, J. A. Catipovic, and J. G. Proakis, "Adaptive multichannel combining and equalization for underwater acoustic communications," *J. Acoust. Soc. Am.* **94**, 1621–1631 (1993).
- ¹¹J. A. Neasham, D. Thompson, A. D. Tweedy, M. A. Lawlor, O. R. Hinton, A. E. Adams, and B. S. Sharif, "Combined equalisation and beamforming to achieve 20 kbit/s acoustic telemetry for ROVs," *Proceedings of OCEANS 1996*, Ft. Lauderdale, FL, pp. 988–993.
- ¹²P. T. Arveson and D. J. Vendittis, "Radiated noise characteristics of a modern cargo ship," *J. Acoust. Soc. Am.* **107**, 118–129 (2000).
- ¹³D. E. Weston, A. A. Horrigan, S. J. L. Thomas, and J. Revie, "Studies of sound transmission fluctuations in shallow coastal waters," *Philos. Trans. R. Soc. London, Ser. A* **265**, 567–606 (1969).
- ¹⁴D. E. Weston and P. D. Hocking, "Interference patterns in shallow-water reverberation," *J. Acoust. Soc. Am.* **87**, 639–651 (1990).
- ¹⁵B. B. Ma, J. A. Nystuen, and R.-C. Lien, "Prediction of underwater sound levels from rain and wind," *J. Acoust. Soc. Am.* **117**, 3555–3565 (2005).
- ¹⁶Y.-C. Wang and C. E. Brennen, "The noise generated by the collapse of a cloud of cavitation bubbles," *Cavitation and gas-liquid flow in fluid machinery and devices*, ASME FED **226**, 17–29 (1995).
- ¹⁷T. H. Eggen, "Phase coherent communication in the presence of ship noise," *Proceedings of OCEANS 2000*, Providence, RI, pp. 1433–1436.
- ¹⁸M. V. Trevorror, S. Vagle, and D. M. Farmer, "Acoustical measurements of microbubbles within ship wakes," *J. Acoust. Soc. Am.* **95**, 1922–1930 (1994).
- ¹⁹M. Trevorror, "High-frequency acoustic scattering and absorption effects within ship wakes," *Proceedings of Boundary Influences in High Frequency, Shallow Water Acoustics*, Bath, UK, 2005, pp. 439–446.
- ²⁰T. C. Yang, "Measurements of spatial coherence, beamforming gain and diversity gain for underwater acoustic communications," *Proceedings of OCEANS 2005*, Washington, D.C., pp. 268–272.

Optimal illumination and wave form design for imaging in random media

Liliana Borcea

Computational and Applied Mathematics, MS 134, Rice University, 6100 Main Street, Houston, Texas 77005-1892

George Papanicolaou

Department of Mathematics, Stanford University, Stanford, California, 94305

Chrysoula Tsogka^{a)}

Department of Applied Mathematics, University of Crete, Heraklion-Crete 71409, Greece

(Received 16 May 2007; revised 30 August 2007; accepted 22 September 2007)

The problem of optimal illumination for selective array imaging of small and not well separated scatterers in clutter is considered. The imaging algorithms introduced are based on the coherent interferometric (CINT) imaging functional, which can be viewed as a smoothed version of travel-time migration. The smoothing gives statistical stability to the image but it also causes blurring. The trade-off between statistical stability and blurring is optimized with an adaptive version of CINT. The algorithm for optimal illumination and for selective array imaging uses CINT. It is a constrained optimization problem that is based on the quality of the image obtained with adaptive CINT. The resulting optimal illuminations and selectivity improve the resolution of the images significantly, as can be seen in the numerical simulations presented in the paper.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799502]

PACS number(s): 43.60.Pt, 43.60.Gk, 43.60.Tj [PEB]

Pages: 3507–3518

I. INTRODUCTION

We consider the imaging of small scatterers in cluttered media from the time traces of echoes recorded at a remote array. By clutter we mean inhomogeneities in the medium that are unknown and cannot be estimated in detail. In regimes with significant multiple scattering of the waves by the clutter, the time traces have considerable delay spread (coda) and their travel-time migration creates images with speckles that are difficult to interpret. The images are also unstable, which means that they change unpredictably with different realizations of the clutter.

To stabilize the imaging process in clutter, we introduced^{1,2} the coherent interferometric (CINT) imaging functional. The images produced by CINT are statistically stable, that is, they have small variance with respect to random background fluctuations. Let us emphasize that it is the images that have small variance and not the background fluctuations, which have a sizable cumulative effect on wave propagation, and hence on the array data. CINT is a smoothed version of travel-time (Kirchhoff) migration^{3,4} in which the delay spread is reduced by using the cross correlations of the traces over appropriate space-time windows,^{1,5,6} rather than the traces themselves. The CINT images are then formed with travel-time migration of the local cross correlations. The choice of the space-time windows depends on the clutter and it affects both the statistical stability and the resolution of the image. We showed in Ref. 5 that there is a trade-off between these two effects, and we

introduced an adaptive algorithm for selecting the size of the space-time windows so as to achieve an optimal compromise between gaining statistical stability and losing resolution by blurring.

In this paper we consider the imaging of a cluster of small scatterers that are close together and, because of the inevitable blur, cannot be distinguished with adaptive CINT. We introduce an algorithm for optimal illumination and selective imaging of the scatterers one at a time, in which we use the singular value decomposition (SVD) of the array response matrix, frequency by frequency.

The SVD of the response matrix has been used before^{7–9} for obtaining optimal illuminations for detection of scatterers from the traces recorded at the array. These illuminations maximize the power of the signals recorded by the array. They are determined by the leading singular vector of the response matrix at the resonant frequency, which is where the leading singular value achieves its maximum over the bandwidth. However, the optimal illumination for detection is not appropriate for imaging because (1) the narrow bandwidth gives poor range resolution of the images and no statistical stability in clutter,^{6,10,11} and (2) the leading singular vector does not necessarily illuminate a single (strongest) scatterer in the cluster. In general, this occurs only when the scatterers are well separated.

Well separated small scatterers can be imaged selectively in known media with the DORT method.¹² DORT stands for decomposition de l'opérateur de retournement temporel in French, which uses one singular vector at a time, over the entire bandwidth. However, DORT does not image selectively in clusters of nearby scatterers because the one-

^{a)}Author to whom correspondence should be addressed. Electronic mail: tsogka@tem.uoc.gr

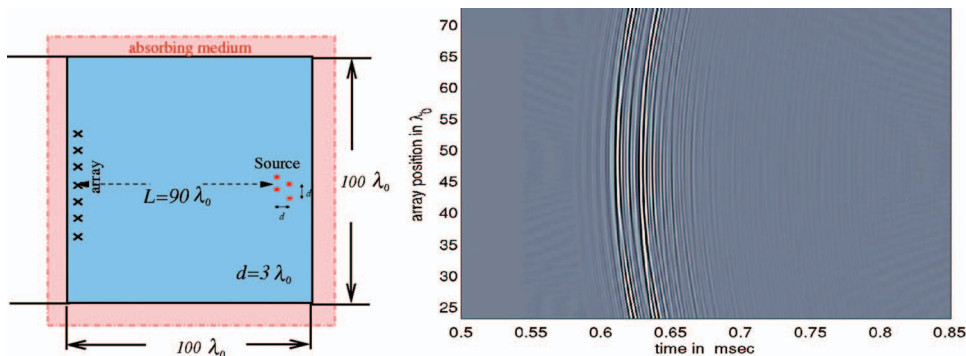


FIG. 1. Schematic of array imaging in a homogeneous medium. The computational setup is on the left. The traces received when illuminating from the central element in the array are on the right.

to-one correspondence between the scatterers and the singular vectors does not hold. It also does not give stable images when there is clutter.

There are three parts to the optimal illumination algorithm for selective imaging of clusters of small scatterers as presented in this paper. First, all imaging is based on the adaptive CINT functional, so that it is statistically stable in clutter. Second, the illumination is not done with one singular vector at a time, but with an optimal convex combination of the leading singular vectors across the bandwidth. The coefficients (subspace weights) are calculated by solving a constrained optimization problem which minimizes the support of the image, which is formed with adaptive CINT. Third, once the optimal subspace selection has been done so that the image is focused on a single scatterer, it is further improved by assigning optimal weights to each illuminating source in the array. These weights are also constrained minimizers of the support of the image.

The paper is organized as follows. In Sec. II we describe the setup for array imaging of a cluster of small scatterers. In Sec. III we describe the adaptive CINT algorithm and its relation to travel-time migration. In Sec. IV we consider the optimal illumination problem. We define the weighted subspaces using the SVD of the array response matrix in Sec. IV A, and then introduce the optimal illumination algorithm in Sec. IV B. We illustrate the performance of the algorithms with numerical simulations in the regime of ultrasonic array imaging in concrete.^{13,14} Details and comments on the numerical results are given in Sec. V. We end with a summary and conclusions in Sec. VI.

II. PROBLEM SETUP

We consider imaging of a cluster of M small (point-like) scatterers at nearby locations $\mathbf{y}^{(j)}$ for $j=1, \dots, M$, in a cluttered medium. The data are gathered with an array \mathcal{A} of N_s sources and N_r receivers that is far from the scatterers. The sources are at points $\mathbf{x}_s \in \mathcal{A}$, for $s=1, \dots, N_s$ and the receivers are at $\mathbf{x}_r \in \mathcal{A}$, for $r=1, \dots, N_r$. The array illuminates the scatterers sequentially by sending a pulse $f(t)$ from the sources at \mathbf{x}_s and recording for each pulse the traces $P(\mathbf{x}_r, \mathbf{x}_s, t)$ of the scattered echoes at the receivers

$$P(\mathbf{x}_r, \mathbf{x}_s, t) = f(t) \star \Pi(\mathbf{x}_r, \mathbf{x}_s, t) \quad (1)$$

for $r=1, \dots, N_r$ and $s=1, \dots, N_s$. Here $\Pi(t) = \{\Pi(\mathbf{x}_r, \mathbf{x}_s, t)\}$ is the $N_r \times N_s$ array impulse response matrix and the symbol \star stands for time convolution. The traces $P(\mathbf{x}_r, \mathbf{x}_s, t)$ are re-

corded over a time window that is assumed for simplicity to be long enough (essentially infinite) so as to neglect it in Eq. (1). We also take for convenience the case of $N_s = N_r = N$ collocated sources and receivers. By reciprocity, this makes the array response matrix $\Pi(t)$ symmetric.

The probing pulse is of the form

$$f(t) = e^{-i\omega_0 t} f_B(t), \quad (2)$$

where ω_0 is the central frequency and $f_B(t)$ is the base-band pulse, with bandwidth B . In the frequency domain we have

$$\begin{aligned} \hat{f}(\omega) &= \int_{-\infty}^{\infty} e^{i(\omega - \omega_0)t} f_B(t) dt \\ &= \hat{f}_B(\omega - \omega_0) \neq 0 \quad \text{for } |\omega - \omega_0| \leq \frac{B}{2}, \end{aligned} \quad (3)$$

where the symmetric interval of negative frequencies is suppressed for simplicity. We choose $\hat{f}_B(\omega)$ as the indicator function of the frequency interval $|\omega| \leq B/2$. This way, the Fourier coefficients of the traces coincide over the bandwidth with the Fourier coefficients of the response matrix

$$\hat{P}(\mathbf{x}_r, \mathbf{x}_s, \omega) = \hat{\Pi}(\mathbf{x}_r, \mathbf{x}_s, \omega) \quad (4)$$

for $s, r=1, \dots, N$ and $|\omega - \omega_0| \leq B/2$.

A. Setup for the numerical simulation

The numerical simulations are in two dimensions, in a regime that is often used in ultrasonic array imaging in concrete.^{13,14} In this paper, the elastic wave propagation in concrete is approximated by a scalar acoustic problem. Therefore, only pressure waves are considered and shear waves, Rayleigh waves, and mode conversion effects are neglected. We solve the acoustic wave equation as a first order velocity-pressure system with the finite element, time domain method given in Ref. 15, in two dimensions. The setup is shown in Figs. 1 and 2 for imaging in homogeneous and cluttered media, respectively. We simulate the wave propagation in an unbounded environment by surrounding the computational domain with a perfectly matched absorbing layer, as shown in Fig. 1.

The array is linear, with $N=100$ transducers at $\lambda_0/2$ apart and with an aperture $a=49.5\lambda_0$. We use the frequency band 150–450 kHz, with bandwidth $B=300$ kHz. The reference sound speed is $c_0=3$ km/s and therefore $\lambda_0=1$ cm. We have $M=4$ small scatterers at range $L=90\lambda_0$ from the array.

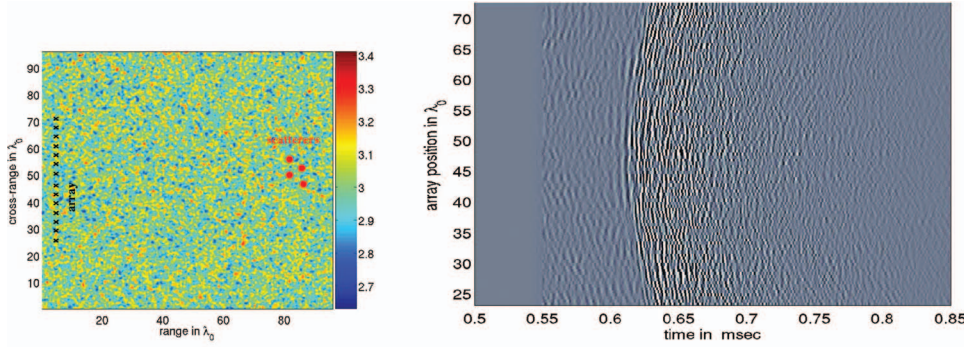


FIG. 2. Array imaging in clutter. The computational setup is shown on the left. The fluctuations in the sound speed are shown in color and the color bar is in km/s. The traces received when illuminating from the central element in the array are on the right.

They are at distance $d=3\lambda_0$ apart and we model them as soft acoustic disks of diameter λ_0 , with the acoustic pressure equal to zero at their boundary. In concrete they correspond to air holes.

We model the clutter as a random process and write the square of the index of refraction $n(\mathbf{x})=c_0/c(\mathbf{x})$ in the form

$$n^2(\mathbf{x}) = n_0^2(\mathbf{x}) \left[1 + \sigma \nu \left(\frac{\mathbf{x}}{\ell} \right) \right]. \quad (5)$$

Here $n_0(\mathbf{x})$ is the smooth and known index of refraction of the background medium. We take $n_0(\mathbf{x})=1$ for simplicity, so that the wave speed $c(\mathbf{x})$ fluctuates about the constant value c_0 . The normalized fluctuations are modeled by $\nu(\mathbf{x})$, which is a statistically homogeneous random process with mean zero and rapidly decaying correlation. The fluctuations have a characteristic length scale ℓ , the correlation length, for example, which can be considered to be the typical size of the inhomogeneities. The parameter σ controls the strength of the fluctuations.

The fluctuation process $\nu(\mathbf{x})$ is isotropic with Gaussian correlation

$$R(\mathbf{x}, \mathbf{x}') = R(|\mathbf{x} - \mathbf{x}'|) = e^{-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\ell^2}}.$$

The correlation length is taken to be $\ell=\lambda_0/2$ and the fluctuation strength $\sigma=0.03$. Note that we are in a regime with small fluctuations $\sigma \ll 1$, as is expected in concrete structures. Nevertheless, because the range L is large with respect to λ_0 and ℓ , there is significant delay spread in the traces, as seen in Fig. 2. The estimated transport mean free path^{16,17} in the clutter is $75\lambda_0$. This is to be contrasted with the time reversal experiments in Ref. 18, where the range is about ten transport mean free paths and all coherence is effectively lost in the echoes. Here the range L is comparable to the mean free path so there is some residual coherence in the data and coherent interferometric imaging can be effective.

III. ADAPTIVE COHERENT INTERFEROMETRIC IMAGING

In travel-time (Kirchhoff) migration the data traces are mapped to an image by the functional

$$\begin{aligned} \mathcal{I}^{\text{KM}}(\mathbf{y}^S) &= \int_{|\omega-\omega_0| \leq B/2} d\omega \sum_{r=1}^N e^{-i\omega\tau(\mathbf{x}_r, \mathbf{y}^S)} \\ &\times \sum_{s=1}^N \hat{P}(\mathbf{x}_r, \mathbf{x}_s, \omega) e^{-i\omega\tau(\mathbf{x}_s, \mathbf{y}^S)} \\ &= \sum_{r=1}^N \sum_{s=1}^N P(\mathbf{x}_r, \mathbf{x}_s, \tau(\mathbf{x}_r, \mathbf{y}^S) + \tau(\mathbf{x}_s, \mathbf{y}^S)). \end{aligned} \quad (6)$$

Here \mathbf{y}^S is a search point in the region where we form the image and $\tau(\mathbf{x}_r, \mathbf{y}^S)$ is the travel time of the waves from the array element \mathbf{x}_r to \mathbf{y}^S , in the background medium with sound speed c_0 . Since we assume a constant c_0 ,

$$\tau(\mathbf{x}_r, \mathbf{y}^S) = \frac{|\mathbf{x}_r - \mathbf{y}^S|}{c_0}.$$

In general but smooth media, τ is given by Fermat's principle.¹⁹

Travel-time migration of the traces when they have significant delay spread due to multiple scattering in clutter produces images with speckles that are difficult to interpret, as seen in the left picture of Fig. 3. The images are also unstable, in the sense that they change unpredictably with the realization of the clutter.

To stabilize the imaging process we introduced^{1,2,6} the coherent interferometric (CINT) approach which migrates to \mathbf{y}^S local cross correlations of the traces, computed over appropriately sized space-time windows. The CINT imaging function is

$$\begin{aligned} \mathcal{I}^{\text{CINT}}(\mathbf{y}^S; \Omega_d, \kappa_d) &= \int_{|\omega-\omega_0| \leq \frac{B}{2}} d\omega \int_{|\omega'-\omega_0| \leq \frac{B}{2}, |\omega-\omega'| \leq \Omega_d} d\omega' \\ &\times \sum_{r, r' \in \mathcal{X} \left(\frac{\omega+\omega'}{2}, \kappa_d \right)} \sum_{s, s' \in \mathcal{X} \left(\frac{\omega+\omega'}{2}, \kappa_d \right)} \\ &\times \overline{\hat{Q}(\mathbf{x}_r, \mathbf{x}_s, \omega; \mathbf{y}^S)} \hat{Q}(\mathbf{x}_{r'}, \mathbf{x}_{s'}, \omega'; \mathbf{y}^S), \end{aligned} \quad (7)$$

where the bar means complex conjugate and $\hat{Q}(\mathbf{x}_r, \mathbf{x}_s, \omega; \mathbf{y}^S)$ is the Fourier transform of the trace $P(\mathbf{x}_r, \mathbf{x}_s, t)$ migrated to \mathbf{y}^S

$$\hat{Q}(\mathbf{x}_r, \mathbf{x}_s, \omega; \mathbf{y}^S) = \hat{P}(\mathbf{x}_r, \mathbf{x}_s, \omega) e^{-i\omega[\tau(\mathbf{x}_s, \mathbf{y}^S) + \tau(\mathbf{x}_r, \mathbf{y}^S)]}. \quad (8)$$

Here the set of indices in the summation is defined by

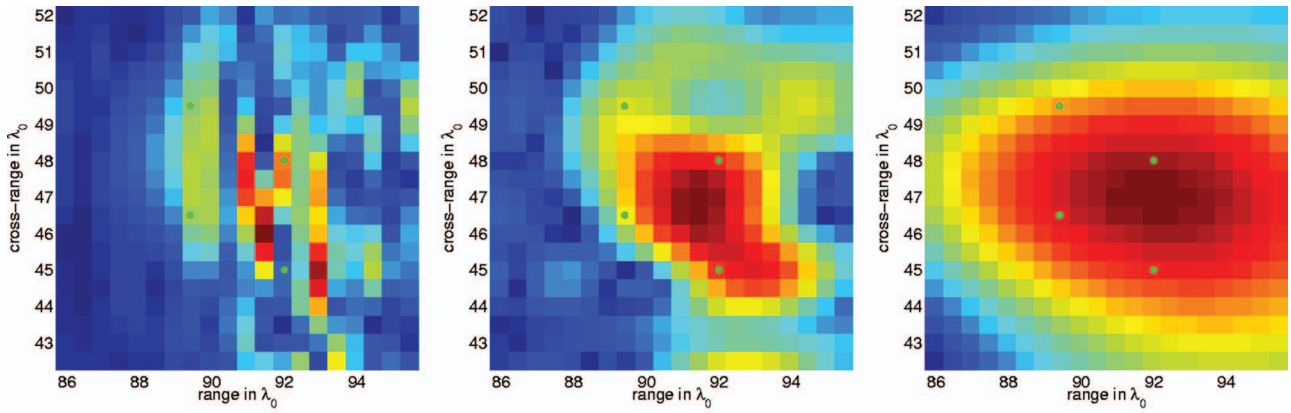


FIG. 3. The square root of the coherent interferometric imaging function for three choices of smoothing parameters. The left picture is with no smoothing, the middle picture is with optimal smoothing, and the right picture with too much smoothing. The four reflectors are indicated with green dots.

$$\chi(\omega, \kappa_d) = \left\{ r, r' = 1, \dots, N; |\mathbf{x}_r - \mathbf{x}_{r'}| \leq X_d(\omega) = \frac{c_0}{\omega \kappa_d} \right\}. \quad (9)$$

The Ω_d and κ_d are clutter-dependent coherence parameters that must be estimated from the data. They determine the size of the time-space windows used to compute the local cross correlations of the traces, as we now explain.

The data appear in $\mathcal{I}^{\text{CINT}}$ in the form $\hat{P}(\mathbf{x}_r, \mathbf{x}_s, \omega) \hat{P}(\mathbf{x}_{r'}, \mathbf{x}_{s'}, \omega')$ for receiver and source location indexes in $\chi[(\omega + \omega')/2, \kappa_d]$, and for frequencies that are not more than Ω_d apart. This Ω_d is the decoherence frequency in clutter, defined as the smallest difference $|\omega - \omega'|$ over which $\hat{P}(\mathbf{x}_r, \mathbf{x}_s, \omega)$ and $\hat{P}(\mathbf{x}_r, \mathbf{x}_s, \omega')$ become statistically uncorrelated. When we restrict $|\omega - \omega'| \leq \Omega_d$ we have significant random phase cancellation in the local cross-correlation $\hat{P}(\mathbf{x}_r, \mathbf{x}_s, \omega) \hat{P}(\mathbf{x}_{r'}, \mathbf{x}_{s'}, \omega')$. Equivalently, in the time domain, when we convolve $P(\mathbf{x}_r, \mathbf{x}_s, t)$ with its time reversed version over a time window of size $O(1/\Omega_d)$, we get a significant reduction of the delay spread.

The parameter κ_d defines the decoherence length $X_d(\omega)$, at frequency ω , as shown in Ref. 9. This decoherence length is the smallest separation between the sources $|\mathbf{x}_s - \mathbf{x}_{s'}|$ and between receivers $|\mathbf{x}_r - \mathbf{x}_{r'}|$ over which $\hat{P}(\mathbf{x}_r, \mathbf{x}_s, \omega)$ and $\hat{P}(\mathbf{x}_{r'}, \mathbf{x}_{s'}, \omega)$ become statistically uncorrelated. Thus, in CINT, we cross correlate the traces over space-time windows that are defined by the decoherence length and frequency in order to have significant reduction of the delay spread in the traces.

We showed^{5,6} that thresholding of the space-frequency windows with X_d and Ω_d amounts to smoothing the image and giving it statistical stability. In very weak clutter, where there is little loss of coherence in the data, X_d and Ω_d can be as large as the aperture a and bandwidth B , respectively. Then $\mathcal{I}^{\text{CINT}}$ becomes simply the square of \mathcal{I}^{KM} . However, in stronger clutter X_d and Ω_d can be small compared to a and B , respectively, and then $\mathcal{I}^{\text{CINT}}$ is a smoothed version of $|\mathcal{I}^{\text{KM}}|^2$. The smoothing occurs by convolution with the window functions whose effective support is controlled by Ω_d and κ_d , as we explain in detail in Refs. 5 and 6. We showed^{1,5} that the CINT point spread function has range resolution c_0/Ω_d and

cross-range resolution $\lambda_0 L/X_d(\omega_0) = 2\pi L \kappa_d$. These results should be compared with the Rayleigh resolution limits c_0/B and $\lambda_0 L/a$ for travel-time migration in homogeneous media.¹⁹ For strong clutter, where $\Omega_d \ll B$ and $X_d(\omega_0) \ll a$, there is considerable loss of resolution.

The key question in the implementation of $\mathcal{I}^{\text{CINT}}$ is how to estimate the clutter dependent parameters Ω_d and κ_d . This can be done directly from the data, using statistical techniques such as the variogram,²⁰ but it is a delicate process. We introduced⁵ an adaptive version of the CINT algorithm that estimates Ω_d and κ_d during the image formation process by optimization of an objective function that quantifies the quality of the image.

CINT is a trade-off between smoothing for statistical stabilization and loss of resolution by blurring. This can be seen in Fig. 3, where we display the $\mathcal{I}^{\text{CINT}}$ image computed with three different sets of parameters Ω_d and κ_d . On the left we show \mathcal{I}^{KM} , which corresponds to setting $\Omega_d = B$ and $X_d = a$ across the bandwidth. On the right we show the square root of the oversmoothed image $\mathcal{I}^{\text{CINT}}$, obtained with values for Ω_d and X_d that are too small. There are no speckles in this image. They have been removed by smoothing but we have also lost a lot of resolution. The middle picture in Fig. 3 shows the square root of $\mathcal{I}^{\text{CINT}}$ for the optimal choice of Ω_d and X_d (i.e., κ_d) obtained by the following adaptive CINT algorithm.

Adaptive CINT algorithm: For given starting values of the parameters Ω_d and κ_d , calculate

$$\mathcal{J}(\mathbf{y}^S; \Omega_d, \kappa_d) = \frac{|\mathcal{I}^{\text{CINT}}(\mathbf{y}^S; \Omega_d, \kappa_d)|^{1/2}}{\sup_{\mathbf{y}^S \in \mathcal{D}} |\mathcal{I}^{\text{CINT}}(\mathbf{y}^S; \Omega_d, \kappa_d)|^{1/2}} \quad (10)$$

for \mathbf{y}^S sweeping a search domain \mathcal{D} . For a suitably chosen parameter $\alpha > 0$, estimate Ω_d and κ_d as the minimizers of

$$\mathcal{O}(\Omega_d, \kappa_d) = \|\mathcal{J}(\cdot; \Omega_d, \kappa_d)\|_{L^1(\mathcal{D})} + \alpha \|\nabla_{\mathbf{y}^S} \mathcal{J}(\cdot; \Omega_d, \kappa_d)\|_{L^1(\mathcal{D})}, \quad (11)$$

over all possible $\Omega_d \in [0, B]$ and $\kappa_d \geq 0$.

The first term in Eq. (11) favors images of small spatial support and the second term induces smoothing. This is essential for penalizing speckled images but the smoothing must be controlled in order to minimize the blurring. The

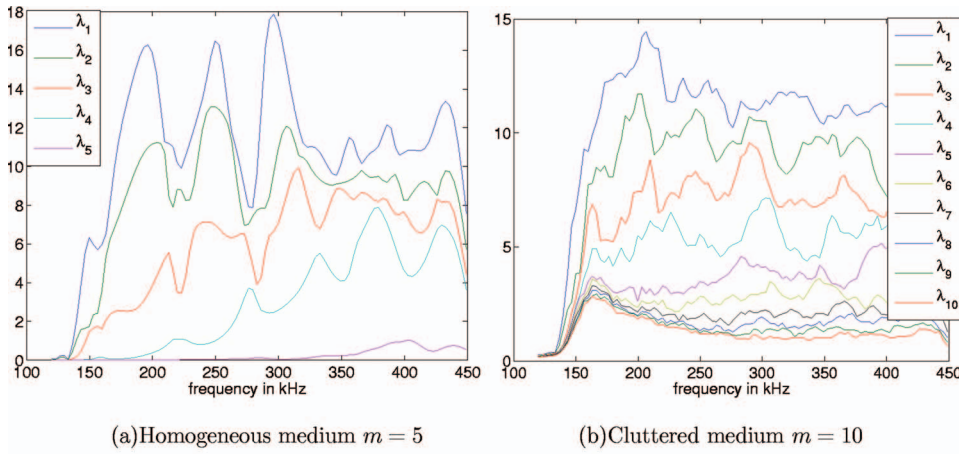


FIG. 4. The first m singular values of the response matrix as a function of frequency.

balance between the two terms in Eq. (11) is done with parameter $\alpha > 0$, which must be chosen in such a way that $\|\mathcal{J}(\cdot; \Omega_d, \kappa_d)\|_{L^1(\mathcal{D})}$ and $\alpha \|\nabla_{\mathbf{y}_s} \mathcal{J}(\cdot; \Omega_d, \kappa_d)\|_{L^1(\mathcal{D})}$ are of the same order. In principle, α can be changed during the estimation process. However, a good indication that the data respond well to the adaptive CINT algorithm is the relative stability of α as the iteration process for the determination of Ω_d and κ_d advances, until a plateau for the objective is reached. In all numerical simulations in this paper, and in previous ones,⁵ $\alpha = 1$ worked well.

Note that in adaptive CINT we work with the square root of $|\mathcal{I}^{\text{CINT}}|$. This is because we compare it with \mathcal{I}^{KM} . When no smoothing is done then $|\mathcal{I}^{\text{CINT}}|^{1/2}$ is equal to $|\mathcal{I}^{\text{KM}}|$.

In the numerical setup considered in this paper (cf. Fig. 2) the scatterers are close together and, because of the inevitable blurring that CINT introduces, they cannot be distinguished with adaptive CINT. This is clear from the middle picture in Fig. 3. In order to image the scatterers separately, we introduce in the next section an algorithm for optimal illumination and selective imaging. In this algorithm we use the singular value decomposition (SVD) of the array response matrix, frequency by frequency. When the scatterers are not so closely spaced, or more precisely, when the blurring introduced is smaller than the distance in between the scatterers, then the adaptive CINT method can effectively distinguish them. Results for such a case are reported in Ref. 5.

IV. OPTIMAL ILLUMINATION AND SELECTIVE IMAGING WITH CINT

We introduce a two-step algorithm for optimal illumination and subspace selection with adaptive CINT in order to image clusters of small scatterers in clutter. The first step is described in Sec. IV A and it uses the singular value decomposition of the array response matrix to image the scatterers in a selective manner, one at a time. The second step is described in Sec. IV B. It is designed to improve the image of each scatterer using an optimal illumination from the array.

A. Selective imaging with CINT

1. Singular value decomposition of the response matrix

The singular value decomposition of the $N \times N$ response matrix $\hat{\Pi}(\omega) = \{\hat{\Pi}(\mathbf{x}_r, \mathbf{x}_s, \omega)\}$ at any frequency ω in the bandwidth is given by

$$\hat{\Pi}(\omega) = \sum_{j=1}^N \sigma_j(\omega) \hat{\mathbf{u}}_j(\omega) \hat{\mathbf{v}}_j^*(\omega) \quad (12)$$

so that

$$\hat{\Pi}(\omega) \hat{\mathbf{v}}_j(\omega) = \sigma_j(\omega) \hat{\mathbf{u}}_j(\omega), \quad j = 1, \dots, N. \quad (13)$$

Here the star stands for complex conjugate and transpose. The singular values $\sigma_j(\omega) \geq 0$ are in decreasing order and $\hat{\mathbf{u}}_j(\omega), \hat{\mathbf{v}}_j(\omega)$ are the orthonormal left and right singular vectors, respectively.

Because the complex matrix $\hat{\Pi}(\omega)$ is symmetric, although not Hermitian, we can determine the left singular vectors as the complex conjugates of the right ones. However, this is true only when the correct phase has been assigned to these vectors. The computation of the SVD with any public software returns

$$\hat{\mathbf{u}}_j(\omega) = e^{i\varphi_j(\omega)} \overline{\hat{\mathbf{v}}_j(\omega)}, \quad j = 1, \dots, N \quad (14)$$

with an ambiguous phase that is difficult to unwrap in a consistent manner across the bandwidth. Nevertheless, the projection matrices

$$\mathcal{P}_j(\omega) = \hat{\mathbf{u}}_j(\omega) \hat{\mathbf{u}}_j^*(\omega) \quad (15)$$

onto the space spanned by the j th left singular vector have no phase ambiguities, and this is what we use in the algorithm described below.

2. Data filtering

Let us assume that the number N of array elements is larger than the number M of small scatterers in the cluster. Then there are $n^*(\omega) < N$ significant singular values of $\hat{\Pi}(\omega)$, with $n^*(\omega) \sim M$. See Fig. 4 for an illustration of this fact, for the numerical setup described in Sec. II A, where $M = 4$. We note in Fig. 4(a) that in the homogeneous medium we can set

$n^*(\omega)=4$. In the random medium Fig. 4(b) shows that we can set $n^*(\omega)=5$ or at most 6, because for all indices beyond this threshold the singular values level off as when there is additive noise across the bandwidth. Note also that this threshold stays basically the same across the bandwidth, so we can write

$$n^*(\omega) \approx n^*(\omega_0). \quad (16)$$

This is because we are dealing with small, point-like scatterers. In the case of extended scatterers $n^*(\omega)$ does vary across the bandwidth, as was shown before,²¹ and this has to be taken into account when doing selective imaging of such reflectors.

Using the threshold $n^*(\omega_0)$, we now define a set of filtering operators that we wish to apply to the data.

Definition 1 Let Δ be the set of non-negative subspace weights

$$\Delta = \{d_j(\omega) \geq 0 \text{ for } j = 1, \dots, n^*(\omega_0), \\ |\omega - \omega_0| \leq \frac{B}{2} \text{ and}$$

$$\sum_{j=1}^{n^*(\omega_0)} \int_{|\omega - \omega_0| \leq \frac{B}{2}} d_j(\omega) d\omega = 1\}.$$

We define the filtering operators $D(\omega, \cdot): \Delta \rightarrow \mathbb{C}^{N \times N}$, which take coefficients $d \in \Delta$ and return a linear combination of the projection matrices (15)

$$D(\omega, d) = \sum_{j=1}^{n^*(\omega_0)} d_j(\omega) \mathcal{P}_j(\omega) \quad (17)$$

for each frequency ω in the bandwidth.

When we apply these filters to the response matrix, we get

$$D(\omega, d) \hat{\Pi}(\omega) = \sum_{j=1}^{n^*(\omega_0)} d_j(\omega) \mathcal{P}_j(\omega) \hat{\Pi}(\omega) \\ = \sum_{j=1}^{n^*(\omega_0)} d_j(\omega) \sigma_j(\omega) \hat{\mathbf{u}}_j(\omega) \hat{\mathbf{u}}_j^T(\omega), \quad (18)$$

where we removed the null space of $\hat{\Pi}(\omega)$. Note that here the singular vectors $\hat{\mathbf{u}}_j(\omega)$ have the correct phase (i.e., $\varphi_j(\omega)=0$ in (14)). This correct phase is obtained by computing the projection $\mathcal{P}_j(\omega) \hat{\Pi}(\omega)$ of $\hat{\Pi}(\omega)$, which does not suffer from the phase ambiguity.

Now if we distinguish the subsets Δ_p of Δ , for $p=1, \dots, n^*(\omega_0)$,

$$\Delta_p = \{d \in \Delta; d_p(\omega) = \frac{1}{B} \text{ for } |\omega - \omega_0| \leq \frac{B}{2} \text{ and}$$

$$d_j(\omega) = 0 \text{ for } j \neq p, j = 1, \dots, n^*(\omega_0)\},$$

we note that $D(\omega, d)$ for $d \in \Delta_p$ filters out the contribution of all singular vectors, except the p th one, uniformly across the bandwidth. This is what is done at the p th step of the basic DORT algorithm. The travel-time migration of the filtered data $D(\omega, d) \hat{\Pi}(\omega)$ gives the imaging function

$$\mathcal{I}(\mathbf{y}^S, p) = \int_{|\omega - \omega_0| \leq \frac{B}{2}} d\omega \sigma_p(\omega) \sum_{s=1}^N \sum_{r=1}^N \hat{u}_p(\mathbf{x}_s, \omega) \\ \times \hat{u}_p(\mathbf{x}_r, \omega) e^{-i\omega[\tau(\mathbf{x}_s, \mathbf{y}^S) + \tau(\mathbf{x}_r, \mathbf{y}^S)]} \\ = \int_{|\omega - \omega_0| \leq \frac{B}{2}} d\omega \sigma_p(\omega) \left[\sum_{s=1}^N \hat{u}_p(\mathbf{x}_s, \omega) e^{-i\omega\tau(\mathbf{x}_s, \mathbf{y}^S)} \right]^2. \quad (19)$$

We call this the DORT image because it uses one singular vector at a time. We recall that in the original DORT method,¹² the singular vectors are computed at each frequency and then a time-domain signal is formed from them and is backpropagated. To construct this signal in time, the phase of the singular vectors is used to compute the time delay that it needs at the array in order to focus. We can also describe DORT as an imaging method in which each singular vector over the bandwidth is backpropagated and the images are then summed over frequencies. The backpropagation is done in a homogeneous medium. In such a medium, DORT and iterative time reversal²² are essentially equivalent and can be used either for detection or for imaging of small targets. In iterative time reversal the backpropagation step is done by *physically* transmitting the time reversed signal from the transducer array, repeatedly until it focuses on the strongest target. In a random medium iterative time reversal is an excellent *detection* method because it is not sensitive to the random inhomogeneities. This is another manifestation of statistical stability in time reversal.^{6,10,11} The issue of iterative time reversal versus DORT as a *detection* method with noisy signals is also discussed in Chapter 10 of Ref. 23. Imaging in clutter is addressed^{24,25} using the singular value decomposition and time correlation techniques. The algorithms based on adaptive CINT that are presented in this paper are, however, more flexible and effective.

Note that when we use just one source in Eq. (19) we get

$$\tilde{\mathcal{I}}(\mathbf{y}^S, p, \mathbf{x}_s) = \int_{|\omega - \omega_0| \leq \frac{B}{2}} d\omega \sigma_p(\omega) [\hat{u}_p(\mathbf{x}_s, \omega) e^{-i\omega\tau(\mathbf{x}_s, \mathbf{y}^S)}] \\ \times \left[\sum_{r=1}^N \hat{u}_p(\mathbf{x}_r, \omega) e^{-i\omega\tau(\mathbf{x}_r, \mathbf{y}^S)} \right]. \quad (20)$$

When $p=1$ and $\sigma_p(\omega)$ is constant in the bandwidth, this can be interpreted as the image produced by travel-time migration of the traces obtained with the illumination produced by iterative time reversal^{7,8} in echo mode. Indeed, suppose that we illuminate the scatterers from the source at $\mathbf{x}_s \in \mathcal{A}$ and we record the echoes

$$\hat{P}(\mathbf{x}_r, \mathbf{x}_s, \omega) = \hat{\Pi}(\mathbf{x}_r, \mathbf{x}_s, \omega), \quad r = 1, \dots, N.$$

This is the s column of the response matrix written in short $\hat{\Pi}(\cdot, \mathbf{x}_s, \omega)$. Now, if we time reverse these signals and we reemit them in the medium, we get back the new vector

$$\overline{\hat{\Pi}(\omega) \hat{\Pi}(\cdot, \mathbf{x}_s, \omega)}.$$

Iterating this way, we converge to the leading singular vector as is well known.^{7,8} However, this singular vector is com-

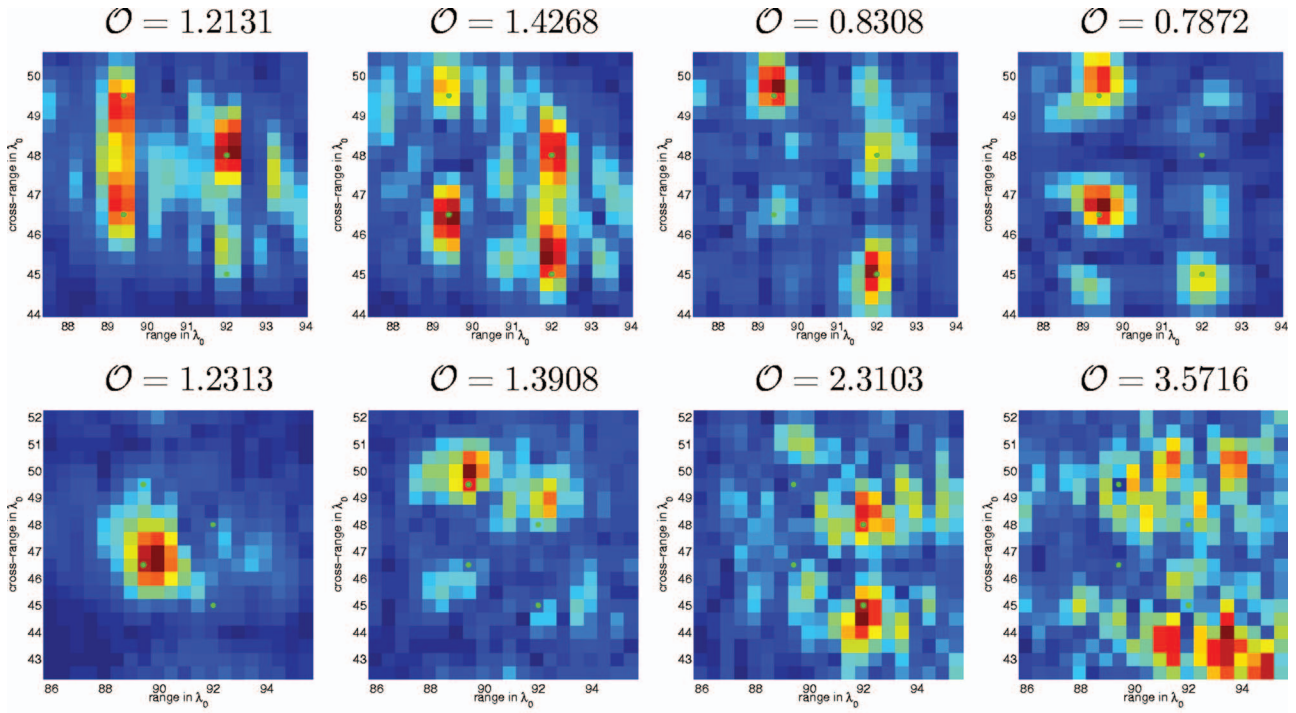


FIG. 5. DORT images using the first, second, third and fourth singular vector across the bandwidth. Top row: homogeneous medium. Bottom row: random medium. Above each plot we give the value of the L^2 norm of the image normalized by its maximum, which we denote by \mathcal{O} .

puted with the phase of $\hat{u}(\mathbf{x}_s, \omega)$ that is inherited from the original illumination from the source at \mathbf{x}_s . This phase is compensated in Eq. (20) with the travel time $\tau(\mathbf{x}_s, \mathbf{y}^S)$ to the search point \mathbf{y}^S where the image peaks. The DORT images (20) peak on one scatterer at a time, if these are well separated, in the sense that the p th singular value is associated with the same scatterer uniformly in the bandwidth. However, in clusters of nearby scatterers DORT does not give selective images because the one-to-one correspondence between the singular values and the scatterers does not hold across the bandwidth. This is seen clearly in the plots of the top row of Fig. 5. Moreover, DORT does not give stable images in clutter as can be seen in plots of the bottom row of Fig. 5.

3. Selective imaging algorithm

To image the scatterers one by one, we formulate an iterative algorithm with two parts that we now describe.

Part 1: We estimate first the clutter-dependent decoherence parameters Ω_d and K_d using adaptive CINT. This estimation is done as described in Sec. III with the CINT imaging function (7) that uses the unfiltered array response matrix $\hat{\Pi}(\omega)$. The resulting image is shown in the middle in Fig. 3. The numerical setup is described in Sec. II A.

We note that we may also be able to estimate Ω_d and κ_d using CINT with array data coming from a known small scatterer, a pilot, whose range is comparable to that of the cluster that we wish to image selectively.

Part 2: With Ω_d and κ_d fixed, as found in Part 1, define the CINT image using the filtered data $D(\omega, d)\hat{\Pi}(\omega)$

$$\begin{aligned} \mathcal{I}^{\text{CINT}}(\mathbf{y}^S; \Omega_d, \kappa_d; d) &= \int_{|\omega - \omega_0| \leq \frac{B}{2}} d\omega \int_{|\omega' - \omega_0| \leq \frac{B}{2}, |\omega - \omega'| \leq \Omega_d} d\omega' \\ &\times \sum_{r, r' \in \chi\left(\frac{\omega + \omega'}{2}, \kappa_d\right)} \sum_{s, s' \in \chi\left(\frac{\omega + \omega'}{2}, \kappa_d\right)} \\ &\times \hat{Q}_d(\mathbf{x}_r, \mathbf{x}_s, \omega; \mathbf{y}^S) \overline{\hat{Q}_d(\mathbf{x}_{r'}, \mathbf{x}_{s'}, \omega'; \mathbf{y}^S)} \end{aligned} \quad (21)$$

for arbitrary $d \in \Delta$. Here $\hat{Q}_d(\mathbf{x}_r, \mathbf{x}_s, \omega; \mathbf{y}^S)$ is the filtered data migrated to \mathbf{y}^S

$$\hat{Q}_d(\mathbf{x}_r, \mathbf{x}_s, \omega; \mathbf{y}^S) = [D(\omega, d)\hat{\Pi}(\omega)]_{r,s} e^{-i\omega[\tau(\mathbf{x}_s, \mathbf{y}^S) + \tau(\mathbf{x}_r, \mathbf{y}^S)]} \quad (22)$$

and the set $\chi(\omega, \kappa_d)$ is defined in Eq. (9) in terms of κ_d . Now we begin the iteration:

Step 1: Find the optimal filter $D(\omega, d)$, for $|\omega - \omega_0| \leq B/2$, by minimizing the objective function

$$\mathcal{O}(d) = \|\mathcal{J}(\cdot; d)\|_{L^1(\mathcal{D})},$$

$$\mathcal{J}(\mathbf{y}^S; d) = \frac{|\mathcal{I}^{\text{CINT}}(\mathbf{y}^S; \Omega_d, \kappa_d; d)|^{1/2}}{\sup_{\mathbf{y}^S \in \mathcal{D}} |\mathcal{I}^{\text{CINT}}(\mathbf{y}^S; \Omega_d, \kappa_d; d)|^{1/2}} \quad (23)$$

over all subspace weights $d \in \Delta$. This optimization seeks images of small spatial support. There is no need to penalize the gradient here because all the statistical smoothing has been done by the thresholding with Ω_d and X_d (i.e., κ_d). The optimization involves $n^*(\omega_0)N_\omega$ variables $d_j(\omega_m)$, for $j = 1, \dots, n^*(\omega_0)$ and $m = 1, \dots, N_\omega$, where N_ω is the number of frequency subbands used to parametrize the bandwidth. De-

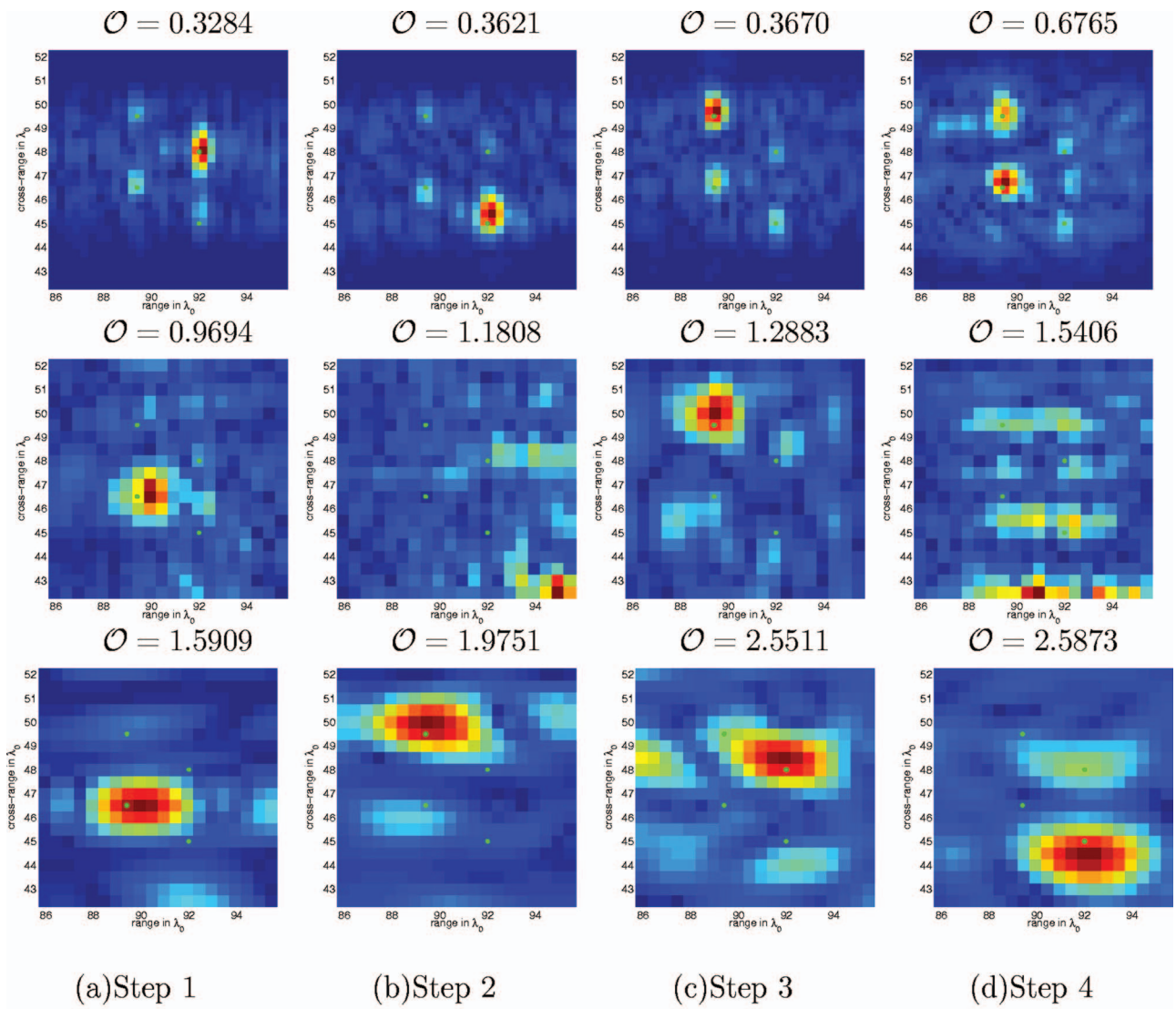


FIG. 6. Images with optimally selected $d_j(\omega)$ and uniform weights ω . Top row: homogeneous medium, middle row: Cluttered medium using $X_d=a$ and $\Omega_d=B$. Bottom row: Cluttered medium using X_d and Ω_d obtained by the adaptive algorithm. The scatterers are indicated with green dots in the image. Above each plot we give the value of the objective function \mathcal{O} .

note the resulting optimal filter by $D(\omega, d^{(1)})$.

Remark 1 We show in the bottom left picture of Fig. 6 the image obtained with the optimal filter $D(\omega, d^{(1)})$, in the setup described in Sec. II A. The image is focused on a single scatterer and the optimally filtered data $D(\omega, d^{(1)})\hat{\Pi}(\omega)$ corresponds to the echoes from this scatterer alone. The optimal subspace weights $d^{(1)}$ are plotted in the bottom left picture of Fig. 7.

Step 2: In order to image the next scatterer, we use filters that mask the scatterer found in Step 1, as follows. Let us denote by $J(\omega, 1)$ the index of the largest subspace weight

$$d_{J(\omega, 1)}^{(1)}(\omega) = \max_{j=1, \dots, n^*(\omega_0)} d_j^{(1)}(\omega). \quad (24)$$

We define the map $\delta^{(1)}: [\omega_0 - \frac{B}{2}, \omega_0 + \frac{B}{2}] \rightarrow \mathbb{N}$ which takes a frequency ω in the bandwidth and returns

$$\delta^{(1)}(\omega) = \begin{cases} J(\omega, 1) & \text{if } d_{J(\omega, 1)}^{(1)}(\omega) > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (25)$$

Now we seek the optimal filter $D(\omega, d^{(2)})$ for $|\omega - \omega_0| \leq B/2$ by minimizing the objective function Eq. (23) over all the subspace weights in

$$\Delta^{(1)} = \{d_j(\omega) \in \Delta, d_j(\omega) = 0 \text{ if } j = \delta^{(1)}(\omega), \text{ for } |\omega - \omega_0| \leq \frac{B}{2}, j = 1, \dots, n^*(\omega_0)\}.$$

Remark 2 We show in Fig. 7 the optimal subspace weights $d^{(2)}$ and in Fig. 6 the images obtained with $D(\omega, d^{(2)})$. The objective function is larger than that at Step 1 because of the additional constraints in $\Delta^{(1)}$. The images are focused on a single scatterer, which is different from the one in Step 1. That scatterer has been masked by restricting the search to the set $\Delta^{(1)} \subset \Delta$.

⋮

Step p: At the pth iteration we mask the scatterers found at Steps 1, ..., p-1 and we seek the optimal filter $D(\omega, d^{(p)})$

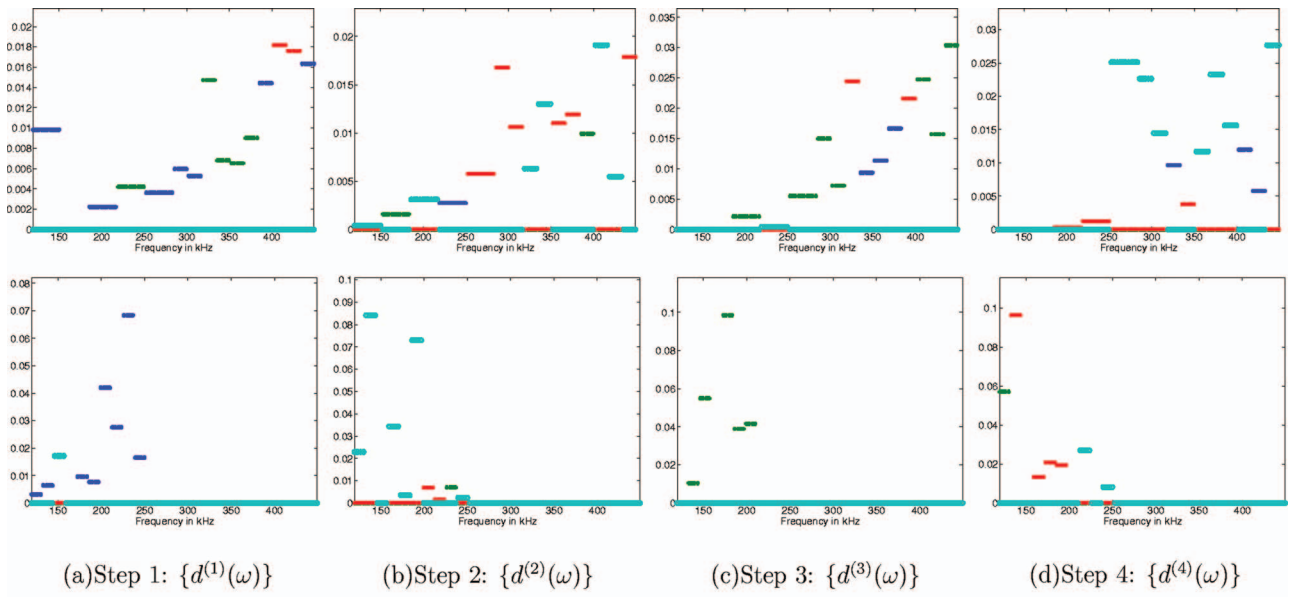


FIG. 7. Optimal $d_j^{(p)}(\omega)$, $j=1, \dots, 4$, $p=1, \dots, 4$. Top row: homogeneous medium. Bottom row: random medium. Only the largest coefficients in each frequency band are shown, and $d_1^{(p)}(\omega)$ are in blue, $d_2^{(p)}(\omega)$ in green, $d_3^{(p)}(\omega)$ in red and $d_4^{(p)}(\omega)$ in cyan (light blue). The p th column is for the p iteration, with $p=1, \dots, 4$.

that minimizes the objective function (23) over the subspace weights in

$$\Delta^{(p)} = \{d_j(\omega) \in \Delta^{(p-1)}, d_j(\omega) = 0 \text{ if } j = \delta^{(p-1)}(\omega),$$

$$\text{for } |\omega - \omega_0| \leq \frac{B}{2}, j = 1, \dots, n^*(\omega_0)\}.$$

⋮

The iteration terminates at the $n^*(\omega_0)$ step, where we have used all the degrees of freedom in the array response matrix.

B. Optimal illumination algorithm

The images obtained with the selective imaging algorithm described in Sec. IV A can be improved further by assigning optimal weights to each source in the array, as we now describe. Let

$$\mathcal{W} = \left\{ w_s \geq 0 \text{ for } s = 1, \dots, N; \sum_{s=1}^N w_s = 1 \right\}$$

be the set of weights that we can assign to the sources in the array and define for any $d \in \Delta$ and any weights in \mathcal{W} the CINT imaging function

$$\mathcal{I}^{\text{CINT}}(\mathbf{y}^S; \Omega_d, \kappa_d; d, w) = \int_{|\omega - \omega_0| \leq \frac{B}{2}} d\omega \int_{\substack{|\omega' - \omega_0| \leq \frac{B}{2} \\ |\omega - \omega'| \leq \Omega_d}} d\omega'$$

$$\times \sum_{r, r' \in \mathcal{X}} \sum_{\left(\frac{\omega + \omega'}{2}, \kappa_d\right)} \sum_{s, s' \in \mathcal{X}} \sum_{\left(\frac{\omega + \omega'}{2}, \kappa_d\right)} w_s$$

$$\times \hat{Q}_d(\mathbf{x}_r, \mathbf{x}_s, \omega; \mathbf{y}^S) w_{s'} \hat{Q}_d(\mathbf{x}_{r'}, \mathbf{x}_{s'}, \omega'; \mathbf{y}^S).$$

(26)

Here $\hat{Q}_d(\mathbf{x}_r, \mathbf{x}_s, \omega; \mathbf{y}^S)$ is the filtered data migrated to \mathbf{y}^S , and given by Eq. (12)

For a given data filter $D(\omega, d)$ with $|\omega - \omega_0| \leq B/2$, the optimal illumination w is the minimizer of the objective function

$$\mathcal{O}(w) = \|\mathcal{J}(\cdot; w)\|_{L^1(\mathcal{D})},$$

$$\mathcal{J}(\mathbf{y}^S; w) = \frac{|\mathcal{I}^{\text{CINT}}(\mathbf{y}^S; \Omega_d, \kappa_d; d, w)|^{1/2}}{\sup_{\mathbf{y}^S \in \mathcal{D}} |\mathcal{I}^{\text{CINT}}(\mathbf{y}^S; \Omega_d, \kappa_d; d, w)|^{1/2}} \quad (27)$$

over the set \mathcal{W} . This can be used in conjunction with the optimal selective illumination algorithm as follows.

The optimal illumination algorithm: For $p = 1, \dots, n^*(\omega_0)$ determine the optimal filters $D(\omega, d^{(p)})$ for $|\omega - \omega_0| \leq B/2$ as described in Sec. IV A. Then compute for any given p the optimal illumination $w^{(p)}$, which minimizes over the set \mathcal{W} the objective function (27) computed with $d = d^{(p)}$.

Remark 3 We illustrate this algorithm with numerical simulations in the setup described in Sec. II A. The source weights are shown in Fig. 9 and the images are shown in Fig. 8.

To explain how the algorithm works, let us consider the source at \mathbf{x}_s . Its contribution to the imaging function (26) is in the term

$$w_s [D(\omega, d) \hat{\Pi}(\omega)]_{r,s} \quad \text{for } r = 1, \dots, N_s. \quad (28)$$

Suppose for simplicity that we have a single scatterer and $M = n^*(\omega_0) = 1$. Then Eq. (28) $d = d^{(1)}$ is just the s th column of the response matrix, corresponding to the illumination of the scatterer from the source at \mathbf{x}_s . The image with the source at \mathbf{x}_s peaks naturally at the location of the scatterer. However, the focusing can be sharpened by using all the sources with optimal weights, as we have shown in Ref. 21.

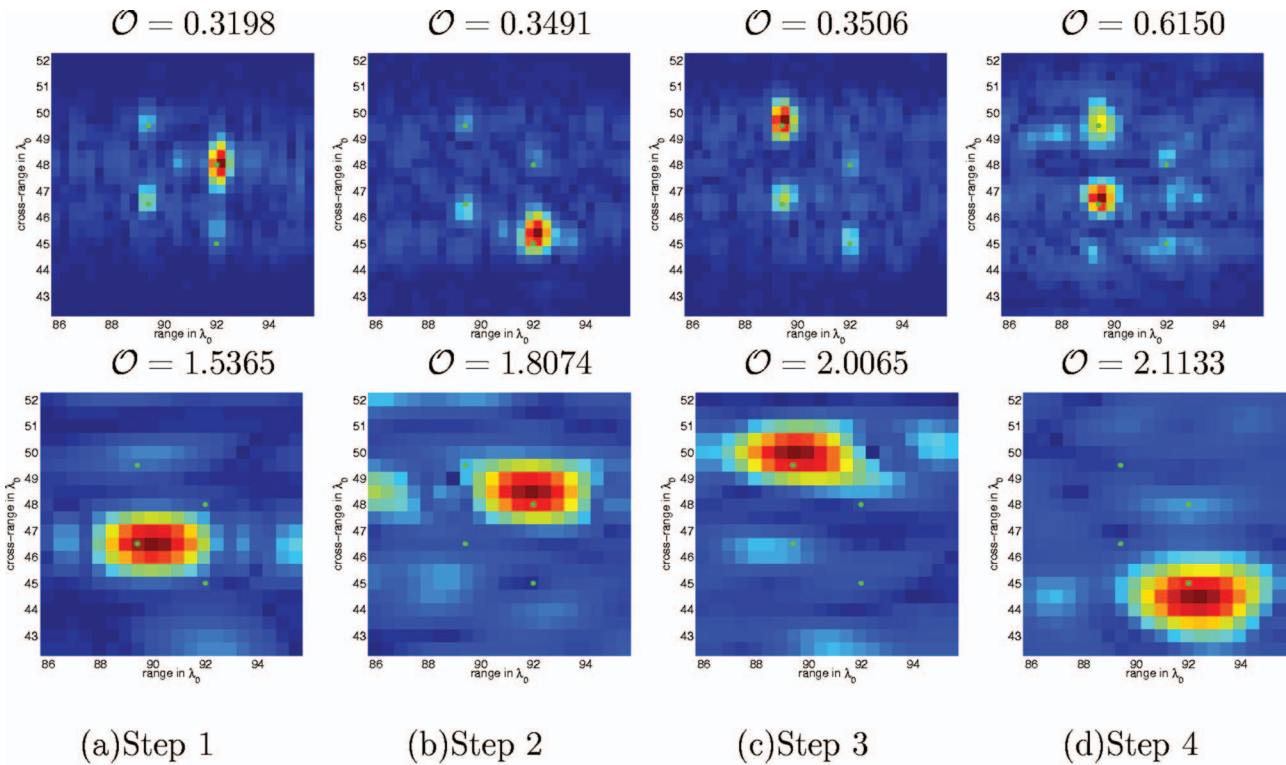


FIG. 8. Images with optimally selected $d_f(\omega)$ and weights ω . Top row: Homogeneous medium, bottom row: Cluttered medium using X_d and Ω_d obtained by the adaptive algorithm. The scatterers are indicated with green dots in the image. Above each plot we give the value of the objective function \mathcal{O} .

When we have a cluster of scatterers the algorithm works in a similar way. The selective imaging algorithm gives at step p an image that is focused on the p th scatterer. The filtered data $D(\omega, d^{(p)})\hat{\Pi}(\omega)$ correspond approximately to the echoes from this scatterer alone, and the image is sharpened by using the optimal illumination $w^{(p)}$ from the array.

V. NUMERICAL IMPLEMENTATION AND RESULTS

In the implementation of the algorithms, the search point \mathbf{y}^S is in a square domain of size $10\lambda_0 \times 10\lambda_0$, centered at the

center of mass of the scatterers. This domain is discretized using a uniform grid of 31×31 points and thus the discretization step is $\lambda_0/3$. It is in this domain that we present all the images in the paper (see Figs. 3, 5, 6, and 8).

Details regarding the implementation of the adaptive CINT algorithm can be found in Ref. 5. In order to save computational time when implementing the selective imaging algorithm, we divide the bandwidth B into 15 subbands. In the lower frequency range, for $\omega \in [150, 300]$ kHz, we take 5 subbands of size 30 kHz while in the higher frequency range, for $\omega \in [300, 450]$ kHz, we take a finer discretization

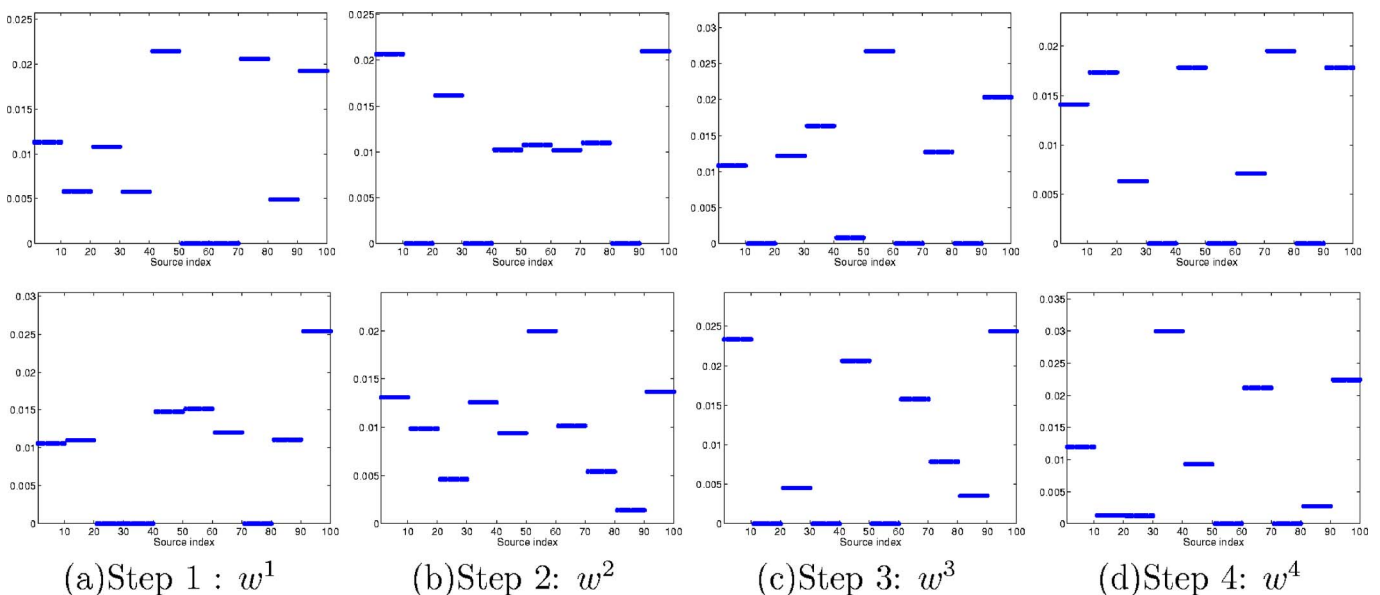


FIG. 9. (Color online) The optimal weights. Top row: homogeneous medium. Bottom row: cluttered medium.

using ten subbands of size 15 kHz. In each subband the subspace weights $d_j(\omega)$ are constant. This means that for the selective imaging algorithm we have 60 unknowns because we use $n^*(\omega_0)=4$. In the cluttered medium, we tried first this discretization of the frequency and noted that the coefficients $d_j(\omega)$ are mostly zero for frequencies ω outside of [150,250] kHz. We therefore changed the discretization. In the cluttered medium we use nonzero $d_j(\omega)$ only in the lower part of the bandwidth, that is, for $\omega \in [150,250]$ kHz. The results shown in the bottom row of Fig. 7 are obtained by dividing this part of the bandwidth into ten subbands of size 10 kHz and taking the subspace weights $d_j(\omega)$ to be constant in each subband. The optimal coefficients $d_j(\omega)$ for homogeneous and random media are shown in Fig. 7. The convention for the colors in Fig. 7 is the following: we use the blue color for $d_1(\omega)$, green for $d_2(\omega)$, red for $d_3(\omega)$ and cyan for $d_4(\omega)$.

In the implementation of the optimal illumination algorithm, we group the array weights into blocks of ten elements and take the weights w to be constant in each block. The number of variables in this optimization problem is ten. The optimal weights for homogeneous and random media are shown in Fig. 9.

Because the scatterers are close to each other in our numerical simulations, we cannot image them separately with the DORT method as can be seen clearly in the top row of Fig. 5, which is in a homogeneous medium. There are more elaborate implementations of DORT that use continuation methods for singular values along frequencies. These may give better results in homogeneous media. However, DORT in its various implementations cannot be used in cluttered media, because it gives noisy and unstable images. In the images in the bottom row of Fig. 5, we note that as we go from left to right the results deteriorate. The last image in particular is just noise.

We see a big improvement in Fig. 6 bottom row, where the images are obtained with the selective subspace algorithm. We note that we image the scatterers one by one, both in the homogeneous and in the cluttered medium. The discretization of the bandwidth in the selective imaging process significantly affects the results. On one hand, coarser parametrizations lead to fewer optimization variables and lower computational cost. On the other hand, the image quality suffers from coarse parametrizations. This can be seen in the images in the top row of figure, which is for a homogeneous medium. They are better than the DORT images shown in the top row of Fig. 5, but we still see some ghosts. However, when we refine the bandwidth parametrization, the results improve.

By comparing the images in the middle and bottom rows of Fig. 6, we see that it is crucial to use a statistically stable method in clutter, such as CINT. When we do not have statistically stable results, the optimal subspace selection does not make sense. This is seen in the middle row of Fig. 6, where travel-time migration is used instead of CINT. The bottom row in Fig. 6 shows how selective imaging works in clutter. Here the resolution is naturally worse than that in the

homogeneous case. This is expected because, as we explained Sec. III, with CINT we give up some resolution to gain stability in clutter.

The results are further improved with the optimal illumination algorithm, as shown in the images of Fig. 8. For both homogeneous and random media, the use of the optimal weights improves the quality of the images by reducing side-lobes. The improvement is more visible in cluttered media.

VI. SUMMARY AND CONCLUSIONS

We introduce a three-part algorithm for imaging small nearby scatterers in a randomly inhomogeneous medium. We assess its performance with numerical simulations in the context of ultrasonic array imaging of concrete structures. Travel-time migration is not effective and gives images that have speckles and are unstable because of significant multiple scattering from the inhomogeneities. Adaptive coherent interferometry (CINT) is effective in stabilizing the image at the expense of some blurring, which makes nearby scatterers difficult to identify. The algorithm presented in this paper uses adaptive CINT combined with (1) optimal subspace selection using the singular value decomposition (SVD) of the array response matrix, and (2) with optimal illumination. This algorithm is much more demanding computationally than travel-time migration, but it is quite effective for imaging in clutter.

ACKNOWLEDGMENTS

The work of L.B. was partially supported by the Office of Naval Research, under Grants No. N00014-05-1-0699 and by the National Science Foundation, Grant Nos. DMS-0604008, DMS-0305056, DMS-0354658. The work of G.P. was supported by Grants Nos. ONR N00014-02-1-0088, 02-SC-ARO-1067-MOD 1 and NSF DMS-0354674-001. The work of C.T. was partially supported by the Office of Naval Research, under Grant No. N00014-02-1-0088 and by 02-SC-ARO-1067-MOD 1.

¹L. Borcea, G. Papanicolaou, and C. Tsogka, "Interferometric array imaging in clutter," *Inverse Probl.* **21**, 1419–1460 (2005).

²L. Borcea, G. Papanicolaou, and C. Tsogka, "Coherent interferometric imaging," *Geophysics* **71**, S1165–S1175 (2006).

³J. F. Claerbout and S. M. Doherty, "Downward continuation of moveout-corrected seismograms," *Geophysics* **37**, 741–768 (1972).

⁴N. Bleistein, J. Cohen, and J. W. Stockwell, *Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion* (Springer, New York, 2001).

⁵L. Borcea, G. Papanicolaou, and C. Tsogka, "Adaptive interferometric imaging in clutter and optimal illumination," *Inverse Probl.* **22**, 1405–1436 (2006).

⁶L. Borcea, G. Papanicolaou, and C. Tsogka, "Asymptotics for the space-time Wigner transform with applications to imaging," in *Stochastic Differential Equations: Theory and Applications. Volume in Honor of Professor Boris L. Rozovskii*, edited by P. H. Baxendale and S. V. Lototsky, Vol. 2 of *Interdisciplinary Mathematical Sciences* (World Scientific, Singapore, 2007), pp. 91–112.

⁷C. Prada, J. L. Thomas, and M. Fink, "The iterative time-reversal process: Analysis of the convergence," *J. Acoust. Soc. Am.* **97**, 62–71 (1995).

⁸G. Montaldo, M. Tanter, and M. Fink, "Revisiting iterative time reversal processing: Application to detection of multiple targets," *J. Acoust. Soc. Am.* **115**, 776–784 (2004).

⁹M. Cheney, D. Isaacson, and M. Lassas, "Optimal acoustic measurements," *SIAM J. Appl. Math.* **61**, 1628–1647 (2001).

- ¹⁰P. Blomgren, G. Papanicolaou, and H. Zhao, "Super-resolution in time-reversal acoustics," *J. Acoust. Soc. Am.* **111**, 238–248 (2002).
- ¹¹G. Papanicolaou, L. Ryzhik, and K. Solna, "Self-averaging from lateral diversity in the Itô-Schrödinger equation," *SIAM J. Multiscale Modeling and Simulation* **6**, 468–492 (2007).
- ¹²C. Prada and M. Fink, "Eigenmodes of the time reversal operator: A solution to selective focusing in multiple-target media," *Wave Motion* **20**, 151–163 (1994).
- ¹³K. J. Langenberg, R. Marklein, K. Mayer, T. Krylov, P. Ampha, M. Krause, and D. Streicher, "Wavefield inversion in nondestructive testing," in *Electromagnetics in a Complex World—Challenges and Perspectives*, edited by I. M. Pinto, V. Galdi, and L. B. Felsen (Springer, New York, 2003), pp. 277–285.
- ¹⁴R. Marklein, K. J. Langenberg, K. Mayer, J. Miao, A. Shilivinski, A. Zimmer, W. Müller, V. Schmitz, C. Kohl, and U. Mletzko, "Recent applications and advances of numerical modeling and wavefield inversion in nondestructive testing," *Adv. Radio Sci.* **3**, 167–174 (2005).
- ¹⁵E. Bécache, P. Joly, and C. Tsogka, "An analysis of new mixed finite elements for the approximation of wave propagation problems," *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **37**, 1053–1084 (2000).
- ¹⁶L. Ryzhik, G. Papanicolaou, and J. B. Keller, "Transport equations for elastic and other waves in random media," *Wave Motion* **24**, 327–370 (1996).
- ¹⁷M. C. W. van Rossum and T. M. Nieuwenhuizen, "Multiple scattering of classical waves: Microscopy, mesoscopy, and diffusion," *Rev. Mod. Phys.* **71**, 313–371 (1999).
- ¹⁸A. Derode, A. Tourin, and M. Fink, "Limits of time-reversal focusing through multiple scattering: Long range correlation," *J. Acoust. Soc. Am.* **107**, 2987–2998 (2000).
- ¹⁹M. Born and E. Wolf, *Principles of Optics* (Academic, New York, 1970).
- ²⁰G. Christakos, *Random Field Models in Earth Sciences* (Dover, New York, 2005).
- ²¹L. Borcea, G. Papanicolaou, and C. Tsogka, "Optimal waveform design for array imaging," *Inverse Probl.* **23**, 1973–2020 (2007).
- ²²E. Kerbrat, C. Prada, D. Cassereau, and M. Fink, "Imaging in the presence of grain noise using the decomposition of the time reversal operator," *J. Acoust. Soc. Am.* **113**, 1230–1240 (2003).
- ²³J.-P. Fouque, J. Garnier, G. Papanicolaou, and K. Solna, *Wave Propagation and Time Reversal in Randomly Layered Media* (Springer, New York, 2007).
- ²⁴L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman, "Imaging and time reversal in random media," *Inverse Probl.* **18**, 1247–1279 (2002).
- ²⁵L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman, "Statistically stable ultrasonic imaging in random media," *J. Acoust. Soc. Am.* **112**, 1509–1522 (2002).

A computer model of medial efferent suppression in the mammalian auditory system

Robert T. Ferry^{a)} and Ray Meddis

Department of Psychology, University of Essex, Colchester, CO4 3SQ, United Kingdom

(Received 26 July 2007; revised 24 September 2007; accepted 26 September 2007)

Stimulation of the olivocochlear bundle reduces basilar membrane displacement, driven auditory nerve activity, and compound action potential (CAP) response to acoustic stimulation. These effects were simulated using a computer model of the auditory periphery. The model simulates the medial efferent activity by attenuating the basilar membrane response. The model was evaluated against three animal studies reporting measurements at three levels of the auditory system; basilar membrane, single auditory nerve fibers and whole auditory nerve CAP. The CAP data included conditions where tones were masked by noise and “unmasked” by stimulation of the olivocochlear bundle. The model was able to simulate the data both qualitatively and quantitatively. As a consequence, it may be a suitable platform for studying the contribution of the efferent system to auditory processing of more complex auditory sounds in distracting backgrounds.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799914]

PACS number(s): 43.64.Bt [BLM]

Pages: 3519–3526

I. INTRODUCTION

The olivocochlear bundle is an important component of the mammalian auditory efferent system (Rasmussen, 1946). It originates in the superior olivary complex (SOC) and terminates in the cochlea. The purpose of this reflex appears to be the regulation of activity in the auditory nerve (AN) through modification of outer hair cell (OHC) electrical and mechanical properties and, more directly, through postsynaptic contacts on the AN itself. The efferent system is normally characterized as two separate olivocochlear systems, medial and lateral (MOC and LOC) according to their origin within the SOC. Although relatively little is known about the LOC system at present, the MOC system is much better understood and is known to act on the OHC so as to reduce the response of the basilar membrane (BM) to acoustic stimulation. The purpose of which may be to shift the dynamic range of hearing, reduce the effects of masking (unmasking), protect from sustained acoustic trauma and aid in selective attention (Rajan, 2000; Delgutte, 1990; Maison *et al.*, 2001).

Recent research (reviewed in Guinan, 2006) has considerably expanded our understanding of the physiology of the efferent system and has created favorable conditions for building models of how it functions. In this study we shall adapt an existing computer model of BM activity to simulate the effect of MOC efferent activity. Any change to the model BM response will necessarily have consequences for activity in the model AN.

The simulations reported here will replicate three empirical studies from different laboratories (BM: Russell and Murugasu, 1997; AN: Guinan and Stankovic, 1996; compound action potential (CAP): Dolan and Nuttall, 1988). The long term aim of the project is to use the modified model to

study the functional role of the efferent system in auditory processing of complex stimuli such as music and speech (clean and with noise). The primary aim of this study, however, is to demonstrate that the new model, implementing the effect of MOC stimulation, is simply able to fit the physiological data demonstrated in each of these studies.

All three studies show that direct electrical stimulation of MOC fibers produces a reduction in the effect of acoustic stimulation. Russell and Murugasu (1997) showed that the displacement of the BM was reduced, whereas Guinan and Stankovic (1996) showed that AN activity was reduced and rate/level functions shifted toward higher intensities. Dolan and Nuttall (1988) showed a reduction in the CAP magnitude in response to efferent stimulation. They also showed a depression of the CAP response to a tone in a noise background. Intriguingly, they found that MOC stimulation reversed the suppressive effect of the noise leading to an increase in the CAP response. This increase confirms the same findings by Kawase and Liberman (1993) and is an example of MOC activity acting on the BM to enhance the response to stimuli presented in an adverse background. This *antimasking* effect is a particular challenge to the computer model.

In this study, an existing computer model of the response of the BM (Meddis *et al.*, 2001) was used after applying a simple modification to reflect the action of MOC activity. An attenuator was applied to the response of the BM so that the attenuation was proportional to the amount of MOC activity. This small change to the model structure was all that was necessary to simulate the physiological effects described. The BM model was then deployed as a component in a general model of the auditory periphery (Meddis, 2006) that simulated the response up to the level of the AN. In a study of speech perception, Ghitza *et al.* (2006) have recently suggested a similar modification to Goldstein's (1990) multiple band pass nonlinear model to achieve the same effect.

^{a)}Current address: Room 2.702, Department of Psychology, University of Essex, Colchester, CO4 3SQ, United Kingdom. Electronic mail: rferry@essex.ac.uk

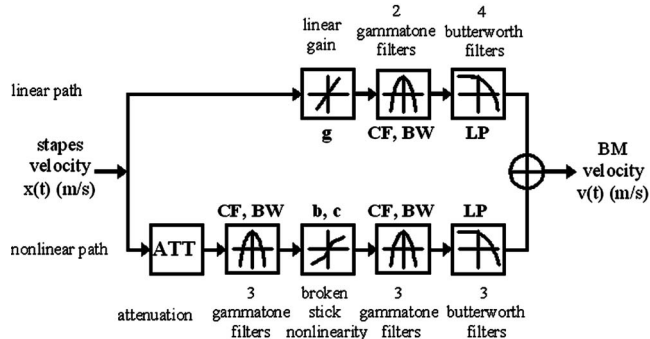


FIG. 1. Schematic of the modified DRNL filterbank. The modification involves the addition of an attenuation module (ATT) in the nonlinear path. Using ATT it is possible to use changes to a single parameter to simulate the suppressive effect of the efferent system. The amount of attenuation is determined by the amount of efferent activity. Parameters associated with each module are shown in bold letters.

It should be noted that the effect of MOC activity can be further subdivided into two processes (fast and slow) according to the time scales upon which they function (Sridhar *et al.*, 1995; Cooper and Guinan, 2003). The purpose of this paper is to model the basic suppressive effect of the MOC system and not to distinguish between these fast and slow systems at this stage. It is, however, reasonable to assume that the effect of the efferent system implemented in our model is the fast effect and is analogous to the effect of electrical shocks to the MOC observed in the studies replicated in the following.

II. THE COMPUTER MODEL

The computer model of the auditory periphery and its parameters are taken from Meddis (2006). It consists of a cascade of modules representing the resonances of the outer/middle ear, the response of the BM, the inner hair cell (IHC) stereocilia, the IHC receptor potential, calcium dynamics, transmitter release and adaptation at the IHC/AN synapse and the generation of AN action potentials. Unless specified, the model is the same in all respects to the detailed description given in the appendix of Meddis (2006) where the parameters are specific to the guinea pig.

The output from the model is a stochastic train of spikes in one or more simulated AN fibers. A schematic diagram of the dual resonance nonlinear (DRNL) BM model is given in Fig. 1. One of the resonances is linear (top) and the other nonlinear (bottom). The nonlinear path has a compressive nonlinearity interpolated between two second-order gammatone filters. The compressive nonlinearity is a broken-stick function where compression is applied only above a certain level. The output from the model is the sum of the two resonances. Detailed discussions of this algorithm can be found in Meddis *et al.* (2001) and 2003.

The novel feature of the model is the attenuator marked ATT at the input to the nonlinear path. The amount of attenuation applied is a free parameter in this study with a value of 0 dB in the absence of efferent activity. When simulating the BM response during MOC stimulation, its value is varied to provide the best fit to the data.

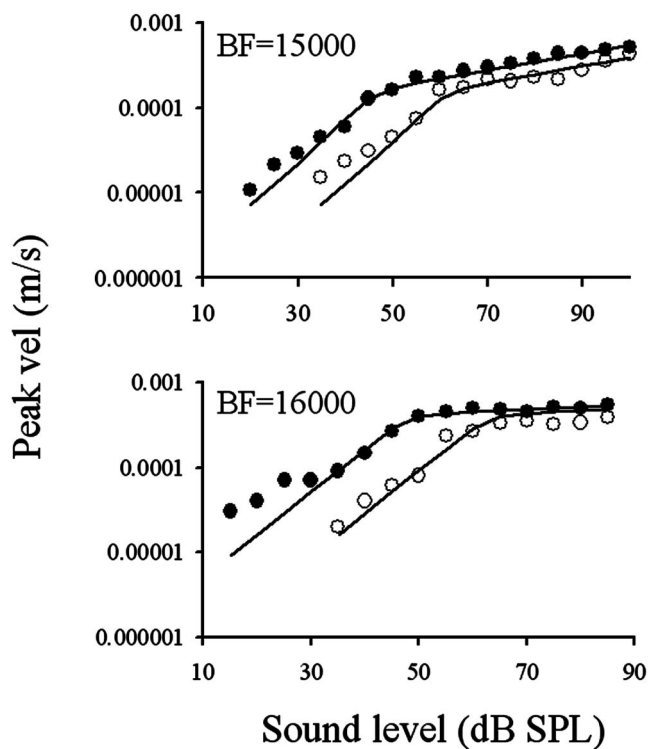


FIG. 2. Basilar membrane peak velocity (m/s) as a function of stimulus sound level for recording sites with BF of (A) 15 000 and (B) 16 000 Hz. Animal control data is shown as filled circles with the model control data as a line connecting these points. Measurements made during stimulation of the MOC are represented using unfilled circles and the model efferent data is shown as a line connecting the data points. Animal data is taken from Russell and Murugasu (1997). Model parameters are given in Table I.

The default parameters of all component modules are those used in Meddis (2006). Some of these parameters were developed in a series of studies by Sumner *et al.* (2002, 2003a, b). All computations were carried out using the MATLAB language and the code is available, on request, from the authors.

III. EVALUATION I: BASILAR MEMBRANE RESPONSE

Russell and Murugasu (1997) measured tone-evoked displacements of the BM in the basal turn of the guinea pig cochlea, with and without electrical stimulation of the olivocochlear bundle. They demonstrated that activation of the medial efferent system results in a reduction in the amount of basilar membrane displacement coupled with a shift of the rate-level function towards higher sound levels (see Fig. 2). Their data are plotted as individual points in Figs. 2(A) and 2(B) (control as filled circles and with MOC stimulation as unfilled circles). Results are shown for two experiments using measurements made at BM sites with BFs (best frequency) of 15 and 16 kHz. Their displacement measurements have been converted to peak velocity.

The data were simulated using a reduced version of the model incorporating only the outer/middle ear and BM functions. All relevant aspects of the animal experiments were replicated exactly. The stimulus was a BF pure tone with duration of 40 ms and rise and fall times of 2 ms. Stimuli were presented at levels ranging from 15 to 100 dB SPL in steps of 5 dB.

TABLE I. Parameter changes required to fit the computer model response to the physiological data of Russell and Murugasu (1997).

Parameter	BF	
	15 000 Hz	16 000 Hz
a	600	1000
b	0.008	0.0025
c	0.2	0.05
ATT (dB)	15	15

A. Model parameters

Parameters were adjusted to provide the best fit to the model in two stages. First the model was adjusted to give the best fit to the control data [solid circles in Figs. 2(A) and 2(B)]. Parameters a and b were free to vary at this stage. Parameter a is the main determinant of the steeply rising slope to the left of the function. Parameter b is the main determinant of the shallow slope to the right of the function. Parameter g (linear path) was set to zero throughout, effectively reducing the model to a single (nonlinear) path. In the second stage, only the ATT parameter was adjusted to give the best fit to the BM measurements during stimulation of the MOC bundle [unfilled circles in Figs. 2(A) and 2(B)]. In this second stage, only the ATT parameter was changed. All parameter changes were guided by a criterion based on the least squares best fit between the model and experimental data. The resulting parameters are given in Table I. The model data are presented in Figs. 2(A) and 2(B) as lines connecting the animal data points.

B. Model results

Comparison of the animal and model data shows that a simple attenuation of the input to the nonlinear path of the model is adequate to model the observed shift in the rate/level function resulting from stimulation of the MOC. An attenuation of 15 dB was found to be optimal for both sets of data.

Although the fit between the model and the animal data was adequate along most of both functions, it is noteworthy that the animal control data could not be adequately fit at the lowest signal levels. The animal data is linear between 15 and 50 dB SPL but shows a reduced gain around 35 dB SPL in both sets of measurements. There is no provision in the model to simulate this. The discrepancy was not pursued as it was not directly relevant to the issue of efferent stimulation.

IV. EVALUATION II: AUDITORY NERVE RESPONSE

Guinan and Stankovic (1996) measured AN fiber responses in the cat for tone bursts with and without electrical stimulation of the olivocochlear bundle. They demonstrated that activation of the medial efferent system results in a shift in the rate level function to higher levels (see Fig. 3). The largest reductions in firing rate occurred in low and medium spontaneous rate fibers at moderate sound levels (45–75 dB SPL) with substantial reductions also present at around 100 dB SPL. Their data are plotted as individual points in Fig. 3 (control data are filled circles, whereas MOC stimulation data are unfilled circles). Results are shown for six ex-

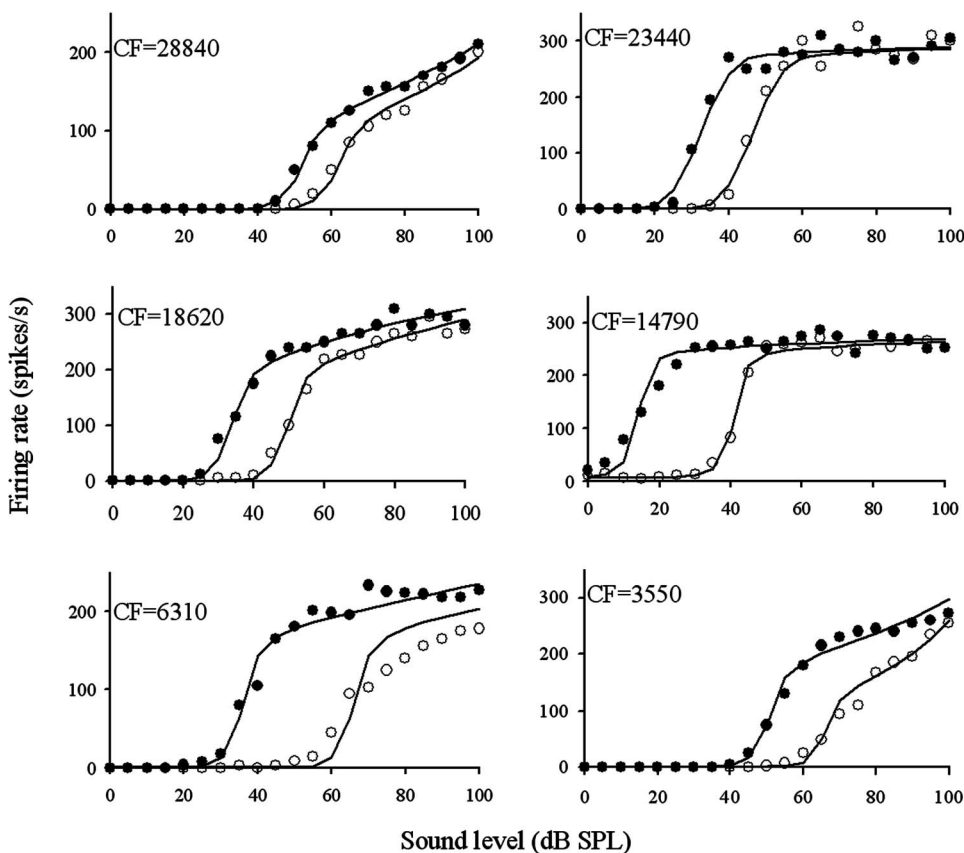


FIG. 3. Auditory nerve response (spikes/s) as a function of stimulus sound level (dB SPL). Individual figures show data from simulation of the model with stimuli of frequencies of 28 840, 23 440, 18 620, 14 790, 6310, and 3550 Hz for auditory nerve fibers with matching CFs (centre frequency). Animal control data is shown as filled circles with the model control data as a line connecting these points. Measurements made during stimulation of the MOC are represented using unfilled circles and the model efferent data is shown as line connecting these points. Animal data is taken from Guinan and Stankovic (1996). Model parameters are given in Table II.

TABLE II. Parameter changes used to fit the computer model to the animal data for the control (no MOC) condition (Guinan and Stankovic, 1996). ATT is the attenuation applied to the DRNL filter to simulate the effect of stimulation of the MOC bundle.

Parameter	CF					
	28 840 Hz	23 440 Hz	18 620 Hz	14 790 Hz	6310 Hz	3550 Hz
a	27 750	170 000	56 000	150 000	23 000	4500
b	0.165	0.6	0.092	0.018	0.012	0.014
G_{lin}	25	0	0	0	0	20
c	0.1	0.1	0.1	0.05	0.05	0.1
M	10	8	10	8	10	10
tauCa	0.00009	0.00015	0.00015	0.0004	0.00015	0.00015
ATT (dB)	10	14	16	27	30	23

periments using measurements recorded from auditory nerve fibers with CFs (centre frequency) of 28840, 23440, 18620, 14790, 6310 and 3550 Hz.

Although their results are qualitatively similar to that of Russell and Murugasu (1997), there are some important differences. The effect of MOC stimulation on the BM response was relatively simple and was similar for the two measurements reported previously. The data of Guinan and Stankovic (1996), however, show substantially different patterns of change for each fiber. No two fibers appear to respond in the same way. Given that the model aims to simulate MOC stimulation by changing only one parameter, it might appear unlikely that this would prove adequate to simulate the full variety of the animal AN data. In fact that is all that is necessary.

The computer model used to simulate the data of Guinan and Stankovic (1996) used the full range of peripheral modules; outer/middle ear, BM, IHC and AN responses. Their experimental protocol was followed as closely as possible. The stimuli were CF pure tones with a duration of 50 ms and rise and fall times of 2.5 ms. Stimuli were presented at levels ranging from 0 to 100 dB SPL in steps of 5 dB.

Guinan and Stankovic (1996) presented tones in succession, separated by 50 ms of silence. To reduce computational load, we evaluated the response of a cohort of 20 fibers with identical characteristics computed in parallel. As a consequence, our method does not include any cumulative adaptation that may have occurred in the animal recordings. Spike rates are based on a total count of all spikes occurring during the presentation of the stimulus.

A. Model parameters

The process of fitting the model parameters to the animal data proceeded in two stages in the same manner as described for the Russell and Murugasu (1997) data above. First, a range of parameters were adjusted to find a best-fit to the animal control data (filled circles in Fig. 3). In the second stage only the attenuation parameter (ATT) was changed to simulate the effect of MOC stimulation (unfilled circles in Fig. 3). All parameters of the model, before modification, are based on Meddis (2006). Table II shows only those parameters that were changed during the course of fitting the model to the data.

In the first stage we adopted various strategies to minimize the number of parameters that needed to be explored.

The gain, g , of the linear path in the BM model normally influences the output of the model only at high signal levels. This parameter was set to zero unless evidence was found of a change in slope of the rate/level function at very high signal levels. Parameter b , the BM gain above the compression threshold, does not influence the AN fiber rate if the fiber saturates at a signal level below that threshold. For two fibers (CF=14 790 and 23 440 Hz) it can be ignored. No changes were made to the compression exponent, c , of the model. Component bandwidths do not influence the model when measurements are made at BF and these were left at their default values. As a result, once the BF of the fiber had been set, only parameters a , b and, sometimes, g were changed in the BM component of the peripheral model. The only other parameters to be changed concerned the IHC/AN synapse. These were m (the maximum number of transmitter vesicles in the IHC/AN synapse) and tauCa (the time constant of clearance of IHC pre-synaptic calcium). The parameter m has a small effect on the saturated firing rate of the model fiber. The parameter tauCa is used to control the spontaneous rate of the fiber. The parameters used for each of the six simulations are given in Table II. The model data are presented in Fig. 3 as lines connecting the animal data points.

B. Model results

Comparison of the animal and model data shows that despite some imperfections, an acceptable fit was achieved in all cases for the purpose of studying the effect of MOC stimulation. The line connecting the unfilled circles shows the effect of attenuating the input to the BM model in order to simulate the effect of MOC stimulation. To achieve a good fit attenuations were required in the region of 10–30 dB (mean=19 dB). This is comparable to an attenuation of 15 dB required to simulate the data of Russell and Murugasu (1997). Four of the six evaluations give an excellent simulation of the effect of MOC stimulation. The two simulations at 6310 and 3550 Hz are probably acceptable but fail to replicate an apparent reduction in rate/level slope after MOC stimulation. This may, however, be the result of an initially imperfect fit to the control condition despite our best efforts.

V. EVALUATION III: CAP RESPONSE

Dolan and Nuttall (1988) recorded CAP responses to brief tones in silence and noise, with and without electrical

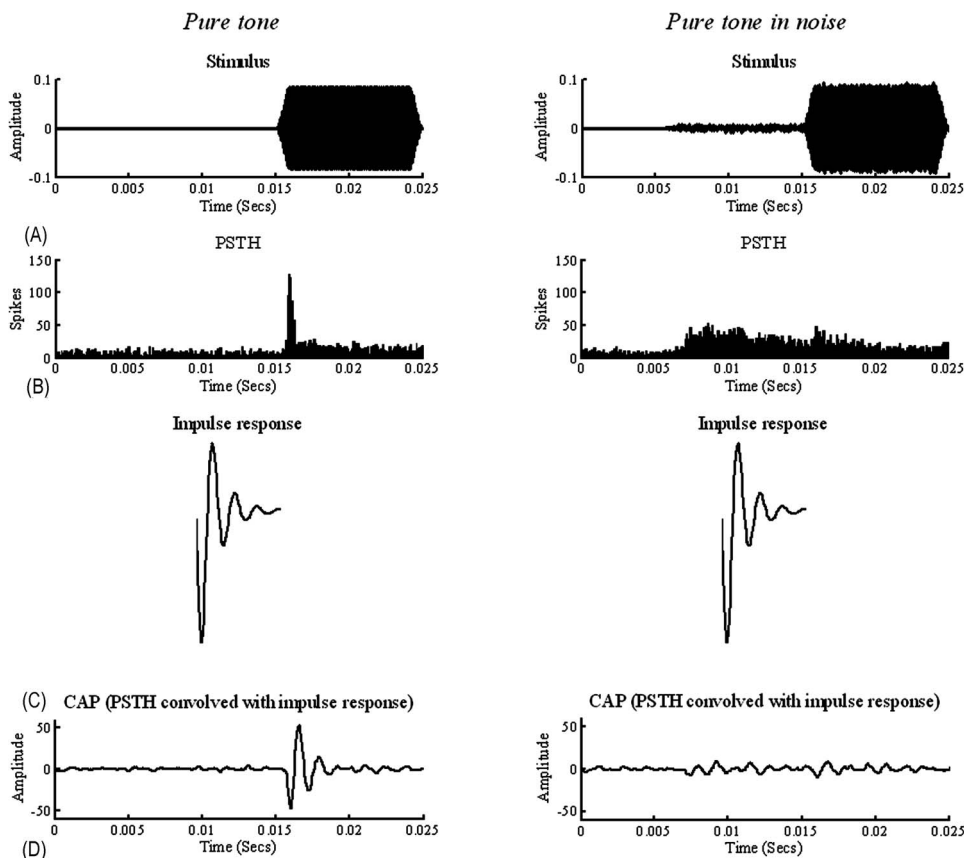


FIG. 4. Method used to simulate the auditory nerve compound action potential (CAP). (Left) Pure tone in silence and (right) pure tone in noise. (A) Initial stimulus. (B) The PSTH output from the auditory nerve stage of the model. (C) Impulse response as defined by Chertoff [2004, Eq. (3)] and (D) final peak CAP for the model during presentation of the tone after each spike in the (B) PSTH has been convolved with the (C) impulse response.

stimulation of the crossed olivocochlear bundle (see Fig. 5). They demonstrated that activation of the medial efferent system produces a shift to higher levels in the rate level function for tones, as demonstrated by Guinan and Stankovic (1996) and Russell and Murugasu (1997). When noise was added to the tones the CAP was reduced, presumably as a consequence of the adaptation caused by the noise. Dolan and Nuttall then demonstrated a partial restoration in the magnitude of the CAP response to tones in noise during MOC stimulation. This latter result is of considerable functional significance because it appears to represent an example of “unmasking” by efferent activity (Kawase and Liberman, 1993; Kawase *et al.*, 1993).

Our simulation of the data of Dolan and Nuttall (1988) used the full model of the auditory periphery including outer/middle ear, BM, IHC, and AN. Parameters for the guinea pig auditory periphery as used for Evaluation II were also used. Because CAP reflects synchronized activity in the AN as a whole, a multi-channel version of the model was implemented with fibers arranged across 30 channels with BFs distributed on a logarithmic scale between 1 and 20 kHz. Each channel had 50 fibers with identical parameters making a total of 1500 fibers altogether. The CAP from the model was computed by substituting every AN spike [across all channels in response to a tone stimulus—Fig. 4(A)] with a single unit impulse response as defined by Chertoff [2004, Eq. (3)]. The result of this is that for every spike in the auditory nerve poststimulus time histogram [PSTH, Fig. 4(B)] we add the impulse response illustrated in Fig. 4(C). Because of synchrony and desynchrony in the firing of each of the auditory nerve fibers the addition of each of these

impulse responses either adds up, or cancels out. This is particularly evident in Fig. 4(D) where synchrony as a result of the tone onset produces a large impulse/CAP response. The CAP measure used was the maximum height of the CAP function measured during the presentation of the tone.

The experimental protocol of Dolan and Nuttall (1988) was followed as closely as possible. Stimuli were brief pure tones with a duration of 10 ms, and rise and fall times of 1 ms. They were presented at levels ranging from 30 to 80 dB SPL. The broadband (white) noise (BBN) had a duration of 20 ms and rise and fall times of 2 ms. It began 5 ms before and ended 5 ms after the tone. The noise was delivered at one of three different levels (21, 26.5, and 31 dB SPL rms) according to the experimental data being simulated.

In total, three simulations were performed (one for each of the three noise masker levels) as for the original study. Four conditions were performed in each simulation: (i) tone alone, (ii) tone in noise, (iii) tone alone with MOC stimulation, and (iv) tone in noise with MOC stimulation.

A. Model parameters

Parameter adjustment was carried out in two stages as for the previous two simulations. In the first stage parameters were chosen to obtain a good fit to the rate/level functions in both the control conditions (tone alone and tone with noise). These parameters were then used without alteration for all conditions including all three noise levels. MOC stimulation was simulated by changing the ATT in the nonlinear path of the BM (see Fig. 1). An attenuation of 8.4 dB was used in all conditions when MOC stimulation was indicated. The pa-

TABLE III. Parameter changes used to fit the computer model response to the physiological data of Dolan and Nuttall (1988).

Parameter	
StapesScalar	5.5E-5
$\log(BWnl)$	$=1.0+0.70^{BF}$
ATT (dB)	8.4

parameter set began with the parameters given in Meddis (2006). Only two parameters required changing. The scalar applied to the stapes velocity was adjusted slightly from $1.4E-4$ to $5.5E-5$ to produce what is effectively a level adjustment of -8.1 dB. The parameters of the formula for computing the component bandwidths of the gammatone filters in the nonlinear path of the DRNL were adjusted to make the tuning curves of the individual channels slightly wider. The resulting parameters are given in Table III.

B. Model results

Figure 5 illustrates model and animal rate/level functions for auditory nerve fiber CAP magnitudes for noise masker levels of 21, 26.5, and 31 dB SPL. Animal data [Fig. 5 (left)] and model data [Fig. 5 (right)] are shown side by side for easier comparison. Data illustrated by unfilled circles shows data for the control condition (without BBN noise and without olivocochlear bundle stimulation). Data illustrated by unfilled squares shows data for the “BBN” condition (with noise masker and without olivocochlear bundle stimu-

lation). Data illustrated by filled circles shows data for the ‘COCB’ condition (without BBN noise and with olivocochlear bundle stimulation). Data illustrated by filled squares are for the “BBN + COCB” condition (with noise masker and with olivocochlear bundle stimulation). The model produced a useful qualitative fit throughout and a good quantitative fit. For noise levels of 21, 26.5, and 31 dB average correlations between the data and model results (across all four conditions) were 0.98, 0.98, and 0.97, respectively.

The effect of the MOC stimulation was an overall reduction in the observed CAP. This is consistent with the data reported earlier for the BM and AN. When noise was added to the tone the CAP was also reduced. This is a consequence of adaptation in the auditory nerve firing caused by the noise. The interesting result is the increase in the CAP for tones in noise (filled squares) when MOC stimulation is applied (filled squares). This occurs because the attenuation applied (to simulate MOC stimulation) reduces the response to the noise. This reduction in response reduces adaptation at the IHC synapse and also may cause desynchronization of the auditory nerve activity. As a result of the reduced activity in the system it is then possible for the presence of the tone to evoke a greater response.

VI. DISCUSSION

The main aim of this study was to evaluate a model of the auditory periphery with respect to its capacity to simulate the effect of MOC stimulation. The model is the same as an

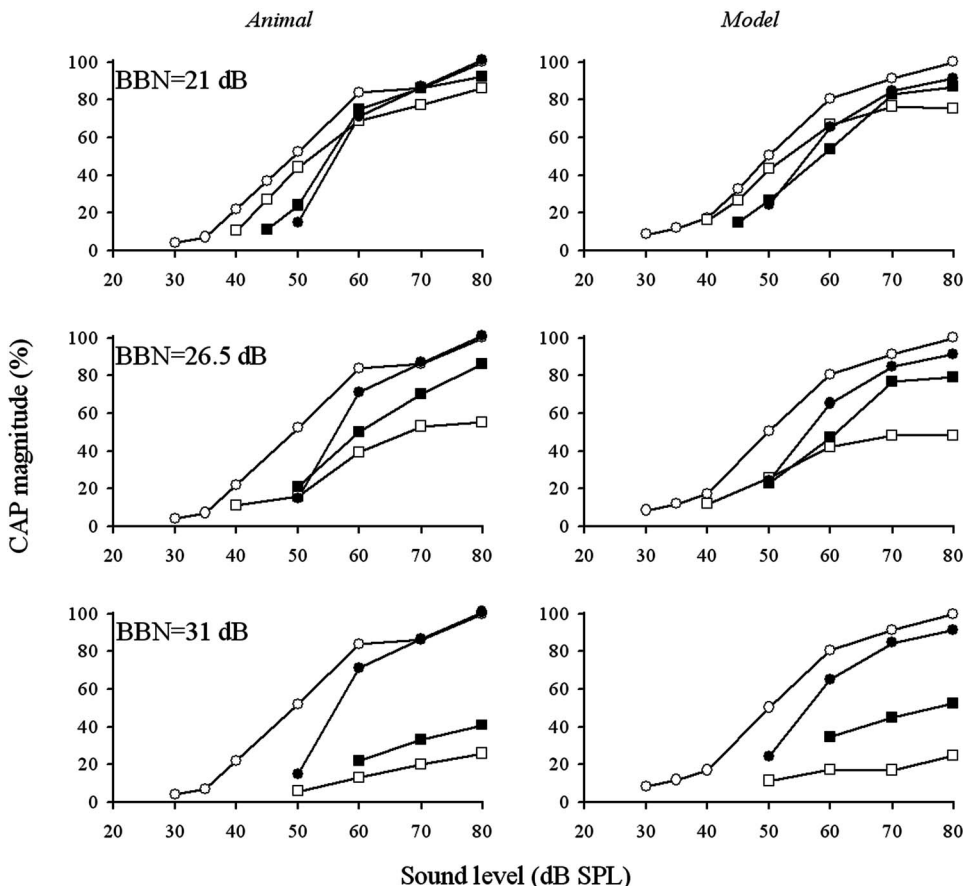


FIG. 5. Auditory nerve compound action potential (CAP) magnitude as a function of (left) stimulus sound level for the animal data and (right) model response. Graph shows the CAP magnitude in response to a 10 kHz tone alone (\circ), a tone in the presence of broadband noise (\square), a tone alone whilst stimulating the crossed olivocochlear bundle (\bullet) and a tone in the presence of broadband noise whilst stimulating the crossed olivocochlear bundle (\blacksquare). Noise levels used are 21, 26.5, and 31 dB SPL. Animal data is taken from Dolan and Nolan (1998). Model parameters are given in Table III.

existing model except for the addition of an attenuator (Fig. 1—ATT) in the nonlinear path of the DRNL module used to simulate the response of the BM. The evaluations involved simulations of animal data collected in different laboratories at three different levels of the auditory periphery (BM, AN, and CAP responses) during stimulation of the medial olivocochlear bundle. The simulations were qualitatively and quantitatively satisfactory. The attenuation required to simulate the observed effects differed between simulations across a range of 7–30 dB. It is possible that these differences can be understood in terms of the different methods used to stimulate the MOC bundle but the published reports do not allow a direct comparison.

Our original expectation was that the algorithm to simulate MOC stimulation would be more complex than a single attenuator. The attenuator was initially positioned prior to both the linear and nonlinear pathways of the DRNL filterbank, this resulted in an overall signal level reduction. Using a Euclidean distance measure of best fit we were able to determine that this method provided as good a fit to the BM data of [Russell and Murugasu \(1997\)](#) as the present method. When we modeled the AN response of [Guinan and Stankovic \(1996\)](#) the model fit to the animal data was worse, particularly at higher signal levels. Moving the attenuator so that it applied only to the nonlinear path rectified this and produced a good fit with the animal data, notably at higher signal levels where it had previously failed to fit for the AN data. As a result of our original approach failing to account for efferent effects it seems necessary that MOC attenuation acts on the nonlinear mechanics of the cochlea.

The present method was able to generate an adequate fit to the BM data of [Russell and Murugasu \(1997\)](#). The AN data reported by [Guinan and Stankovic \(1996\)](#) showed considerable variation between the response of AN fibers to MOC stimulation. Nevertheless the model was able to simulate these changes without any special reference to the fiber type. Once the fiber had been modeled for the control condition, an appropriate response to efferent stimulation followed naturally.

The modified model is intended as a tool for the study of the contribution of efferent effects in the processing of complex sounds such as speech and particularly speech in a masking background. A similar model has already been implemented in the context of speech recognition by [Ghitza et al. \(2006\)](#). Their aim was to demonstrate enhanced representations of speech in background noise. The positive outcome of this evaluation offers support to his strategy. In this respect simulating the data of [Dolan and Nuttall \(1988\)](#) is particularly relevant. They showed that the CAP amplitude/level function for tones is reduced when studied in a fixed noise background. However, they also demonstrated that the rate/level function is partly restored when the MOC bundle is stimulated and confirmed the relevance of the efferent system to the unmasking of sounds in noisy backgrounds ([Kawase et al., 1993](#)). It is encouraging that the model can simulate both of these effects; the suppression by noise and the unmasking by MOC activity.

VII. CONCLUSIONS

- (1) A simple attenuation of the basilar membrane response of the model is all that is required to reliably model the effect of stimulating the MOC system.
- (2) The model can be used to demonstrate the unmasking of signals in masking background noise by (efferent) attenuation of the input.
- (3) This makes this model a suitable platform on which we can further study the functional role of the MOC system in auditory perception.

ACKNOWLEDGMENTS

The authors are grateful to David Dolan, John Guinan, and Susan Shore for their insightful comments on an earlier version of the manuscript.

- 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.
- Cooper, N. P., and Guinan, J. J. (2003). "Separate mechanical processes underlie fast and slow effects of medial olivocochlear efferent activity," *J. Physiol. (London)* **548**, 307–312.
- Delgutte, B. (1990). "Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers," *J. Acoust. Soc. Am.* **87**, 791–809.
- Dolan, D. F., and Nuttall, A. L. (1988). "Masked cochlear whole-nerve response intensity functions altered by electrical stimulation of the crossed olivocochlear bundle," *J. Acoust. Soc. Am.* **83**, 1081–1086.
- Ghitza, O., Messing, D., Delhorne, L., Braida, L., Bruckert, E., and Sondhi, M. (2006). "Towards predicting consonant confusions of degraded speech," 14th International Symposium on Hearing (ISH), Cloppenburg, Germany.
- Goldstein, J. L. (1990). "Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering," *Hear. Res.* **49**, 39–60.
- Guinan, J. J. (2006). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," *Ear Hear.* **27**, 589–607.
- Guinan, J. J., and Stankovic, K. M. (1996). "Medial efferent inhibition produces the largest equivalent attenuation at moderate to high sound levels in cat auditory nerve fibres," *J. Acoust. Soc. Am.* **100**, 1680–1690.
- Kawase, T., Delgutte, B., and Liberman, C. M. (1993). "Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones," *J. Neurophysiol.* **70**, 2533–2549.
- Kawase, T., and Liberman, C. M. (1993). "Antimasking effects of the olivocochlear reflex. I. Enhancement of compound action potentials to masked tones," *J. Neurophysiol.* **70**, 2519–2532.
- Maison, S., Micheyl, C., and Collet, L. (2001). "Influence of focused auditory attention on cochlear activity in humans," *Psychophysiology* **38**, 35–40.
- Meddis, R. (2006). "Auditory-nerve first-spike latency and auditory absolute threshold: A computer model," *J. Acoust. Soc. Am.* **119**, 406–417.
- Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (2001). "A computational algorithm for computing nonlinear auditory frequency selectivity," *J. Acoust. Soc. Am.* **109**, 2852–2861.
- Rajan, R. (2000). "Centrifugal pathways protect hearing sensitivity at the cochlea in noisy environments that exacerbate the damage induced by loud sound," *J. Neurosci.* **20**, 6684–6693.
- Rasmussen, G. (1946). "The olivary peduncle and other fiber projections of the superior olivary complex," *J. Comp. Neurol.* **84**, 141–219.
- Russell, I. J., and Murugasu, E. (1997). "Medial efferent inhibition suppresses basilar membrane responses to near characteristic frequency tones of moderate to high intensities," *J. Acoust. Soc. Am.* **102**, 1734–1738.
- Sridhar, T. S., Liberman, M. C., Brown, M. C., and Sewell, W. F. (1995). "A novel cholinergic 'slow effect' of efferent stimulation on cochlear potentials in the guinea pig," *J. Neurosci.* **15**, 3667–3678.
- Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (2003a).

“Adaptation in a revised inner-hair cell model,” *J. Acoust. Soc. Am.* **113**, 893–901.
Sumner, C. J., O’Mard, L. P., Lopez-Poveda, E. A., and Meddis, R. (2002).
“A revised model of the inner-hair cell and auditory nerve complex,” *J.*

Acoust. Soc. Am. **111**, 2178–2189.
Sumner, C. J., O’Mard, L. P., Lopez-Poveda, E. A., and Meddis, R. (2003b).
“A non-linear filter-bank model of the guinea-pig cochlea,” *J. Acoust. Soc. Am.* **113**, 3264–3274.

Multifield coupled finite element analysis for sound transmission in otitis media with effusion

Rong Z. Gan^{a)} and Xuelin Wang

School of Aerospace and Mechanical Engineering, University of Oklahoma, Norman, Oklahoma 73019

(Received 24 May 2007; revised 9 September 2007; accepted 10 September 2007)

In this paper, a newly constructed three-dimensional finite element (FE) model of the human ear based on histological sections of a left ear temporal bone is reported. The otitis media with effusion was simulated in the model with variable fluid levels in the middle ear. The interfaces among the air, structure, and fluid in the ear canal and middle ear cavity were identified and the acoustic-structure-fluid coupled FE analysis was conducted when the middle ear fluid level was varied from zero to full fill of the cavity. The results show how the displacements of the tympanic membrane and stapes footplate or the middle ear transfer function is affected by fluid in the cavity across the auditory frequencies. Comparison of model results with measured data in temporal bones indicates that this model has the capability to extend FE analysis into pathological ears such as otitis media with visualized fluid-air interfaces inside the middle ear structures.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2793699]

PACS number(s): 43.64.Ha, 43.64.Bt [BLM]

Pages: 3527–3538

I. INTRODUCTION

The human ear includes complex morphological structures of the external ear canal, middle ear, and inner ear (or cochlea) to transfer sound from environment into central auditory system. Sound signals stimulate the air media in the ear canal and induce vibration of the tympanic membrane or eardrum. Movement of the tympanic membrane (TM) initiates the acoustic-mechanical transmission through three ossicular bones (i.e., malleus, incus, and stapes), which are connected by two joints (i.e., incudomalleolar and incudostapedial joints) and suspended in an air-filled cavity (i.e., middle ear cavity) by ligaments and muscle tendons. The output portion of the middle ear is the stapes footplate, which sets into the oval window and transmits vibrations into cochlear fluid.

In normal ears, the magnitude of stapes vibration is about 1 nm at the threshold of hearing measured from human temporal bones. In the ears of middle ear disease such as the otitis media with effusion (OME), a common middle ear disease diagnosed with fluid in the middle ear, the disease-induced change of middle ear structure affects the TM and stapes movements and results in conductive hearing loss. To detect the effect of fluid in the middle ear on TM movement, Ravicz *et al.* (2004) reported their measurement of umbo velocity in response to variation of fluid volume in human temporal bones. Recently, Gan *et al.* (2006a) published their measurements of both TM and stapes footplate movements in temporal bones with two laser Doppler vibrometers when the fluid level in the middle ear was varied from zero to full fill of the cavity. However, the studies on OME are limited in temporal bone and animal measurements or simple circuit model analysis without structure components of the ear.

The finite element (FE) method as a general numerical procedure has distinct advantages in modeling complex biological system such as the ear. Using the combined technologies of three-dimensional (3D) reconstruction of human ear and FE analysis, we have published two FE models of the human ear based on two complete sets of histological section images of temporal bones (Sun *et al.*, 2002; Gan *et al.*, 2004). These models have been employed for structure analysis, “one-chamber” (ear canal) acoustic-structural coupled analysis, and “two-chamber” (ear canal and middle ear cavity) air-structure-air coupled FE analysis (Sun *et al.*, 2002; Gan *et al.*, 2002, 2004, 2006b). In this study, we report a newly constructed 3D FE model of a left ear based on histological sections of a temporal bone. This model not only adds another individual model into our ear model database, but also extends the FE analysis into a new dimension for modeling pathological ears such as the OME. The multiple fields including the air in the canal and middle ear cavity, the fluid in the cavity, and the middle ear structures (i.e., bones and soft tissues), are coupled in the model to simulate various stages of the OME. The movements of the TM and stapes footplate and the transfer function of the middle ear in OME are derived from the model and compared with the data measured in human temporal bones. The first ever multicoupled FE analysis for sound transmission in OME is reported in this paper.

II. METHODS

A. 3D finite element model

1. Model geometry construction

An extracted fresh human temporal bone (male, age 52, left ear) obtained at the University of Oklahoma Health Sciences Center was used to prepare 720 histological sections of 20 μm thickness in celloidin, which were scanned into computer as section images (Fig. 1). The images were segmented to construct a 3D geometric model of the ear following the

^{a)}Author to whom correspondence should be addressed. Electronic mail: rgan@ou.edu

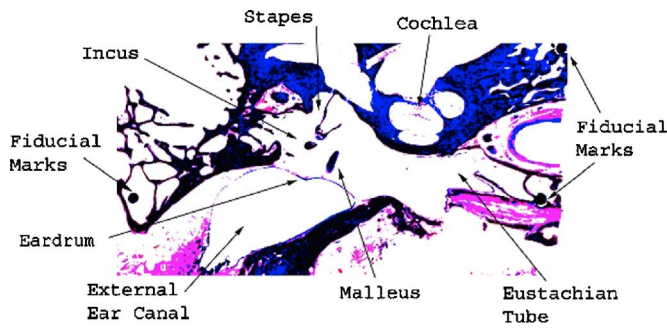


FIG. 1. (Color online) A typical histological section image of a human temporal bone (left ear, male, age 52).

details which can be found in our previous publications (Gan *et al.*, 2002; Sun *et al.*, 2002). Briefly, all the images were aligned with a template constructed from a typical section image using the fiducial marks on each section. The aligned images were then trimmed as standard-sized images and brought onto a sketch plane in CAD software, SOLIDWORKS (SolidWorks, Inc., Concord, MA). The aligned and treated images were then digitized by marking points along the outlines of the middle ear structures including the TM, three ossicles, attached ligaments and muscle tendons, middle ear cavity wall, and external ear canal wall. Finally, the images

were constructed to a 3D CAD model of human ear. Table I lists characteristic dimensions of the middle ear components, ear canal, and middle ear cavity measured from the model and compared with our previous FE model and published anatomical data.

2. Finite element modeling

The geometric models of the ear structures were imported into HYPERMESH (Altair Computing, Inc., Troy, MI) to generate the FE mesh and apply boundary conditions. Considering the restriction for element type used for fluid-structural interfaces in FE analysis software, ANSYS (ANSYS Inc., Canonsburg, PA), all structures including the ear tissues, bones, air in the canal and cavity, and fluid in the cavity were meshed with eight-noded hexahedral elements. Figure 2 shows the FE model of human left ear including the external ear canal, middle ear ossicles with attached ligaments, middle ear cavity, and cochlear load in anterior-medial view. The middle ear cavity was displayed transparently. The openings on the middle ear cavity wall: The oval window with stapedial annulus and stapes footplate sited on, the round window, and Eustachian tube opening, are also illustrated in Fig. 2. The areas of the oval window, round window, and Eustachian tube opening were measured as 3.86,

TABLE I. Dimensions and physical properties of human ear finite element (FE) model in comparison with published data.

Structure	The FE model	Our previous FE model (Gan <i>et al.</i> , 2004)	Published data
Eardrum			
Diameter along manubrium	9.762 mm	10.86 mm	8.0–10.0 mm (Gray, 1918 ^a)
Diameter perpendicular to manubrium	9.011 mm	9.24 mm	7.5–9.0 mm (Helmholtz, 1863 ^a)
Height of the cone	2.421 mm	1.46 mm	2.0 mm (Siebenmann, 1897 ^a)
Surface area	83.73 mm ²	72.01 mm ²	55.8–85.0 mm ² (Wever and Lawrence; Keith 1918; ^a Békésy 1941 ^a)
Thickness	0.05–0.10 mm	0.05–0.10 mm	0.1 mm (Helmholtz, 1863 ^a), 0.04–0.075 mm (Kirikae, 1960)
Malleus			
Length from the end of manubrium to the end of lateral process	4.764 mm	4.71 mm	5.8 mm (Stuhlman, 1937 ^a)
Total length	9.463 mm	8.11 mm	7.6–9.1 mm (Bast and Anson 1949 ^a)
Density (average)	3.59×10^3 kg/m ³	3.59×10^3 kg/m ³	3.59 kg/m ³ (Kirikae, 1960)
Weight	30.01 mg	30.42 mg	23–27 mg (Stuhlman, 1937; ^a Wever and Lawrence)
Incus			
Length along long process	6.661 mm	6.02 mm	7.0 mm (Stuhlman, 1937 ^a)
Length along short process	4.584 mm	4.58 mm	5.0 mm (Stuhlman, 1937 ^a)
Density (average)	3.23×10^3 kg/m ³	3.23×10^3 kg/m ³	3.23×10^3 kg/m ³ (Kirikae, 1960)
Weight	35.17 mg	26.47 mg	25–32 mg (Stuhlman, 1937; ^a Wever and Lawrence)
Stapes			
Height	4.18 mm	2.66 mm	2.5–4.0 mm (Stuhlman, 1937; ^a Wever and Lawrence)
Length of footplate	2.827 mm	2.64 mm	2.64–3.36 mm (Wever and Lawrence), 2.5 mm (our measurement)
Width of footplate	1.616 mm	1.32 mm	0.7–1.66 mm (Helmholtz, 1863; ^a Wever and Lawrence)
Density	2.20×10^3 kg/m ³	2.20×10^3 kg/m ³	2.20×10^3 kg/m ³ (Kirikae, 1960)
Weight	3.639 mg	1.93 mg	2.05–4.35 mg (Wever and Lawrence)
Ear canal			
Length (from umbo to entrance)	29.65 mm	30.22 mm	25–31 mm (Donaldson and Miller, 1973)
Volume	1.540×10^3 mm ³	1.657×10^3 mm ³	$0.83–1.972 \times 10^3$ mm ³ (Donaldson and Miller, 1973; Zwislocki, 1962)
Middle ear cavity			
Volume	678 mm ³	455 mm ³	600–1200 mm ³ (Ravicz <i>et al.</i> , 2004)

^aThe data from Wever and Lawrence (1954).

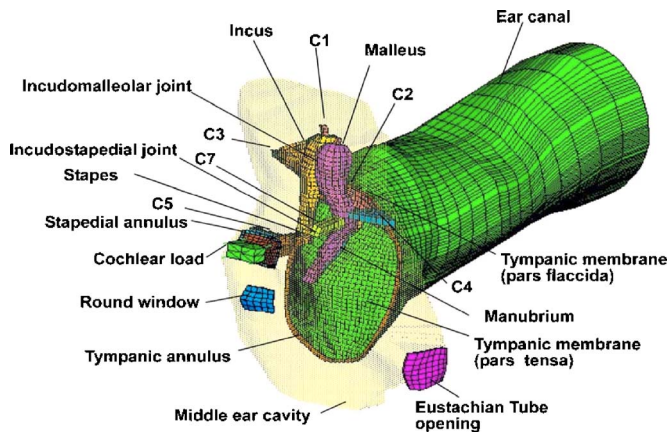


FIG. 2. (Color online) Finite element (FE) model of human left ear in anterior-medial view. The model consists of the tympanic membrane, three ossicles (malleus, incus, and stapes), two joints and manubrium, six ligaments and muscles tendons (C1, C2, C3, C4, C5, C7), tympanic annulus, stapedial annular ligament (or annulus), external ear canal, middle ear cavity, and cochlear load. The middle ear cavity was assumed transparent. The locations of three cavity openings: Oval window through the stapes footplate and stapedial annulus, round window, and Eustachian tube opening are displayed on the cavity wall. Explanations of C1, C2, C3, C4, C5, and C7 are given in Table III.

2.76, and 6.38 mm² from the model, respectively.

Following the common rule of thumb that at least six linear elements (usually seven to ten elements) per shortest acoustic wavelength are adopted for acoustic scattering (Ihlenbury, 1998), the maximum element edge size of 0.0043 m was estimated at 8000 Hz, i.e., $\frac{1}{10}(c/f) = \frac{1}{10}(343 \text{ m/s}/8000/\text{s}) = 0.0043 \text{ m}$, where c is sound speed in air and f is frequency. The hexahedral mesh generation is a time-consuming process and it is difficult to refine the mesh for such a complicated ear system. In fact, the maximum element size for the model is much smaller than 0.004 m and the mesh is fine enough. The TM was meshed by 1151 solid elements (ANSYS Solid 45) and the ossicular bones were meshed by a total of 1585 elements. The incudomalleolar joint, incudostapedial joint, and manubrium that connects the malleus handle to the TM, were meshed by a total of 109 elements. The air in the external ear canal with a volume of 1540 mm³ was meshed by eight-node acoustic elements at a total of 31,652 elements. The volume of the middle ear cavity was measured as 678 mm³, which did not include the mastoid air cells in the model. In the normal ear condition, the air was filled in the cavity and meshed with 15,820 eight-node acoustic elements. The FE mesh for the ear with OME is described in Sec. 2 B.

The human middle ear was described as a linear system under regular sound intensity of the hearing level for acoustic-mechanical transmission from the TM to cochlea. The ossicles, ligaments, and tendons were assumed as isotropic materials, while the TM was assumed as orthotropic material. The mechanical properties of the TM, ossicles, joints, and manubrium in the model (Table II) were adopted based on our previous investigation (Gan *et al.*, 2004), except that the Young's modulus of incudomalleolar joint was assumed as $0.14 \times 10^9 \text{ N/m}^2$. The Poisson's ratio was assumed 0.3 for all materials of the middle ear. The Raleigh damping parameters α and β for all materials were assumed to be 0 s^{-1} and

$0.75 \times 10^{-4} \text{ s}$, respectively. Note that the properties of fluid in middle ear cavity listed in Table II were used for analysis of the otitis media with effusion in the ear.

The structural boundaries of the middle ear include tympanic annulus, middle ear suspensory ligaments or muscles, stapedial annular ligament, and cochlear fluid. Since the vibration modes of the stapes in response to acoustic stimulation show complex movement patterns and significant rocking is observed at higher frequencies, a mass block with ten dashpots (five on each side) was used to model the effect of cochlear fluid on stapes vibration. The damping coefficient of each dashpot was calculated as 0.02 N s/m. This assumption and calculation of cochlear load was based on the experimental studies of cochlear impedance on human temporal bones (Aibara *et al.*, 2001; Rosowski and Merchant, 1995) and the circuit model calculation for effective mass of the cochlear fluid (Zwislocki, 1962; Kringlebotn and Funderston, 1988; Merchant *et al.*, 1996). The Young's moduli for suspensory ligaments and muscle tendons, tympanic annulus, and stapedial annulus (Table III) were assumed as the same as our previous structural analysis (Gan *et al.*, 2007a). The ear canal was open to the atmosphere, and the boundary condition at the canal entrance was set free. The round window and Eustachian tube was fully restrained at the peripheries.

B. 3D FE modeling of OME

To simulate the OME with variable fluid volume in the middle ear (Gan *et al.*, 2006a), the entire middle ear cavity of the FE model was meshed or divided into nine subdivisions as shown in Fig. 3. The volume of each subdivision V_i was listed in Table IV as well as the summed up volume $\sum V_i$, $i = 1-9$, and percentage of the summed up volume to the whole cavity volume, $\sum V_i/V$. The fluid level filled in the cavity was increased stepwise from the inferior bottom (V_1) to the superior top (V_9) and full fill of the cavity. Fluid in the cavity was meshed with eight-node hexahedral fluid elements (Fluid 80 in ANSYS). There are four types of fluid-structural interfaces (FSIs) created in the model for FE analysis: (1) Air-fixed structure FSI, such as the bony walls of the canal and middle ear cavity; (2) air-movable structure FSI, such as the TM, ossicles, and suspensory ligaments contacted to air in the ear canal and middle ear cavity; (3) viscous fluid-structure FSI, such as the bony wall, TM, and ossicles contacted to fluid in the cavity; and (4) viscous fluid-air FSI, such as the middle ear cavity partially filled with fluid. The no-slip condition was assumed between the fluid elements and structure solid elements (e.g., the TM and middle ear cavity bony wall contacted to fluid).

Corresponding to the fluid levels divided in the cavity, the fluid level on the TM was meshed into six layers starting one level above the bottom of the cavity as displayed in Fig. 4. Note that fluid in the cavity represents liquid unless there is a special notice for air. The height of each fluid layer from the bottom L1 to the top L6 was 1.75, 1.55, 1.45, 1.55, 1.50, and 2.25 mm, respectively. The fluid level approaching the umbo or just below the umbo was approximately equal to 38% of the cavity volume filled up with fluid; the fluid level reaching the oval window or covering the umbo totally was

TABLE II. Material properties used for middle ear structures of the FE model.

Structure	The FE model	Sources
Eardrum		
Density	$1.2 \times 10^3 \text{ kg/m}^3$	$1.2 \times 10^3 \text{ kg/m}^3$ (Wada and Metoki, 1992)
Young's modulus: Pars tensa	$3.5 \times 10^7 \text{ N/m}^2$ (radial)	$2.0 \times 10^7 \text{ N/m}^2$ (von Békésy, 1960)
	$2.0 \times 10^7 \text{ N/m}^2$ (circumferential)	$4.0 \times 10^7 \text{ N/m}^2$ (Kirikae, 1960)
Pars flaccida	$1.0 \times 10^7 \text{ N/m}^2$ (radial)	
	$1.0 \times 10^7 \text{ N/m}^2$ (circumferential)	
Malleus		
Density:		
Head	$2.55 \times 10^3 \text{ kg/m}^3$	$2.55 \times 10^3 \text{ kg/m}^3$ (Kirikae, 1960)
Neck	$4.53 \times 10^3 \text{ kg/m}^3$	$4.53 \times 10^3 \text{ kg/m}^3$ (Kirikae, 1960)
Handle	$3.70 \times 10^3 \text{ kg/m}^3$	$3.70 \times 10^3 \text{ kg/m}^3$ (Kirikae, 1960)
Young's modulus	$1.41 \times 10^{10} \text{ N/m}^2$	$1.41 \times 10^{10} \text{ N/m}^2$ (Herrmann and Liebowitz, 1972)
Incus		
Density:		
Body	$2.36 \times 10^3 \text{ kg/m}^3$	$2.36 \times 10^3 \text{ kg/m}^3$ (Kirikae, 1960)
Short process	$2.26 \times 10^3 \text{ kg/m}^3$	$2.26 \times 10^3 \text{ kg/m}^3$ (Kirikae, 1960)
Long process	$5.08 \times 10^3 \text{ kg/m}^3$	$5.08 \times 10^3 \text{ kg/m}^3$ (Kirikae, 1960)
Young's modulus	$1.41 \times 10^{10} \text{ N/m}^2$	$1.41 \times 10^{10} \text{ N/m}^2$ (Herrmann and Liebowitz, 1972)
Stapes		
Density	$2.20 \times 10^3 \text{ kg/m}^3$	$2.20 \times 10^3 \text{ kg/m}^3$ (Kirikae, 1960)
Young's modulus	$1.41 \times 10^{10} \text{ N/m}^2$	$1.41 \times 10^{10} \text{ N/m}^2$ (Herrmann and Liebowitz, 1972)
Incudomalleolar joint		
Density	$3.2 \times 10^3 \text{ kg/m}^3$	$3.2 \times 10^3 \text{ kg/m}^3$ (Sun <i>et al.</i> , 2002)
Young's modulus	$0.14 \times 10^9 \text{ N/m}^2$	$1.41 \times 10^{10} \text{ N/m}^2$ (Sun <i>et al.</i> , 2002)
Incudostapedial joint		
Density	$1.2 \times 10^3 \text{ kg/m}^3$	$1.2 \times 10^3 \text{ kg/m}^3$ (Sun <i>et al.</i> , 2002)
Young's modulus	$6.0 \times 10^5 \text{ N/m}^2$	$6.0 \times 10^5 \text{ N/m}^2$ (Wada <i>et al.</i> , 1997; Prendergast <i>et al.</i> , 1999)
Manubrium		
Density	$1.0 \times 10^3 \text{ kg/m}^3$	$1.0 \times 10^3 \text{ kg/m}^3$ (Sun <i>et al.</i> , 2002)
Young's modulus	$4.7 \times 10^9 \text{ N/m}^2$	$4.7 \times 10^9 \text{ kg/m}^2$ (Sun <i>et al.</i> , 2002)
Fluid in middle ear cavity		
Density	$1.0 \times 10^3 \text{ kg/m}^3$	
Bulk Modulus	$2.2 \times 10^9 \text{ N/m}^2$	
Viscosity	$1.0 \times 10^{-3} \text{ N s/m}^2$ or 1 cp	
Damping (β)	$1.5 \times 10^{-4} \text{ s}$	

approximately equal to 50% of the cavity volume filled with fluid; and the fluid level covering the stapes was equivalent to 61% of the cavity volume filled by fluid. The fluid level covering the entire TM was equal to 75% of the cavity volume filled up by fluid.

Fluid in the middle ear may cause alterations in geom-

etry and materials of the middle ear structure. Mechanical properties of the soft tissue may change when exposed to liquid, and on the other hand, hydrostatic pressure acted on the TM may not be uniformly distributed. These alterations result in deformation of the TM and ossicular chain away from the initial shape. Especially, the deformation could lead

TABLE III. Boundary conditions of human ear FE model.

Ligaments	Young's modulus (N/m ²)	
	The FE model	Our previous FE model
Superior malleolar ligament (C1)	4.9×10^6	1.0×10^5 (Gan <i>et al.</i> , 2002, 2004)
Lateral malleolar ligament (C2)	6.7×10^6	1.0×10^5 (Gan <i>et al.</i> , 2002, 2004)
Posterior incudal ligament (C3)	6.5×10^6	6.5×10^5 (Wada <i>et al.</i> , 1997; Prendergast <i>et al.</i> , 1999)
Anterior malleolar ligament (C4)	8.0×10^6	8.0×10^6 (Gan <i>et al.</i> , 2007a)
Posterior stapedial tendon (C5)	10.0×10^6	10.0×10^6 (Gan <i>et al.</i> , 2007a)
Tensor tympani tendon (C7)	8.0×10^6	8.0×10^6 (Gan <i>et al.</i> , 2007a)
Tympanic annulus	6.0×10^5	6.0×10^5 (Gan <i>et al.</i> , 2002; 2004; Sun <i>et al.</i> , 2002)
Stapedial annular ligament	1.2×10^6	2.0×10^5 (Gan <i>et al.</i> , 2006a, b)

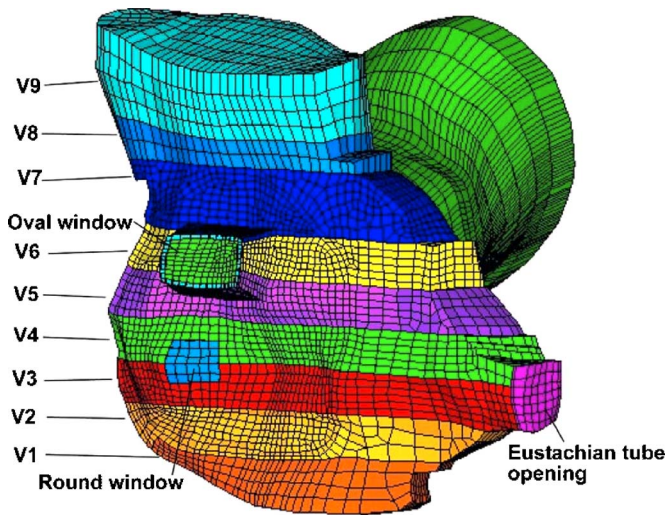


FIG. 3. (Color online) Finite element mesh of the middle ear cavity in anterior-medial view. The cavity was meshed into nine subdivisions of V1–V9 from the inferior bottom to the superior top to simulate the variation of fluid levels inside the cavity.

to a change of the incudostapedial joint status. Thus, the middle ear structural change induced by otitis media with effusion appears to be more complicated than the normal middle ear structure for sound transmission. Material properties of the soft tissue and joint may be different from the normal condition.

As a preliminary study conducted in this model, we assumed that the material properties of most middle ear components were not affected by fluid in the cavity, but the stiffness of incudostapedial (IS) joint was changed when middle ear fluid level reached to 50% of the cavity volume. This assumption was based on the following observations that the volume reduction of the middle ear air space by fluid could reduce the middle ear input compliance reported by Ravicz *et al.* (2004). The compliance of the middle ear air space was decreased by a factor of 10 when the saline solution was filled in the middle ear from 0 to 90% of the volume. However, the experimental measurement on transfer function of the middle ear such as the displacement transmission ratio of the tympanic membrane to stapes footplate (DTR) by Gan *et al.* (2006a) suggested that only a very small increase of DTR at frequencies below 1000 Hz and a maximum variation between 1.5 and 4 were observed as fluid filled up the cavity. The fluid changes the middle ear stiffness or compliance, but the DTR is not significantly altered. Therefore, a possible change of the IS joint stiffness due to the fluid level increasing in the middle ear was proposed in this study, which reflected the reduction of relative displacement between the incus and stapes by fluid accumulated in the middle ear cavity. We assumed that the stiffness of the IS joint is a variable

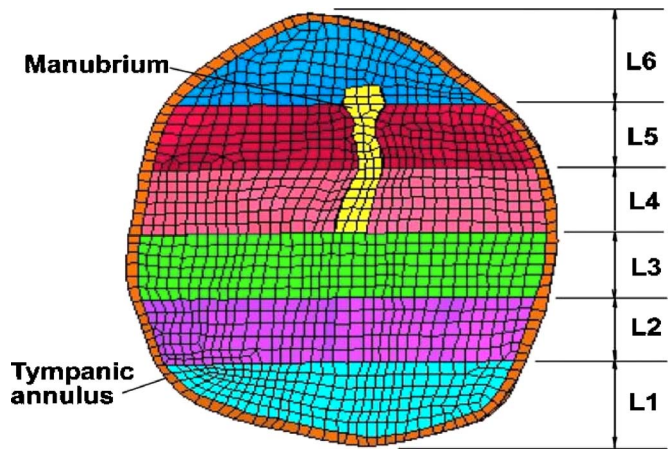


FIG. 4. (Color online) Finite element mesh of the tympanic membrane (TM) in medial view. The TM was meshed into six layers of L1–L6 corresponding to the subdivisions of the middle ear cavity shown in Fig. 3.

with respect to fluid in the cavity. To test this hypothesis, we compared the measured DTR in temporal bones with that calculated by the FE model. The elastic modulus of the IS joint was considered to vary nonlinearly with fluid volume. The elastic modulus of the IS joint was 1.2, 2.8, 8.0, 11.0, and 15.0 MPa when fluid volume in the middle ear cavity was increased to 50%, 61%, 75%, 82%, and 100% of the cavity volume, respectively.

C. Acoustic-structure-fluid coupled FE analysis

1. Two approaches for fluid-structure coupling

In this study, the canal-TM and ossicles-middle ear cavity or “two-chamber” (ear canal and middle ear cavity) coupled analysis was extended to include fluid in the cavity. This resulted in a difficult task to predict the response of an acoustic-structure-fluid coupled system with four types of FSIs in ANSYS. There are two approaches to describe the motion of fluid: Eulerian and Lagrangian methods (Zienkiewicz and Bettess, 1978), when an elastic structure and fluid are coupled. In the Eulerian approach the governing equation can be written as a wave equation in terms of pressure (or velocity potential). The use of pressure variable reduces the number of degrees of freedom at the node to one and the coupling is achieved by consideration of interface forces (e.g., element Fluid 30 for acoustic analysis used in this study). In the Lagrangian approach, the fluid region is written in terms of displacement. The governing equation for fluid is the same as that for solid and the element matrices in fluid and solid regions can be easily assembled (e.g., element Fluid 80 for fluid analysis and Solid 45 for structure analysis used in this study). Thus, the compatibility and equilibrium are automatically satisfied along the interface between the

TABLE IV. Volume data of nine subdivisions of the middle ear cavity shown in Fig. 3.

	V1	V2	V3	V4	V5	V6	V7	V8	V9
Volume (mm ³)	36	58	79	86	78	77	96	49	119
ΣV_i (mm ³)	36	93	172	258	336	412	509	559	678
$\Sigma V_i/V$	5%	14%	25%	38%	50%	61%	75%	82%	100%

fluid and structure. In this study, the air in the ear canal and inside the cavity was described with the Eulerian approach; the fluid in the middle ear cavity was described with the Lagrangian approach. In other words, two different elements in ANSYS (i.e., Fluid 30 and Fluid 80) were used for acoustic analysis from the ear canal to the cavity.

2. Acoustic-structure or air-structure coupling

If the convective acceleration and viscous effects are neglected, the well-known governing equation for pressure distribution within the air region is

$$\frac{\partial^2 P}{\partial t^2} - c^2 \nabla^2 P = 0, \quad (1)$$

where P is acoustic pressure, c is speed of sound and $c = \sqrt{k/\rho_0}$ in fluid medium, ρ_0 is mean density of air, k is bulk modulus of air, and t is time. The speed of sound and density of the air were assumed as 343 m s^{-1} and 1.21 kg m^{-3} , respectively. At the acoustic-structure interfaces (e.g., the lateral surface of the TM and the medial surface of the TM without fluid attached) where the pressure gradient of the air is related to the normal acceleration of the structure by fluid momentum equations, the discretized wave equation can be coupled with discretized structural equation. Thus, coupling is handled by calculating element stiffness matrices and load vectors that contain all necessary degrees of freedom.

The surface of acoustic elements (air) next to the fixed structure, such as the canal and middle ear cavity bony walls, was the first type FSI and named the ‘‘impedance surface.’’ The surface of acoustic elements next to the movable structure such as the TM, ossicles, and suspensory ligaments, was defined as the second type FSI where the acoustic pressure distribution was coupled into structural analysis as the force input in ANSYS. Since sound absorption at the fluid-structure interface exists, the acoustic absorption coefficient is specified to the elements contacting the structure as material properties. The absorption coefficient values for the TM, canal wall, and cavity wall are 0.007, 0.02, and 0.04, respectively, as reported by Gan *et al.* (2006b). The TM (both pars tensa and pars flaccida) had FSIs on its lateral and medial sides.

3. Fluid-structure and air-fluid coupling

In the fluid region which was used to simulate otitis media with effusion, the motion of fluid or effusion is described by

$$P = -k \nabla \cdot \mathbf{u}, \quad (2)$$

where \mathbf{u} is the displacement vector of fluid and k is bulk modulus of fluid. In the present FE modeling, the damping matrix related to the viscosity of fluid was based on

$$\begin{pmatrix} \nabla \dot{\mathbf{u}} \\ \dot{\gamma}_{xy} \\ \dot{\gamma}_{yz} \\ \dot{\gamma}_{xz} \\ \dot{R}_x \\ \dot{R}_y \\ \dot{R}_z \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \eta^{-1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \eta^{-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \eta^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta^{-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta^{-1} \end{pmatrix} \begin{pmatrix} P \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{xz} \\ M_x \\ M_y \\ M_z \end{pmatrix}, \quad (3)$$

where $(\dot{\cdot})$ represents differentiation with respect to time. γ_{xy} , γ_{yz} , and γ_{xz} are shear strains, τ_{xy} , τ_{yz} , and τ_{xz} are shear stresses, $R_i (i=x, y, z)$ is the twisting force about i axis, M_i is the rotation about i axis, η is the viscosity of fluid, and $\theta = \eta \times 10^{-5}$. In this study, the fluid in the middle ear cavity was assumed as normal saline solution with a density of 1000 kg/m^3 and viscosity of $1.0 \times 10^{-3} \text{ N s/m}^2$ or 1 cp. The bulk modulus and damping coefficient (β) was assigned as $2.2 \times 10^9 \text{ N/m}^2$ and $1.5 \times 10^{-4} \text{ s}$, respectively (see Table II).

Because the motion of fluid was described by its displacement in the same manner as the motion of structure, the compatibility and equilibrium were automatically satisfied along the interface between the fluid and solid (the third type FSI in this study). Similarly, the interface between the fluid and air in the cavity (the fourth type FSI in this study) was considered in the same manner as the interface of acoustic-structure coupling.

III. RESULTS

A. Validation of the FE model with temporal bone experimental data

The FE model of normal ear was first validated by comparison of displacement curves of the TM (at the umbo) and stapes footplate derived from the model with that measured from human temporal bones. A uniform harmonic sound pressure of 90 dB SPL (0.632 N/m^2 , rms value) was applied at 2 mm away from the umbo in the ear canal of the FE model. The magnitudes of displacement at the TM and footplate were calculated from the model using ANSYS software and projected to the piston-direction (i.e., the direction perpendicular to the footplate), which is the same direction as that measured from temporal bones using laser Doppler vibrometers (Gan *et al.*, 2006a, 2007b).

Figure 5 shows the model-derived frequency response curve of the TM displacement at the umbo (thick solid line) in comparison with the umbo curves measured from seven temporal bones when 90 dB SPL was applied on the TM in the ear canal. Figure 5(a) displays the magnitudes and Fig. 5(b) shows the phase angles. The mean amplitude and phase curves obtained from seven bones are also displayed in Figs. 5(a) and 5(b) (dashed line). The data measured from temporal bones were reported by Gan *et al.* (2007b) using two laser vibrometers on the umbo and stapes footplate to measure vibrations of the TM and stapes footplate simultaneously.

The model-derived frequency response curves of the stapes footplate displacement are compared with the experimental data measured from seven bones in Fig. 6. The mean

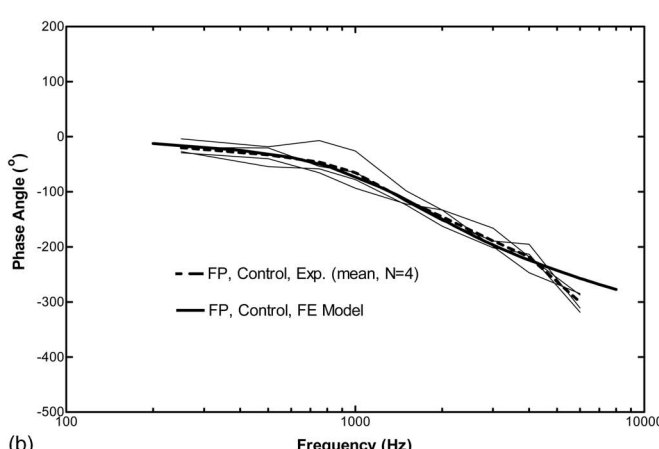
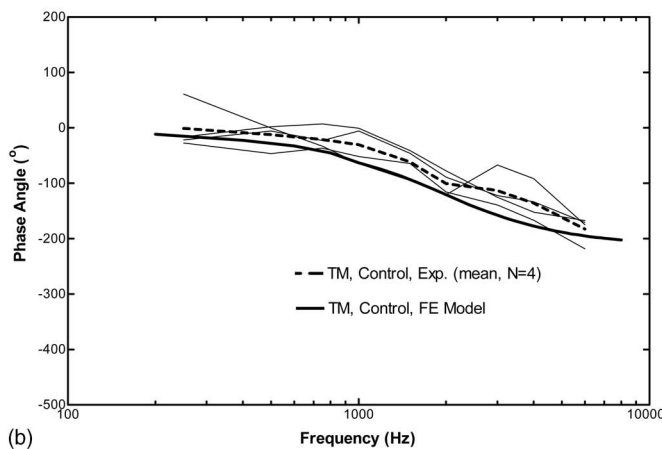
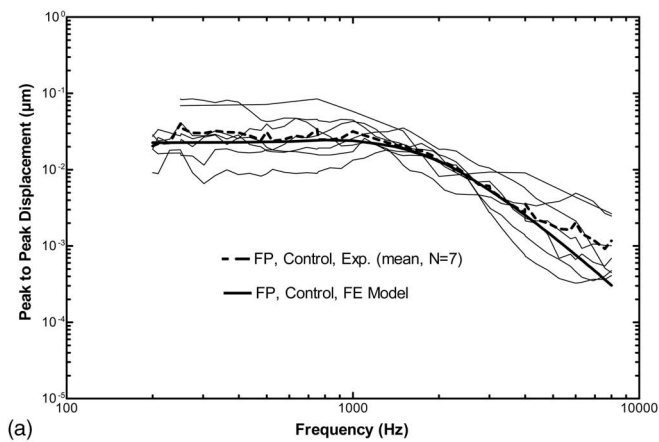
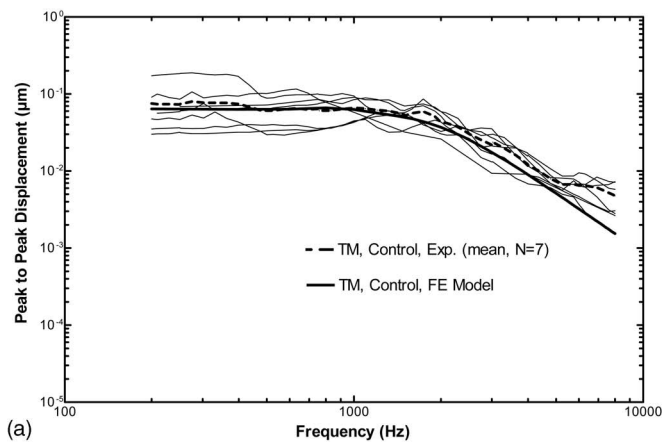


FIG. 5. Comparison of the FE model-derived TM (at the umbo) displacement data with the measurements from human temporal bones. The input sound pressure level was 90 dB at 2 mm away from the umbo in the ear canal. The solid lines represent the model results and the broken lines represent the mean curves obtained from temporal bones reported by Gan *et al.* (2007b). (a) Magnitude and (b) phase angle.

FIG. 6. Comparison of the FE model-derived stapes footplate displacement data with the measurements from human temporal bones. The input sound pressure level was 90 dB at 2 mm away from the umbo in the ear canal. The solid lines represent the model results and the broken lines represent the mean curves obtained from temporal bones reported by Gan *et al.* (2007b). (a) Magnitude and (b) phase angle.

amplitude and phase curves of the footplate displacement from seven bones are also displayed in Figs. 6(a) and 6(b), respectively. As can be seen in Figs. 5 and 6, the FE model-predicted umbo and stapes footplate displacement curves well fall into the range of seven temporal bone experimental curves across the frequency range of 200–8000 Hz. Good agreement between the model and temporal bone data obtained from the normal middle ear without fluid in the cavity indicates that the FE model is suitable to conduct the study on pathological middle ear such as the OME.

B. Effect of fluid in the middle ear cavity on TM and stapes movements

Figure 7 shows the effect of fluid (e.g., normal saline solution) in the middle ear cavity on displacement of the TM at the umbo predicted by the FE model. The amount of fluid in the cavity was increased stepwise from 0 (control) to 25, 38, 50, 61, 75, 82, and 100% of the cavity volume. Figure 7(a) displays the frequency response curves of the displacement magnitude and Fig. 7(b) presents the phase angle.

The results in Fig. 7 show that significant reductions of the umbo displacement were predicted by the model at frequency greater than 300 Hz and the umbo displacement was

strongly affected by the volume of fluid in the cavity. At frequency range of 1000 to 8000 Hz, the TM displacement at the umbo was decreased by 8–10 dB, when 50% of the cavity was filled with fluid. When cavity was totally occupied by fluid, a 26–32 dB reduction of the umbo movement was calculated from the model. There are some variations of the umbo displacement at low frequencies (e.g., 400 Hz) when fluid volume was below 50% of the cavity volume.

The model-derived umbo displacements in response to fluid level of 50% and 100% of the middle ear cavity are consistent with the data measured in human temporal bones by Gan *et al.* (2006a). A critical volume of fluid in the cavity (0.3 ml) at which the umbo velocity or displacement started decreasing significantly in comparison with the normal, no fluid condition was identified from the experimental data shown in Fig. 9a of Gan *et al.* (2006a). The volume 0.3 ml is approximately equal to the fluid level filled up to the umbo or about 50% of the middle ear cavity volume. This experimental observation was predicted by our FE model in which the fluid level reaching 50% of the cavity volume resulted in a significant reduction of the umbo displacement (8–10 dB).

The phase change induced by fluid in the middle ear [Fig. 7(b)] shows that the phase became more negative as the

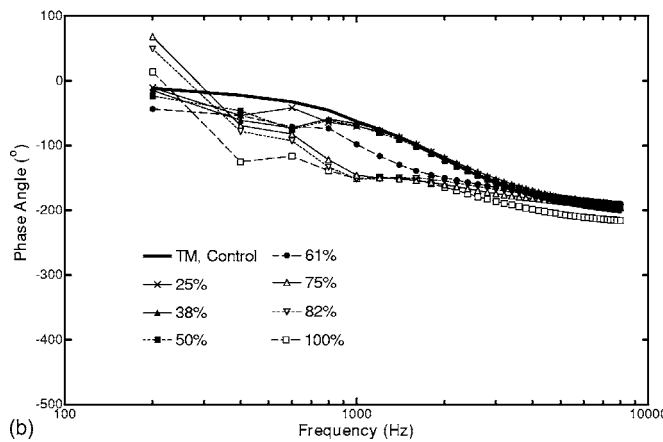
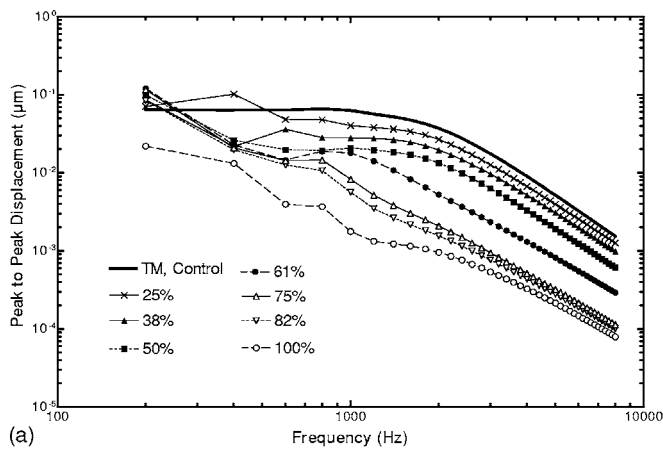


FIG. 7. FE model-predicted TM (at the umbo) displacement curves when fluid volume in the middle ear cavity varied from 25% to 100% of cavity volume. The input sound pressure was 90 dB SPL at 2 mm away from the umbo in the ear canal. (a) Magnitude and (b) phase angle.

amount of fluid in the cavity increased at frequency over 300 Hz, which is similar to the results obtained from temporal bone experiments in Fig. 9b of Gan *et al.* (2006a). However, the FE model shows that the phase change was not sensitive at high frequencies ($f > 4000$ Hz), which is different from the bone results.

The effect of fluid in the middle ear on displacement of the footplate is displayed in Figs. 8 and 9. The results in Fig. 8 were derived from the model with cochlear load or with intact cochlea when fluid volume in the middle ear was varied from 0 (control) to 25, 38, 50, 61, 75, 82, and 100% of the cavity volume. Figure 8(a) displays the magnitude curves of footplate displacement and Fig. 8(b) shows the phase angle curves. As can be seen in Fig. 8(a), the footplate displacement at frequency greater than 300 Hz was decreased when the cavity was filled with fluid at different levels. The reduction of footplate displacement varied with the fluid volume, which is similar to the results measured from temporal bones by Gan *et al.* (2006a). The FE model predicted that the amount of displacement reduction across 1000–6000 Hz, on average, was 8.3, 15.1, 20, and 20.3 dB in response to the change of fluid level from 0 to 50%, 61%, 75%, and 100% of the cavity, respectively.

The results in Fig. 9 were derived from the model without cochlear load. Figure 9(a) displays the magnitude curves

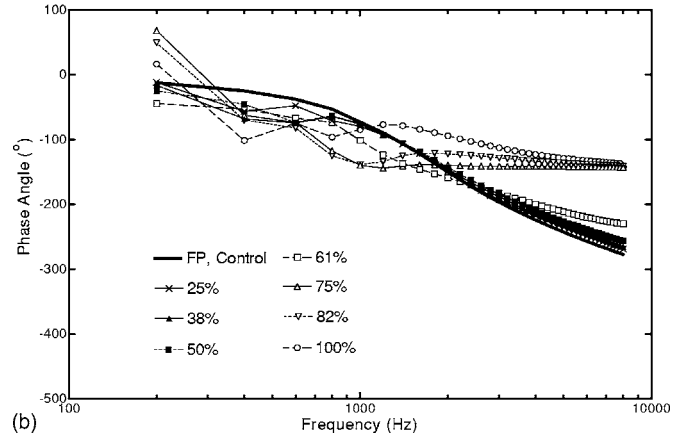
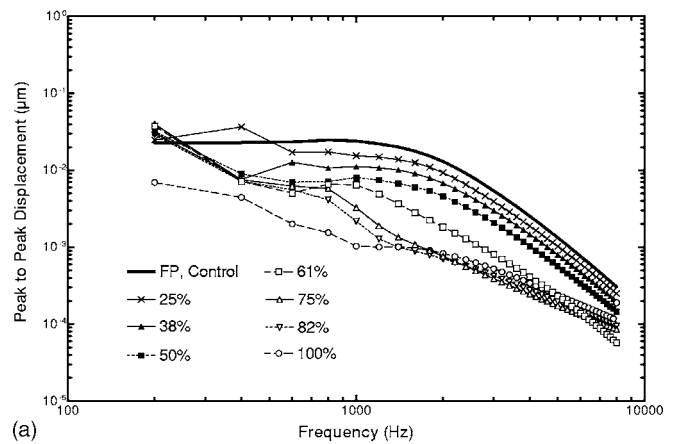


FIG. 8. FE model-predicted stapes footplate displacement curves (FP) when fluid volume in the middle ear cavity varied from 25% to 100% of cavity volume. The input sound pressure was 90 dB SPL at 2 mm away from the umbo in the ear canal. (a) Magnitude and (b) phase angle.

of footplate displacement when fluid in the middle ear was stepwise changed from zero (control) to full fill of the cavity. The footplate displacement curve of the model with intact cochlea is also included in Fig. 9(a). Comparing two control curves in Fig. 9(a), one found that release of cochlear load resulted in a 3 dB increase of the stapes footplate displacement at high frequencies ($f \geq 1000$ Hz). This is in agreement with the results measured in temporal bones by Gan *et al.* (2004, 2006a).

The curves in Fig. 9(a) demonstrate that fluid in the middle ear cavity caused reduction of the footplate displacement, which was strongly dependent on frequency as well as the amount of fluid in the cavity. This is consistent with the data measured in 11 temporal bones with open cochlea when fluid volume was increased from zero to a full middle ear cavity (Gan *et al.*, 2006a). Compared with the curves in Fig. 8(a) (with intact cochlea), the cochlear load had minor effect (about 1 to 2 dB) on reduction of the footplate movement when fluid was filled in the cavity. Similarly, the cochlear load did not affect the phase angle curves when comparing the results between Figs. 8(b) and 9(b).

Finally, the middle ear transfer function DTR, a displacement transmission ratio of the TM to footplate, derived from the model without cochlear load in response to fluid volume variation from 0 (control) to 25, 38, 50, 61, 75, 82, and 100% of cavity volume, is displayed in Fig. 10. The

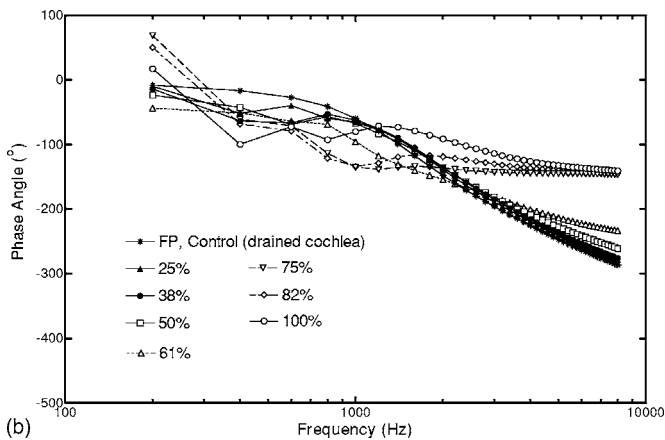
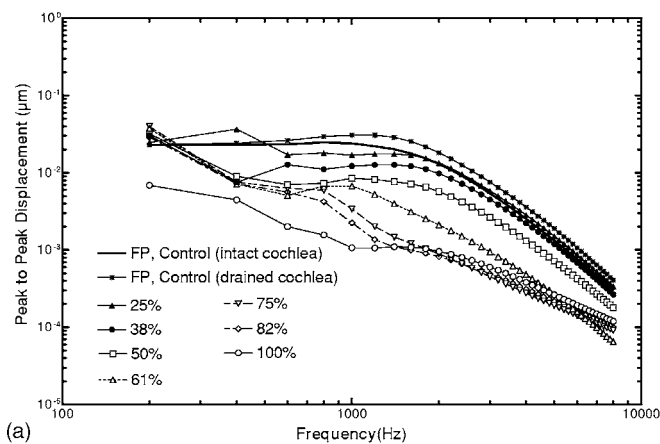


FIG. 9. The footplate displacement curves derived from the model without cochlear load (i.e., drained cochlea) when fluid volume in the cavity varied from 25% to full fill of the cavity (100%). The displacement magnitude curve obtained from the model with intact cochlea (solid line) is also included. (a) Magnitude and (b) phase angle.

DTR was not affected much by a small amount of fluid at low frequencies. However, at high frequencies the DTR curves show different behavior as fluid volume increased to over 75% in comparison with that below 61%. This prediction from the model is consistent with the experimental measurements obtained from temporal bones with open cochlea by Gan *et al.* (2006a). The DTR values around 3 observed in

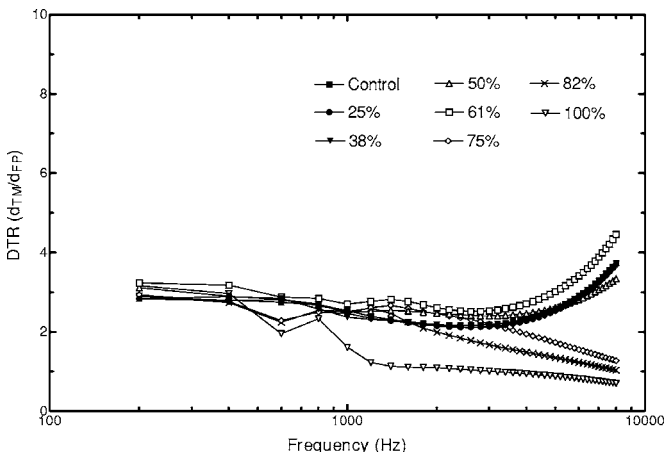


FIG. 10. FE model-predicted displacement transmission ratio (DTR) of the middle ear with open cochlea when fluid volume in the cavity changed from 0 (or control) to 25, 38, 50, 61, 75, 82, and 100% of the cavity volume.

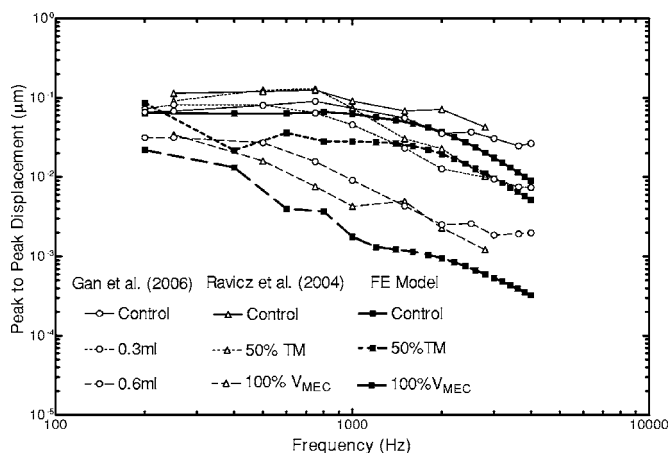


FIG. 11. Comparison of model-derived umbo displacement curves with the mean curves obtained in seven bones with intact cochlea by Gan *et al.* (2006a) and the data reported by Ravicz *et al.* (2004) when saline was introduced to the middle ear. There are three fluid levels for comparison: Control (no fluid), fluid level to umbo (0.3 ml or 50% TM), and fluid filled cavity (0.6 ml or 100% V_{MEC}).

Fig. 10 were reported by Gan *et al.* (2006a) in temporal bones. The variation of DTR with respect to fluid volume at high frequencies was also seen from both experimental and modeling results. This agreement suggests that the assumption on elastic modulus of the IS joint for the FE model was suitable to the fluid level change in the cavity.

IV. DISCUSSION

A. Comparison of results with published data

A newly constructed 3D FE model of the human ear is reported in this paper, which added another individual model of the human ear into our ear model database. The geometry differences between individual models are shown in Table I with the present FE model, previous model, and published geometry data of human ear. The main differences between this model and the previous one (Gan *et al.*, 2004) are the size of the eardrum, stapes, and middle ear cavity. The TM surface area (83.73 mm²) and height of the cone (2.421 mm) of this model are larger than the previous model of surface area 72.01 mm² and height of the cone 1.46 mm. The height of the stapes of this model is 4.18 mm, which is larger than that of the previous model (2.66 mm). The length and width of the stapes footplate of this model are also slightly larger than that of the previous model. Thus, the weight of stapes in this model is larger than the previous model. Another difference is the volume of middle ear cavity: 678 mm³ vs 455 mm³ of the previous model. These cavity values do not include the mastoid air cells. The geometry difference between ear models represents the variation of individual ears and affects the transfer function derived from the model in response to sound stimulus in the ear canal.

The TM displacement data (at the umbo) obtained from the FE model at different fluid levels (Fig. 7) were compared with the data reported by Gan *et al.* (2006a) and Ravicz *et al.* (2004) in Fig. 11. The curves measured by Gan *et al.* on seven bones (Fig. 9 of their paper) with intact cochlea at control (no saline), 0.3, and 0.6 ml of saline and the curves measured by Ravicz *et al.* on one bone (Fig. 4 of their paper)

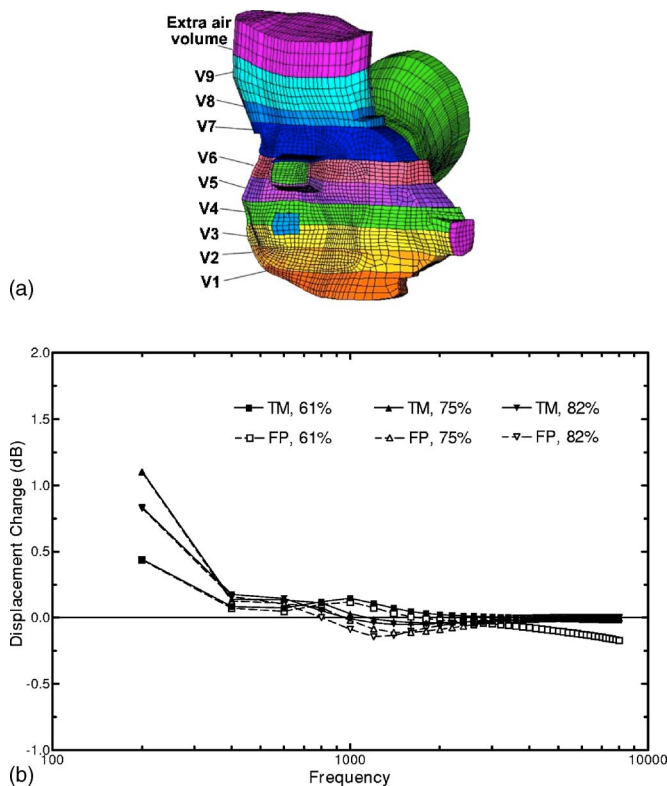


FIG. 12. (Color online) Effect of residual air volume in the middle ear on TM and footplate (FP) displacement. (A) Modified FE model with extra air volume of 112.5 mm^3 added to the top of the middle ear cavity. (B) Change of the displacement at the TM (umbo) and footplate induced by extra air volume when three fluid levels: 61%, 75%, and 82% of the cavity volume were tested.

with no saline (control), saline on 50% of TM, and 100% middle ear cavity were compared with curves obtained from the FE model. As can be seen in Fig. 11, there are some variations between modeling results and experimental data, but the model-predicted results generally agree with the data obtained by Gan and Ravicz.

To verify the influence of residual air volume in the middle ear on TM and stapes footplate movement, we modified the model by adding a volume of 112 mm^3 (i.e., about 17% of the middle ear cavity volume) into the top of the cavity or superior tympanum and calculated the TM and footplate displacements with the extra-air volume [see Fig. 12(a)]. Figure 12(b) shows the TM and footplate displacement changes in decibels due to the extra-air volume. There is no obvious difference between the change of TM or footplate displacement at frequency over 400 Hz. Both TM and footplate displacements at low frequency ($f < 1000 \text{ Hz}$) increased in three tested fluid levels: 61%, 75%, and 82% of the cavity volume when the extra-air volume was added to the model. The maximum increase of the TM and footplate displacement with extra air volume was about 1 dB at 200 Hz or below 400 Hz. These results suggest that the residual air volume in the middle ear affects transfer function of the middle ear at low frequencies when fluid is accumulated in the cavity. In other words, the FE results seem to indicate that the compliance of residual middle-ear air does not affect the TM and footplate displacements significantly.

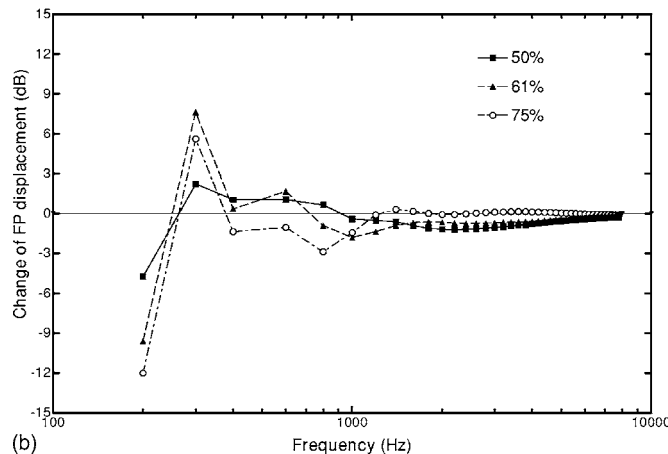
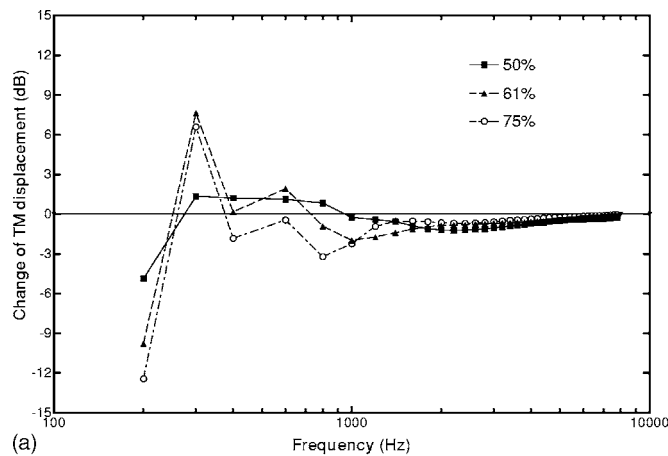


FIG. 13. Effect of fluid viscosity on the TM and footplate displacements when amount of fluid varied from 50% to 61%, and 75% of the cavity volume. (a) Change of the TM (at the umbo) displacement in decibels derived with fluid of 1000 cp viscosity relative to saline (1 cp) and (b) change of the footplate displacement in decibels derived with fluid of 1000 cp viscosity relative to saline of 1 cp.

B. Effects of fluid viscosity on TM and footplate movements

The results reported in Figs. 7–12 were obtained under the assumption that fluid in the middle ear was saline solution with viscosity of 1 cp. To investigate how the viscosity of fluid affects the TM and footplate movement, we changed the fluid viscosity to 1000 cp, which was substantially higher than that of saline, and performed the FE analysis for three fluid filling levels: 50%, 61%, and 75% of the cavity volume. Figure 13(a) shows the change of the TM displacement (in decibels) when the fluid viscosity increased to 1000 cp in reference to the saline viscosity. Figure 13(b) shows the change of the footplate displacement when fluid viscosity increased to 1000 cp with respect to 1 cp of saline solution.

As can be seen in Fig. 13, at high frequencies ($f > 1000 \text{ Hz}$) both TM and footplate displacements were reduced, but the amount of reduction was less than 2 dB as viscosity increased by 1000 times and there was not much difference between three fluid levels tested. At low frequencies ($f < 1000 \text{ Hz}$), the TM and footplate displacements varied from enhancement (particularly at 300 Hz) to reduction as fluid viscosity increased to 1000 cp. When the amount of fluid increased, the displacement reduction increased and a

maximum reduction of 12 dB at the TM and footplate were observed at the fluid level of 75% or the whole TM was under fluid. The observations reported by Ravicz *et al.* (2004) when a wide range of silicone viscosity was tested in one bone showed large variation of the TM or umbo velocity with fluid viscosity at frequency of 100–300 Hz. A reduction of 20 dB at 200 Hz was measured by Ravicz when the silicone viscosity was 1000 cp. There was no substantive difference among the TM velocity change when fluid of different viscosity was used at frequency between 300 and 3000 Hz.

As the middle ear is filled with fluid, the maximum displacements of the TM and footplate move to a lower frequency or the mass-dominated region extends to a lower frequency (Figs. 7–9). Note that as the driving frequency approaches the resonant frequency, the maximum effect of damping on vibration magnitude occurs. The damping matrix of the system is a function of fluid viscosity only as expressed in Eq. (3). Therefore, the present results imply that the effect of viscosity on TM and footplate displacement is dependent on the resonant vibration of the structure-air-fluid (liquid) coupled system. Middle ear fluid results in a greater reduction of vibration amplitudes of the TM and footplate near the resonant frequency than that at other frequencies. This suggests that the influence of fluid viscosity on the TM and footplate motion is both frequency and fluid volume dependent even at low frequencies.

C. Effects of TM and IS joint Young's modulus on TM and footplate movements

The effect of fluid on mechanical properties of the soft tissue in the middle ear remains unclear, and there is no analytical model of how middle ear fluid causes the Young's modulus and structure changes in the middle ear. However, using the FE model in this study, the influence of Young's modulus of the TM and IS joint on middle ear transfer function were estimated.

The effect of TM material properties on TM and footplate movement was first examined when the amount of fluid in the middle ear cavity reached 50%, 61%, and 75% of the cavity volume, which represents the fluid level covering the umbo, stapes footplate, and whole TM, respectively. Figure 14(a) and 14(b) show the change of TM and footplate displacements (in decibels) at three fluid levels when the Young's modulus of the TM (both radial and circumferential directions) were increased by 1.5 times of the base values listed in Table II. There were variations of the TM or footplate displacement change at low frequencies ($f < 800$ Hz, especially at 300 Hz) when the TM became stiffer or having higher Young's modulus. At high frequencies ($f > 3000$ Hz), there was no substantive difference between different fluid levels when the TM became stiffer and a slight increase of the TM and footplate displacement was observed.

Between 800 Hz and 3000 Hz, the TM or footplate displacement was reduced by a limit of 2.5 dB when the amount of fluid reached 50%. When fluid level reached 61%, there was no reduction of the TM or footplate displacement and a maximum increase of the TM or footplate displacement (1.5–2 dB) was observed at 1200 Hz. The TM and footplate

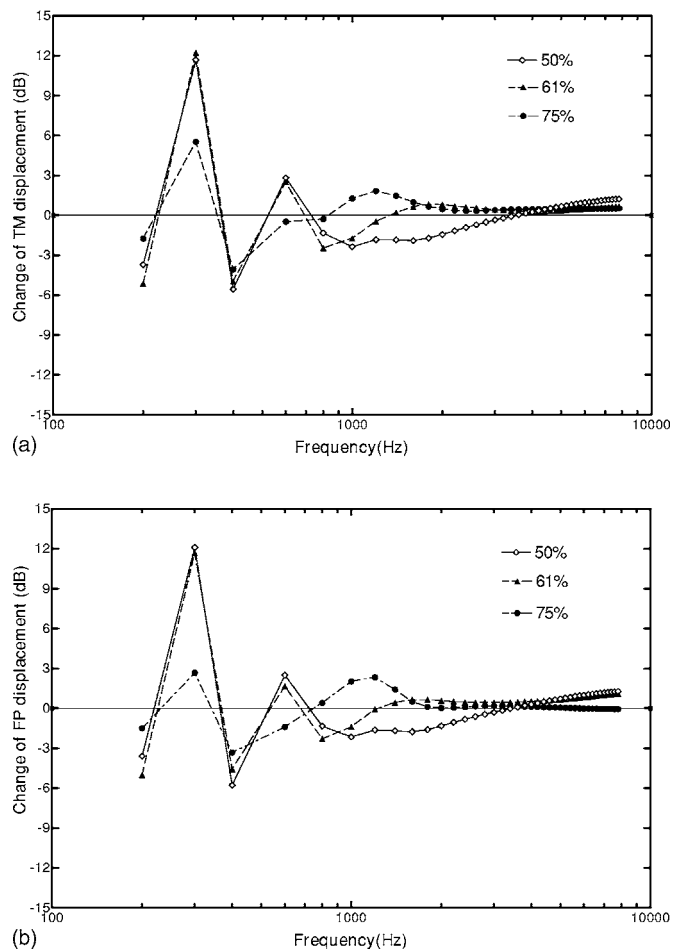


FIG. 14. Effect of TM material properties on TM and footplate displacements when amount of fluid varied from 50% to 61%, and 75% of the cavity volume. (a) Change of the TM (at the umbo) displacement derived with higher Young's modulus of the TM relative to normal modulus of the TM and (b) change of the footplate displacement derived with higher Young's modulus of the TM relative to normal modulus of the TM.

displacements were changed from reduction (1.5 dB) to enhancement (0.8 dB) as frequency increases from 800 to 3000 Hz when the fluid level reached 75%. The results in Fig. 14 indicate that the effect of TM mechanical properties on middle ear transfer function vary with frequency and amount of fluid in the middle ear cavity, especially at frequency below 3000 Hz.

The effect of the IS joint material properties on middle ear transfer function was examined when Young's modulus of the joint was maintained as a constant value of 0.6 MPa (Case 1) or a variable changing with the amount of fluid in the middle ear as described in Sec. II B (Case 2). Figure 15 displays the DTR curves predicted by the FE model for these two cases. When the Young's modulus of the IS joint was not changing with fluid level (Case 1), the DTR curve shows significant difference from the results measured in temporal bones (Fig. 15C of Gan *et al.*, 2006a). A large variation of DTR between three middle ear fluid levels was observed across the frequency range of 200–8000 Hz. However, when the Young's modulus of the IS joint was varied to 1.2, 2.8, and 8.0 MPa corresponding to fluid level of 50%, 61%, and 75%, respectively (Case 2), the DTR value was between 2 and 4 over 200–6000 Hz, which is in a good agreement with

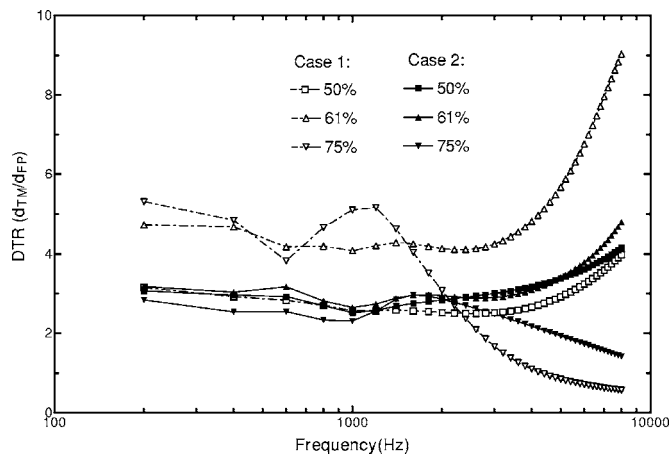


FIG. 15. FE model-predicted DTR curves of the middle ear with three fluid levels in the middle ear: 50%, 61%, and 75% of the cavity volume. The Young's modulus of the IS joint was defined as different values in case 1 and case 2. Case 1: The Young's modulus of the joint was a constant of 0.6 MPa. Case 2: The Young's modulus was increased from 0.6 MPa to 1.2, 2.8, and 8.0 MPa when the fluid level in the cavity reached 50%, 61%, and 75% of the cavity volume, respectively.

the temporal bone experimental results reported by Gan *et al.* (2006a). Thus, the assumption of the IS joint Young's modulus change due to the middle ear fluid level was reasonably well supported by experimental data on variation of middle ear transfer function measured in temporal bones. One possible explanation for nonlinear change of the IS joint stiffness in response to larger fluid volume in the cavity is the degree of contact between the lenticular plate and stapes head. This contact is caused by displacements of the TM and ossicular chain due to the hydrostatic pressure of fluid in the middle ear.

We have shown the effect of stiffness of the TM and IS joint on middle ear transfer function when fluid was accumulated in the cavity. However, more studies need to be further conducted in the FE model regarding the relationship between the stiffness change of the TM and other middle ear soft tissues and the effusion level in the middle ear.

V. CONCLUSION

A 3D FE model of a left human ear was reported in this paper based on histological section images of a human temporal bone. The multiple fields including air in the canal and middle ear cavity, fluid in the cavity, and middle ear structures (i.e., bones and soft tissues) are coupled in the model to simulate various stages of the otitis media with effusion. The movements of the TM and stapes footplate or the transfer function of the middle ear with variable fluid levels in the middle ear are derived from the model and compared with the data measured in human temporal bones. The results indicate that this model has the capability to extend FE analysis into pathological ears such as otitis media with visualized fluid-air interfaces inside the middle ear structures.

ACKNOWLEDGMENTS

The research contributions of Caleb Jones, B.S., Danny Ferguson, B.S., and Tao Cheng, Ph.D., at the University of

Oklahoma and Robert Dyer, Jr., M.D. at the Hough Ear Institute are gratefully recognized. This work was supported by Grant Nos. NIH/NIDCD R01DC006632 and NSF/CMS 0510563.

- Aibara, R., Welsh, J. T., Puria, S., and Goode, R. L. (2001). "Human middle-ear sound transfer function and cochlear impedance," *Hear. Res.* **152**, 100–109.
- Donaldson, J. A., and Miller, J. M. (1973). "Anatomy of the Ear," in *Basic Sciences and Related Disciplines, Otolaryngology* (Saunders, Philadelphia), Vol. 1, pp. 75–110.
- Gan, R. Z., Cheng, T., and Wood, M. W. (2007b). "Acoustic-structural coupled finite element analysis for sound transmission in human ear-Middle ear transfer function," in *Middle Ear Mechanics in Research and Otolaryngology*, A. Huber and A. Eiber (Eds.) (World Scientific Publishing, Singapore), pp. 205–211.
- Gan, R. Z., Dai, C., and Wood, M. W. (2006a). "Laser interferometry measurements of middle ear fluid and pressure effects on sound transmission," *J. Acoust. Soc. Am.* **120**, 3799–3810.
- Gan, R. Z., Feng, B., and Sun, Q. (2004). "Three-dimensional finite element modeling of human ear for sound transmission," *Ann. Biomed. Eng.* **32**, 847–859.
- Gan, R. Z., Nakmali, D., and Wood, M. W. (2007a). "Effects of middle ear suspensory ligaments on acoustic-mechanical transmission in human ear," in *Middle Ear Mechanics in Research and Otolaryngology*, A. Huber and A. Eiber (Eds.) (World Scientific Publishing, Singapore), pp. 212–221.
- Gan, R. Z., Sun, Q., Dyer, R. K., Chang, K.-H., and Dormer, K. J. (2002). "Three dimensional modeling of middle ear biomechanics and its application," *Otol. Neurotol.* **23**, 271–280.
- Gan, R. Z., Sun, Q., Feng, B., and Wood, M. W. (2006b). "Acoustic-structural coupled finite element analysis for sound transmission in human ear-Pressure distributions," *Med. Eng. Phys.* **28**, 395–404.
- Herrmann, G., and Liebowitz, H. (1972). "Mechanics of bone fractures," in *Fracture: An advanced treatise*, edited by H. Liebowitz (Academic, New York), Vol. 7, pp. 772–840.
- Ihlenbury, F. (1998). *Finite Element Analysis of Acoustic Scattering* (Springer, New York).
- Kirikae, I. (1960). *The Structure and Function of the Middle Ear* (University of Tokyo Press, Tokyo).
- Kringlebotn, M., and Gundersen, T. (1988). "Network model for the human middle ear," *Scand. Audiol.* **17**, 75–85.
- Merchant, S. N., Ravicz, M. E., and Rosowski, J. J. (1996). "Acoustic input impedance of the stapes and cochlea in human temporal bones," *Hear. Res.* **97**, 30–45.
- Prendergast, P. J., Ferris, P., Rice, H. J., and Blayney, A. W. (1999). "Vibro-acoustic modeling of the outer and middle ear using the finite-element method," *Audiol. Neuro-Otol.* **4**, 185–191.
- Ravicz, M. E., Rosowski, J. J., and Merchant, S. N. (2004). "Mechanisms of hearing loss resulting from middle-ear fluid," *Hear. Res.* **195**, 103–130.
- Rosowski, J. J., and Merchant, S. N. (1995). "Mechanical and acoustic analysis of middle ear reconstruction," *Am. J. Otol.* **16**, 486–497.
- Sun, Q., Gan, R. Z., Chang, K. H., and Dormer, K. J. (2002). "Computer-integrated finite element modeling of human middle ear," *Biomechan. Model. Mechanobiol.* **1**, 109–122.
- Von Békésy, G. (1960). *Experiments in Hearing* (McGraw-Hill, New York).
- Wada, H., Koike, T., and Kobayashi, T. (1997). "Three-dimensional finite-element method (FEM) analysis of the human middle ear," in *Research and Otolaryngology: Proceedings of the International Workshop on Middle Ear Mechanics in Research and Otolaryngology*, edited by K. B. Hüttenbrink (Dresden University Press, Dresden, Germany), pp. 76–80.
- Wada, H., and Metoki, T. (1992). "Analysis of dynamic behavior of human middle ear using a finite method," *J. Acoust. Soc. Am.* **92**, 3157–3168.
- Wever, E. G., and Lawrence, M. (1954). *Physiological Acoustics* (Princeton University Press, Princeton, NJ).
- Zienkiewicz, O. Z., and Bettess, P. (1978). "Fluid-structure dynamic interaction and wave forces. An introduction to numerical treatment," *Int. J. Numer. Methods Eng.* **13**, 1–16.
- Zwislocki, J. (1962). "Analysis of the middle ear function. I. Input impedance," *J. Acoust. Soc. Am.* **34**, 1514–1523.

Distortion product otoacoustic emissions: Cochlear-source contributions and clinical test performance

Tiffany A. Johnson,^{a)} Stephen T. Neely, Judy G. Kopun, Darcia M. Dierking, Hongyang Tan, Connie Converse, Elizabeth Kennedy, and Michael P. Gorga
Boys Town National Research Hospital, 555 N. 30th Street, Omaha, Nebraska 68131

(Received 18 May 2007; revised 11 September 2007; accepted 14 September 2007)

It has been proposed that the clinical accuracy of distortion product otoacoustic emissions (DPOAEs) is affected by the interaction of distortion and reflection sources contributing to the response. This study evaluated changes in dichotomous-decision test performance and threshold-prediction accuracy when DPOAE source contribution was controlled. Data were obtained from 205 normal and impaired ears with L_2 ranging from 0 to 80 dB SPL and $f_2=2$ and 4 kHz. Data were collected for control conditions (no suppressor, f_3) and with f_3 presented at three levels that previously had been shown to reduce the reflection-source contribution. The results indicated that controlling source contribution with a suppressor did not improve diagnostic accuracy (as reflected by relative operating characteristic curve area) and frequently resulted in poorer test performance compared to control conditions. Likewise, correlations between DPOAE and behavioral thresholds were not strengthened when using the suppressors to control source contribution. While improvements in test accuracy were observed for a subset of subjects (normal ears with the smallest DPOAEs and impaired ears with the largest DPOAEs), the lack of improvement for the larger, unselected subject group suggests that DPOAEs should be recorded in the clinic without attempting to control the source contribution with a suppressor.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799474]

PACS number(s): 43.64.Jb [BLM]

Pages: 3539–3553

I. INTRODUCTION

Current theories of otoacoustic emission (OAE) generation suggest that there are multiple mechanisms by which otoacoustic emissions are generated (e.g., Shera and Guinan, 1999; Shera, 2004). These mechanisms include nonlinear distortion and coherent reflection. Different types of OAEs may be dominated by either one or both mechanisms. For example, as typically recorded in the ear canal, distortion product otoacoustic emissions (DPOAEs) are believed to include contributions from both a nonlinear-distortion source and a coherent-reflection source (e.g., Talmadge *et al.*, 1998, 1999; Mauermann *et al.*, 1999a; Shera and Guinan, 1999; Stover *et al.*, 1999; Konrad-Martin *et al.*, 2001, 2002).

In the case of DPOAEs, the components from these two sources are generated at different locations on the basilar membrane. The nonlinear-distortion component is believed to arise near the f_2 place, where the primary nonlinear interaction between f_1 and f_2 occurs. The nonlinear interaction of the two primaries produces the $2f_1-f_2$ distortion product that propagates in two directions: basally toward the ear canal and apically toward its tonotopic place on the basilar membrane, where it will be reflected by the coherent-reflection mechanism (Zweig and Shera, 1995). DPOAEs, therefore, contain contributions from disparate locations on the basilar membrane. Furthermore, the phases of the nonlinear-distortion and the coherent-reflection components rotate at different rates with respect to one another as a function of

frequency. This difference in the rate of phase rotation with frequency results in the two components interacting with each other in a manner that alternates between constructive and destructive interference and produces a quasisinusoidal pattern of maxima and minima in DPOAE level as frequency is varied in small increments. The pattern of alternating maxima and minima is typically referred to as DPOAE fine structure or microstructure. While the interaction of the distortion and reflection components is believed to be the primary mechanism responsible for DPOAE fine structure, periodicity in the reflection component alone also has been observed (e.g., Kalluri and Shera, 2001; Mauermann and Kollmeier, 2004; Wilson and Lutman, 2006). The periodicity in the reflection component may contribute to DPOAE fine structure, as well as fine structure observed in other otoacoustic emission types (e.g., stimulus-frequency OAEs, transient-evoked OAEs) and in behavioral thresholds. In the case of DPOAEs, the fine structure in the response is due primarily to the interaction of the distortion and reflection components, with the inherent periodicity in the reflection component adding additional variability to the measure.

The idea that DPOAEs contain contributions from two different cochlear locations generated by two different mechanisms has led a number of investigators to speculate that DPOAE test performance may be affected by the interaction of these sources (e.g., Heitmann *et al.*, 1998; Shera and Guinan, 1999; Shaffer *et al.*, 2003; Mauermann and Kollmeier, 2004; Shera, 2004). This hypothesis can be understood by considering how DPOAE fine structure varies across ears and across stimulus levels. The frequency at which a given maximum or minimum occurs is essentially

^{a)}Electronic mail: tiffany-johnson@ku.edu

randomly distributed across ears. For example, when $f_2 = 2$ kHz, one ear with normal hearing may produce a minimum DPOAE level, while another normal-hearing ear might produce a maximum level, and a third ear may be intermediate between maximum and minimum levels. These differing patterns of fine structure might lead to the erroneous conclusion that the normal-hearing ear producing a minimum at 2 kHz is impaired. Additionally, the frequencies at which maxima and minima occur within an individual ear may shift as level is increased (e.g., He and Schmiedt, 1993, 1997; Heitmann *et al.*, 1996; Johnson *et al.*, 2006b). This shift will influence the slope of the DPOAE input/output (I/O) function and, therefore, may not only affect predictions of auditory status (i.e., normal versus impaired), but also may affect the accuracy with which behavioral thresholds can be predicted from DPOAE I/O functions (e.g., Boege and Janssen, 2002; Gorga *et al.*, 2003; Oswald and Janssen, 2003). Indeed, Mauermann and Kollmeier (2004) report less variability in thresholds predicted from DPOAE I/O functions across frequency when the reflection-source contribution was reduced or eliminated, although they do not describe the correspondence between these predicted thresholds and measured behavioral thresholds.

The speculation that multiple-source contributions and the resulting fine structure affect DPOAE test performance has both intuitive and theoretical appeal; however, this hypothesis has not been directly tested. There is indirect evidence that source interactions and the resulting fine structure may not play a large role in limiting DPOAE test performance. For example, Dhar and Shaffer (2004) and Shaffer and Dhar (2006) failed to find a statistically significant correlation between behavioral threshold and DPOAE level for moderate-level stimuli, even when attempting to reduce or eliminate the contribution from the reflection source. However, these data were collected in a small group of normal-hearing subjects and, therefore, are not well suited to testing hypotheses regarding the influence of source contribution on test performance, which would require data for both normal and impaired ears. Additionally, both Dhar and Shaffer (2004) and Johnson *et al.* (2006b) reported that DPOAE fine structure was not observed in the majority of normal-hearing ears in the vicinity of 4 kHz, suggesting that it is unlikely that uncontrolled source interaction plays a major role in limiting test performance when testing at or around 4 kHz. While the Dhar and Shaffer (2004), Johnson *et al.* (2006b), and Shaffer and Dhar (2006) data cast doubt on the hypothesis that fine structure contributes to diagnostic errors, they do not directly address the question of whether uncontrolled source interaction influences DPOAE test performance. Additional data from both normal and impaired ears are necessary to answer this question.

There are several methods by which the sources contributing to the DPOAE can be controlled. In one approach, DPOAE responses are recorded for many closely spaced frequencies. The response is then transformed to the time domain by computing an inverse fast Fourier transform (IFFT). The long-latency components (thought to be generated by the reflection source) are mathematically eliminated from the response and the response is then converted back into a fre-

quency representation by computing a fast Fourier transform (FFT). This method, often referred to as time windowing, has been shown to be effective at eliminating the fine structure in DPOAE responses (e.g., Stover *et al.*, 1996b; Kalluri and Shera, 2001; Knight and Kemp, 2001; Konrad-Martin *et al.*, 2001; Mauermann and Kollmeier, 2004), and represents the gold-standard approach for restricting the DPOAE source contribution to the distortion source. While effective at separating source contributions, this approach is not suitable for clinical applications because it is necessary to record DPOAE responses for many closely spaced frequencies in order to have sufficient precision to compute the IFFT; the associated increase in test time makes this approach impractical in the clinic.

An alternate approach, selective suppression, involves presenting a third tone, a suppressor, in combination with the primary stimuli used to elicit the response. The suppressor is chosen to have a frequency close to the $2f_1-f_2$ frequency in order to reduce or eliminate the contribution from the reflection source. With appropriately chosen suppressor levels, this technique also has been shown to reduce fine structure in DPOAE responses (e.g., Heitmann *et al.*, 1998; Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001; Johnson *et al.*, 2006b). While the suppression technique has been shown to produce more variable reductions of fine structure compared to time windowing (e.g., Kalluri and Shera, 2001; Johnson *et al.*, 2006b), it is possible to implement this technique under clinical conditions because the suppressor and primary stimuli are presented simultaneously, and, therefore, the technique does not require additional test time. When making DPOAE measurements with neonates, infants, and young children (clinical populations with whom DPOAE testing is frequently used), limiting test time is important. Because the suppression technique can be implemented without increasing test time, it is ideally suited for use when examining the influence of cochlear-source contribution on DPOAE test performance. The application of the suppression technique requires knowledge of the combinations of stimulus and suppressor levels that result in optimal effects, which may also depend on stimulus frequency.

In a previous study, we investigated the effect of a broad range of suppressor levels on DPOAE fine structure for stimulus levels ranging from 20 to 80 dB SPL (Johnson *et al.*, 2006b). From this work, we were able to determine the suppressor levels resulting in optimal reduction of fine structure for each stimulus-level condition. Optimal suppressor level was defined as the suppressor level that resulted in the maximum reduction in fine structure with minimum influence on overall level. The purpose of the present study was to evaluate changes in DPOAE test performance and threshold-prediction accuracy when optimal suppressors were used to limit DPOAE cochlear-source contribution. We hypothesized that auditory status would be more accurately identified when cochlear-source contribution was controlled (i.e., when the suppressor was presented in conjunction with the primary stimuli) because the variability in DPOAE level resulting from fine structure would be eliminated. It was also hypothesized that behavioral threshold would be more accurately predicted from DPOAE I/O functions when cochlear-

source contribution was restricted to the distortion source because nonmonotonicities and/or other irregularities in the I/O functions would be reduced as well.

We have chosen the suppression technique to explore these questions because, as previously described, it can be implemented under clinical conditions where maximizing test efficiency is important. Additionally, while we acknowledge that fine structure exists in behavioral thresholds, in addition to OAE responses (e.g., Elliott, 1958; Long, 1984; Zwicker and Schloth, 1984; Zwicker, 1986; Long and Tubis, 1988a, b; Furst *et al.*, 1992; Kapadia and Lutman, 1999; Lutman and Deeks, 1999; Horst *et al.*, 2002; Mauermann *et al.*, 2004), it typically is not measured during clinical assessments where only fixed octave and interoctave frequencies are tested. Therefore, because our primary goal is evaluating the influence of cochlear-source contribution on the clinical test performance of DPOAE responses, we have not focused on the influence of behavioral-threshold fine structure on diagnostic errors when DPOAEs are measured.

II. METHODS

A. Subjects

Data were obtained from 98 ears of 56 subjects with normal hearing and 107 ears of 60 subjects with sensorineural hearing losses. Normal hearing was defined as pure-tone behavioral thresholds ≤ 20 dB HL (re: ANSI, 1996) for the octave and interoctave frequencies between 0.25 and 8 kHz. For purposes of counting the number of normal versus impaired ears participating, an ear was considered hearing impaired if one or more pure-tone behavioral thresholds were > 20 dB HL for the same frequency interval. The majority of these ears (102/107) had behavioral thresholds exceeding 20 dB HL at 2 and/or 4 kHz. A sensorineural hearing loss was defined as air-bone gaps < 15 dB at the octave frequencies of 0.5 through 4 kHz and normal 226 Hz tympanograms. Please note, however, that when evaluating the accuracy of the dichotomous decision, the classification of normal versus impaired was made on a frequency-by-frequency basis (see the following). Both normal-hearing and hearing-impaired subjects were required to have normal 226 Hz tympanograms prior to each data-collection session.

Efforts were made to recruit subjects whose hearing losses were primarily sloping in configuration. The reason for this emphasis was, in part, due to the observation that DPOAE responses from ears with sloping high-frequency hearing losses are more likely to include contributions from the reflection source (Konrad-Martin *et al.*, 2002; Stover *et al.*, 1999) and, thus, corresponding fine structure (Mauermann *et al.*, 1999b). Additionally, in order to maximize the likelihood that the impaired ears would produce several points on the DPOAE I/O function, only ears with pure-tone behavioral thresholds ≤ 60 dB HL at 2 and 4 kHz were included.

B. Procedures

1. Equipment and calibration

Custom-designed software (EMAV; Neely and Liu, 1993) was used for data collection. This software controlled

a 24 bit soundcard (CardDeluxe, Digital Audio Labs) housed in a PC. Separate channels of the soundcard were used to generate the two primary tones (f_1 and f_2), which were then mixed acoustically in the ear canal. When a suppressor (f_3) was presented along with the primaries, it was generated on the same channel as f_2 . In the experimental paradigm used for data collection, f_2 and f_3 have a larger frequency separation than f_1 and f_3 . By generating f_2 and f_3 on the same channel, we ensured that intermodulation distortion products due to interactions between f_2 and f_3 did not occur at the same frequency as the DPOAE response ($2f_1 - f_2$).

An ER-10C (Etymotic Research) probe-microphone system was used for stimulus presentation and response recording. Because the ER-10C had been modified to remove 20 dB of attenuation from each channel, we were able to achieve primary and suppressor levels as high as 80 dB SPL in each subject's ear canal.

Prior to data collection, stimuli were calibrated (in the ear canal) in sound pressure level at the plane of the ER-10C probe. Concerns have been raised regarding *in situ* pressure calibration (Siegel, 1994; Siegel and Hirohata, 1994; Neely and Gorga, 1998), although similar concerns apply to all pressure-calibration methods. While it is the current standard of practice for DPOAE measures, *in situ* pressure calibration may introduce variability into the measure as a result of the calibration, particularly at 4 kHz. We acknowledge this potential source of error and have attempted to address this by collecting data on subjects ($N=205$) with a range of different ear canal dimensions.

Measurement-based stopping rules were used during data collection, such that averaging continued until the noise level was ≤ -25 dB SPL or 32 s of artifact-free averaging was achieved, whichever occurred first. Measurement-based stopping rules may increase test time but have the advantage of producing more consistent and lower noise levels across subjects because averaging time is increased whenever the noise level exceeds -25 dB SPL. Noise level was estimated from the $2f_1 - f_2$ frequency bin. This was accomplished by alternately storing 0.25 s samples of the recorded response in one of two buffers. The buffers were summed to provide an estimate of the DPOAE level and were subtracted to provide an estimate of the noise level in the same frequency bin.

2. DPOAE I/O functions

DPOAE I/O functions were recorded from each subject with $f_2=2$ and 4 kHz. At each f_2 , L_2 ranged from 0 to 80 dB SPL in 5 dB increments. The relationship between L_1 and L_2 and f_1 and f_2 was specified according to the following (Johnson *et al.*, 2006a):

$$L_1 = 80 + 0.137 \log_2(18/f_2)(L_2 - 80), \quad (1)$$

$$f_2/f_1 = 1.22 + \log_2(9.6/f_2)(L_2/415)^2. \quad (2)$$

We chose the stimulus conditions described by Eqs. (1) and (2) (i.e., the L_1 for each L_2 and the f_2/f_1 for each f_2) because they result in the largest average L_{dp} for normal-hearing subjects (Johnson *et al.*, 2006a).

For each f_2 , four different conditions were evaluated. In the control condition, the I/O function was recorded with no

TABLE I. Optimal suppressor levels for reducing DPOAE fine structure.

L_2 (dB SPL)	$f_2=2$ kHz	$f_2=4$ kHz
	C	C
20	40	45
30	35	30
40	30	30
50	30	20
60	20	15
70	15	15
80	10	15

suppressor present. In the three experimental conditions, a suppressor (f_3), whose frequency was 16 Hz below the $2f_1-f_2$ frequency, was presented at one of three levels. The three suppressor levels were chosen based on results described previously (Johnson *et al.*, 2006b) so as to assure the likelihood that the response coming from the reflection source was suppressed. More specifically, the suppressor that was optimal at reducing fine structure in the Johnson *et al.* (2006b) average data was used as well as suppressors 10 dB above and 10 dB below this optimal level. This range of suppressor levels was used as a way of attempting to assure that at least one suppressor level was likely to be optimal in this new group of subjects.

When exploring the suppressor-level space (Johnson *et al.*, 2006b), the suppressor level (L_3) was set according to

$$L_3 = 0.75L_2 + C. \quad (3)$$

A broad range of C values were used for each L_2 . The optimal values of C for each f_2 and L_2 (as determined from Johnson *et al.*, 2006b) are shown in Table I. When determining the optimal value of C from the Johnson *et al.* (2006b) data, three rules were applied. Priority was given to suppressor levels (and corresponding C values) where there was a minimum in the magnitude of the fine structure *and* a minimal reduction in the overall level of the DPOAE. In some cases, consideration also was given to the suppressor level producing the smallest intersubject variability. In addition to these rules, the C value was required to decrease monotonically or to remain constant as L_2 increased. The C values shown in Table I were those meeting all of the above-mentioned conditions. The three suppressor levels used in the present study correspond to the L_3 values resulting from the C values in Table I, which was the base condition (C0). Data also were collected with C set 10 dB above (Cp10) and 10 dB below (Cm10) this value, with the constraint that L_3 never exceeded 80 dB SPL. Thus, the suppressor levels were chosen as the previously determined optimal levels for each $L_2 \pm 10$ dB. Recording the I/O functions in the control condition and with the three different suppressor levels allowed us to evaluate the influence that the suppressors had on the accuracy of the dichotomous-decision task (i.e., normal or impaired hearing) as a function of L_2 . It also enabled us to evaluate the accuracy with which behavioral threshold was predicted from the DPOAE I/O function for each of the four conditions, using previously described procedures (e.g., Boege and Janssen, 2002; Gorga *et al.*, 2003).

3. DPgrams

In addition to recording DPOAE I/O functions for each ear, DPgrams were recorded on a subset of ears recruited to return for additional data collection in efforts to understand why diagnostic errors occurred in some subjects. The 28 normal ears producing the smallest responses and the 23 impaired ears producing the largest responses participated in this additional data-collection effort. These ears were recruited based on the assumption that these would be ears for which diagnostic errors were most likely. DPgrams were recorded while f_2 was varied in 1/64 octave steps for the 1/3 octave interval centered at either 2 or 4 kHz. L_2 was fixed at 50 dB SPL for the interval surrounding 2 kHz and at 35 dB SPL for the 4 kHz interval. These level conditions were selected for more detailed exploration because they resulted in the best performance (defined as the largest area under the relative operating characteristic curve, AROC) in the control condition. The DPgrams were recorded in four different suppressor conditions: the control condition, where no suppressor was present, and suppressor conditions, in which the suppressor was simultaneously presented at each of the three above-described suppressor levels (C0, Cp10, Cm10).

4. General test procedures

A sound-treated room was used for data collection. During data collection, subjects were seated in a reclining chair and slept, read quietly, or watched a silent, captioned movie. The DPOAE I/O-function data typically were collected during two to three 2 h sessions. The DPgram data collection required an additional session typically lasting no more than 1 or 2 h. The amount of data-collection time per subject was variable because of our use of measurement-based stopping rules. Additionally, when recording DPgrams, some subjects had ears that met our inclusion criteria at both 2 and 4 kHz. This resulted in a longer test session than when testing a subject on whom data were collected at only one f_2 frequency. When recording both the DPOAE I/O functions and the DPgrams, the control (no suppressor) condition was always tested first. In-the-ear calibration of level was repeated before testing each condition. In this way, we were able to monitor the stability of the probe fit over the entire test session and to make adjustments if the fit changed.

5. Discrete cosine transform

As will be described in the following, the discrete cosine transform (DCT) (Rao and Yip, 1990) was computed on the DPgram data. The DCT was used to quantify the effectiveness of the suppressors with regard to how completely they reduced the fine structure in the DPgram, an approach we have used previously (Johnson *et al.*, 2006b). The DCT is similar in concept to the FFT and is essentially a FFT for which the components with sine-function symmetry have been dropped. It provides a quantification of the overall level of the response (the DC component of the DCT) as well as the levels of the higher "frequency" (cycles/octave) components that ride on the DCT component. The elimination of sine-symmetry components reduces frequency-domain windowing artifacts. We interpret the DC component of the DCT

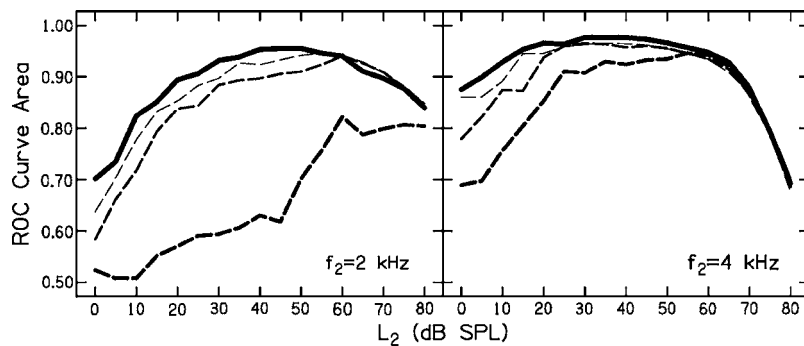


FIG. 1. ROC curve area as a function of L_2 (dB SPL). Results for $f_2=2$ and 4 kHz are shown in the left and right panels, respectively. The parameter in each panel is test condition, with the control condition represented by the thick, solid line and the suppression conditions indicated by the dashed lines. Increasing dash thickness indicates increasing suppressor level.

(or overall level of the response) as an estimate related to the level of the distortion-source component. Phase information is not used in the DCT analysis. The magnitude of the “high frequency” coefficients (coefficients between 1.5 and 22.5 cycles per octave) provides an estimate related to the amount of the fine structure that is present in each condition and, thus, can be used to evaluate the extent to which the suppressors have reduced the fine structure¹. The DCT analysis method was selected to quantify our visual assessment of fine structure based on plots of DPOAE level versus frequency.

III. RESULTS

A. Accuracy of the dichotomous decision

The influence of controlling the DPOAE source contribution on test performance was quantified using clinical decision theory (CDT) (Swets and Pickett, 1982; Swets, 1988). This approach, which is well suited to assessing the accuracy with which diagnostic tests provide dichotomous classifications (i.e., normal or impaired hearing in the present case), has been used previously to evaluate DPOAE test performance (e.g., Gorga *et al.*, 1993, 1997, 1999, 2000, 2005; Kim *et al.*, 1996; Stover *et al.*, 1996a; Dorn *et al.*, 1999). In these applications of CDT, audiometric threshold served as the gold standard to which the results from the experimental measurement (typically DPOAE level) were compared. As in previous applications of CDT to assess DPOAE test performance, behavioral thresholds were used in the present study to classify an ear as either normal (thresholds ≤ 20 dB HL) or impaired (thresholds > 20 dB HL). This classification was made on a frequency-specific basis. In other words, if an ear had a behavioral threshold ≤ 20 dB HL at 2 kHz but > 20 dB HL at 4 kHz, it would be considered normal at 2 kHz and impaired at 4 kHz. By computing hit rates (i.e., correct identification of hearing loss or sensitivity) and corresponding false-alarm rates for all possible DPOAE levels, a complete description of test performance was obtained. The results from the DPOAE I/O function measurements were summarized in the form of relative operating characteristic (ROC) curves, which are plots of hit rate as a function of false-alarm rate. The area under the ROC curve (A_{ROC}) provided a single estimate of test accuracy. Possible A_{ROC} values range from 0.5, which indicates chance performance, to 1.0, which indicates perfect test performance, where the hit rate is 100% for all false-alarm rates. In the present study, ROC curves were

constructed for each of the four test conditions (control, C0, Cp10, and Cm10) at each f_2 and L_2 , and A_{ROC} was computed for every condition.

Figure 1 plots A_{ROC} as a function of L_2 based on data collected from the entire sample of subjects. Results for $f_2=2$ and 4 kHz are plotted in the left and right panels, respectively. The parameter in each panel is test condition. The thick, solid line represents results for the control condition. The dashed lines represent results for the three suppressor conditions, with increasing line thickness indicating increasing suppressor level, from a minimum for Cm10 (thinnest dashed line) to a maximum for Cp10 (thickest dashed line). For virtually every condition (the exception being the highest suppressor level, Cp10, when $f_2=2$ kHz), A_{ROC} increases with L_2 , reaches a maximum over a range of moderate stimulus levels, and then decreases with further increases in L_2 . This pattern of results was observed for both $f_2=2$ and 4 kHz; this dependence on L_2 has been observed previously (e.g., Stover *et al.*, 1996a). When $f_2=2$ kHz, the largest A_{ROC} was observed for the control condition when $L_2=50$ dB SPL. The largest A_{ROC} when $f_2=4$ kHz occurred when $L_2=35$ dB SPL in the control condition, although, at both f_2 frequencies, there was a range of L_2 's for which similar A_{ROC} 's were observed. Regardless of L_2 , A_{ROC} for the suppression conditions (dashed lines), in which the contribution from the reflection source presumably has been reduced, is less than or equal to the A_{ROC} for the control condition (solid line). This result was observed for both $f_2=2$ kHz and $f_2=4$ kHz and demonstrates that test performance was either unchanged or decreased when suppressors were presented simultaneously with the eliciting stimuli.

B. Generalizability of predetermined suppressors

If A_{ROC} had increased when suppressors were presented, then this would have been consistent with the view that uncontrolled source contribution increased the variability (and decreased the accuracy) of DPOAE measurements as a diagnostic tool. If the A_{ROC} were the same with or without the presentation of the suppressors, then this would have indicated that contributions from the reflection source have no influence on DPOAE test performance or the suppressors used to control source contribution were not effective. However, the results shown in Fig. 1 indicate that suppressors, which were used to reduce or eliminate contributions from

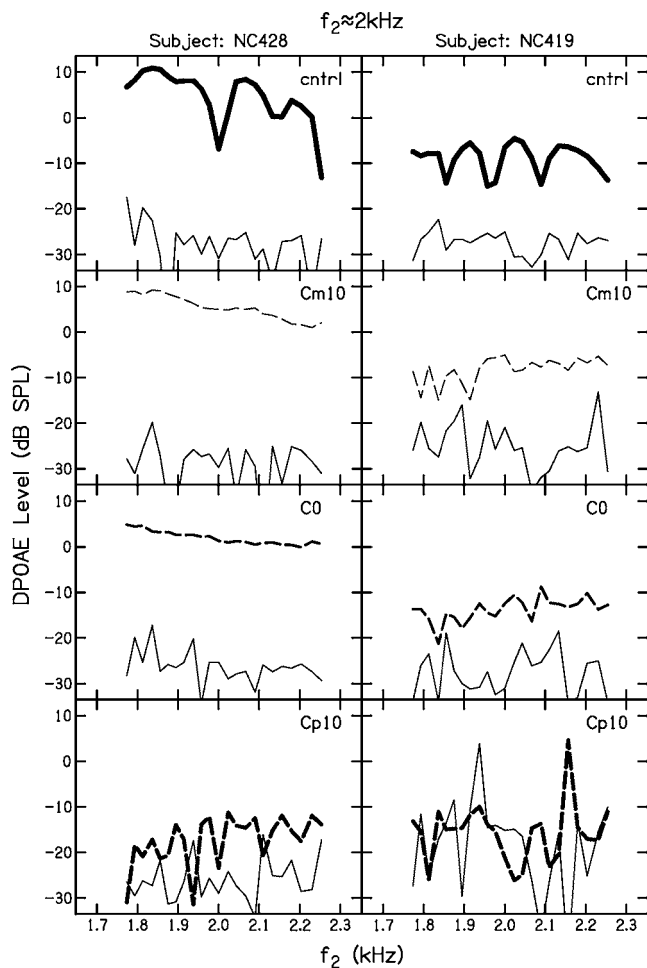


FIG. 2. DPOAE level (dB SPL) as a function of f_2 (kHz) for the 1/3 octave interval surrounding 2 kHz with $L_2=50$ dB SPL. Each column represents data for a different normal-hearing subject with each row representing a different test condition. Data for the control condition are plotted in the top row, with suppressor level increasing from top to bottom, as indicated by the inset in each panel. Within each panel the thick solid lines and dashed lines represent the DPOAE level and the thin solid lines near the bottom of each panel represent the associated noise.

the reflection source, most frequently had a deleterious effect on test performance, as shown by the reduction in A_{ROC} for essentially all suppressor conditions.

The data plotted in Figs. 2–4 represent an effort to better understand the results that were described in Fig. 1. These additional data were collected for the subset of subjects who participated in the main experiment for whom it was most likely that diagnostic errors would have occurred based on the original set of measurements. Specifically, the normal-hearing ears producing the smallest DPOAE levels and the hearing-impaired ears producing the largest DPOAE levels for the control (no suppressor) condition with the largest A_{ROC} ($L_2=50$ dB SPL for $f_2=2$ kHz; $L_2=35$ dB SPL for $f_2=4$ kHz) were recruited to return for additional data collection. Sixteen normal ears and 10 impaired ears for the 2 kHz conditions and 12 normal ears and 13 impaired ears for the 4 kHz conditions participated in this additional study. The decision to recruit in this manner was based on the idea that the normal-hearing ears with the smallest DPOAE levels were most likely to be misidentified as impaired, whereas the

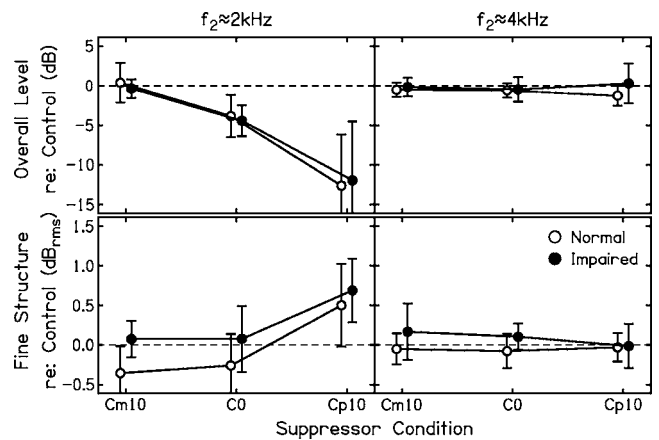


FIG. 3. The average overall DPOAE level (top row) and average fine-structure depth (bottom row) for the three suppression conditions relative to the control condition for the small group of subjects in whom diagnostic errors were most likely. Standard error bars represent ± 1 s.d. Results for $f_2=2$ kHz are plotted in the left column with corresponding data for $f_2=4$ kHz in the right column. The results for the normal and impaired ears are represented by the open and closed circles, respectively. The horizontal dashed line represents the point at which the control and suppression conditions are equivalent. These data were obtained from the DCT analyses as described in the text.

hearing-impaired ears with the largest DPOAE levels were most likely to be erroneously labeled as normal.

Figure 2 plots examples of DPOAE level as a function of f_2 for closely spaced frequencies in the 1/3 octave band surrounding 2 kHz for two subjects with normal hearing. This f_2 was selected because it is more likely that fine structure will be observed at this frequency, compared to when $f_2=4$ kHz (Dhar and Shaffer, 2004; Johnson *et al.*, 2006b). Each column in Fig. 2 plots results for a different subject and each row represents data for a different test condition, as indicated in each panel. Results for the control condition are plotted in the top row, with the results for conditions in which suppressor level increased presented in subsequent rows. Within each panel, DPOAE level is plotted (thick solid or dashed line) along with the associated noise levels (thin, solid line). For the subject whose data are shown in the left column, the Cm10 suppressor was effective at reducing the fine structure, without reducing the overall DPOAE level. The C0 and Cp10 suppressor levels produced differing amounts of over-suppression in which the overall level of the DPOAE was reduced. The data shown in the right column suggest that the predetermined suppressor levels were less effective for this subject. Fine structure is reduced in the Cm10 condition, particularly for $f_2 > 2$ kHz, but not as completely as for the subject whose data are shown in the left column. The C0 and Cp10 suppressors reduced overall DPOAE level, but were no more effective at smoothing the fine structure. Additionally, the Cp10 condition would be expected to result in a false-positive error because the overall level of the response has been suppressed to the noise floor. These data suggest that there is between-subject variability in the effectiveness of the predetermined suppressors, making it difficult to select conditions for general use, as would be needed for clinical measurements.

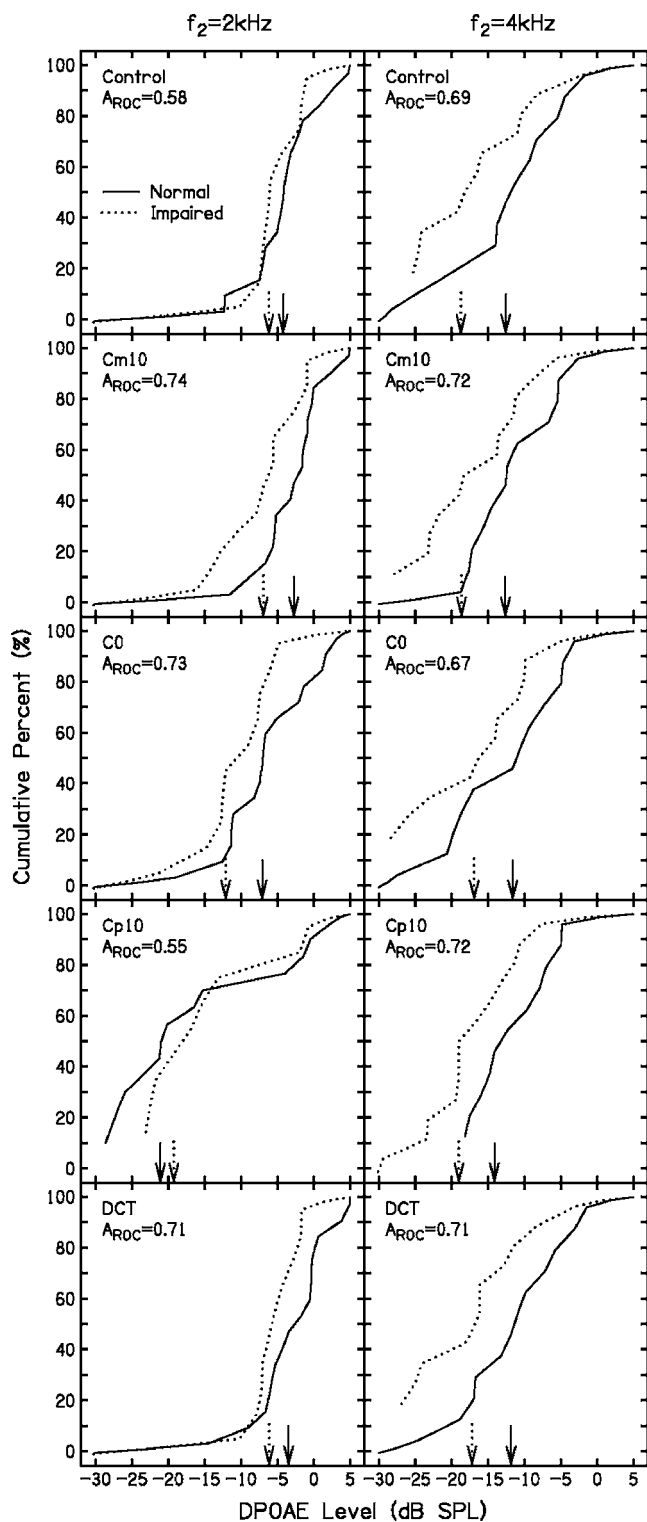


FIG. 4. Cumulative percentages of DPOAE level (dB SPL) for the 2 kHz (left column) and 4 kHz (right column) conditions for the small group of subjects (see the text for details). The rows represent, from top to bottom, results for the control condition, the three suppression conditions, and the frequency-smoothing (DCT) condition. The parameter in each panel is subject group (normal or impaired), as indicated in the legend. The arrows in each panel indicate the 50th percentile of DPOAE level for the normal and impaired ears. Also shown as an inset in each panel is the A_{ROC} for each test condition for this group of subjects.

The mean data (± 1 s.d.) shown in Fig. 3 describe the effectiveness of the suppressors for the entire group of subjects who returned for additional data collection. These data

were obtained by computing a DCT on the DPOAE fine-structure measures (see Fig. 2). The left column of Fig. 3 plots data for the 2 kHz conditions, with data for the 4 kHz conditions plotted in the right column. The top row plots the overall DPOAE level (the DC component) in decibels re: the control condition as a function of suppressor level (i.e., the three suppressor conditions used in the present study). The data in the bottom row describe the change in fine-structure depth for the three suppressor conditions relative to the control condition. These data were obtained by computing the rms level of the “high frequency” DCT coefficients, which provides a single value corresponding to fine-structure depth for each condition. These values for the suppressor conditions were then subtracted from the control condition value. The data plotted in the bottom row of Fig. 3, therefore, represent the difference in rms level in the suppressor conditions relative to the control condition. The parameter in each panel is group (open circle=normal, closed circle=impaired) with the symbols indicating the mean value (± 1 s.d.). The dashed horizontal line indicates the point at which the responses for control and suppressor conditions are equivalent.

In the top row, points below the dashed line indicate conditions where suppressors reduced the overall level. In the bottom row, points below the dashed line indicate conditions where suppressors reduced the fine-structure depth. As shown in Fig. 3, the lowest suppressor level (Cm10) was most effective at reducing the fine-structure depth without reducing the overall level for this subgroup of normal ears in the frequency interval surrounding 2 kHz, although there was variability across subjects, as represented by the error bars. The higher-level suppressors (C0 and Cp10) produced reductions in the overall level for both the normal and impaired sub-groups. A different pattern emerged for frequencies surrounding 4 kHz. In the 4 kHz interval, the suppressors had little influence on either fine-structure depth or overall level in both subject groups. The lack of a suppression effect at 4 kHz may be due to the lower prevalence of fine structure observed at 4 kHz as compared to 2 kHz (i.e., 1 of 12 normal ears had fine structure at 4 kHz as compared to 7 of 16 normal ears at 2 kHz), a frequency-dependent pattern that has been previously reported (Dhar and Shaffer, 2004; Johnson *et al.*, 2006b).

Figure 4 shows cumulative distributions of DPOAE level (dB SPL) for the subset of subjects who returned for additional data collection. Recall that these are the ears most at risk for diagnostic errors during the primary analysis based on the measurements of unsuppressed DPOAE level (Fig. 1). Results for $f_2=2$ and 4 kHz are plotted in the left and right columns, respectively. These data were collected for the primary levels at which the A_{ROC} was largest in the control condition in the larger sample of subjects ($L_2=50$ and 35 dB SPL at 2 and 4 kHz, respectively). From top to bottom, the rows represent results for the control condition, the three suppressor conditions, and a frequency-smoothing condition (the DCT condition). In the frequency-smoothing condition, the fine structure was eliminated through the use of the DCT as described in Johnson *et al.* (2006b). When using the DCT to eliminate DPOAE fine structure, the coefficients of the high-frequency components are set to zero so that only the

DC and slowly varying components remain. This preserves the overall level of the response but eliminates the fine structure.

Within each panel of Fig. 4, arrows indicate the DPOAE levels associated with the 50th percentile of the distribution of responses from both the normal (solid line) and impaired (dashed line) subgroups of subjects. The A_{ROC} values associated with each test condition for these subsets of subjects are shown as insets in each panel. These A_{ROC} values are lower than the best values shown in Fig. 1 because they represent results for the small group of ears on which diagnostic errors were most likely to occur. Stated another way, these ears were specifically selected because they represented cases in which there was the greatest overlap between normal and impaired responses, which, in turn, must result in a lower A_{ROC} .

In this group of subjects, the separation between the 50th percentiles for the suppressor and DCT conditions tended to be larger than or equivalent to the separation observed in the control condition, but only when $f_2=2$ kHz. At this frequency, two of three suppression conditions and the DCT condition resulted in larger A_{ROC} (by 12%–15%), compared to the control condition. No differences in A_{ROC} were observed when $f_2=4$ kHz, regardless of test condition. Here, the lack of a difference between the control and any of the conditions in which the reflection source was removed is expected, given the previous observations that fine structure (which presumably results from source interactions) is infrequently observed at 4 kHz (Dhar and Shaffer, 2004; Johnson *et al.*, 2006b). Furthermore, even the results from this small subset of ears at 2 kHz ($N=26$) should be viewed cautiously because they differ from the trends observed in the large sample ($N=205$). For the small group of ears on which diagnostic errors were most likely to occur, the data suggest that controlling source contribution, through either the use of a suppressor or using a frequency-smoothing approach (the DCT), resulted in fewer errors. The contrasting result in the large group, where no difference or poorer performance was observed, suggests that the use of suppressors to control cochlear source in a more general population of subjects introduced errors for cases where errors were not made in the control condition (i.e., the condition where source contributions were uncontrolled).

C. Threshold prediction

Data describing the relationship between behavioral thresholds and DPOAE thresholds are plotted in Figs. 5 and 6. Figure 5 plots these data for the $f_2=2$ kHz conditions, with comparable data for the $f_2=4$ kHz conditions plotted in Fig. 6. In both Figs. 5 and 6, behavioral threshold (dB HL) is plotted as a function of DPOAE threshold (dB SPL) for two different approaches to obtaining a DPOAE threshold (right and left columns). In both Figs. 5 and 6, the top row provides data for the control conditions and each subsequent row represents a different suppression condition, with suppressor level increasing as one moves down each column. Solid lines

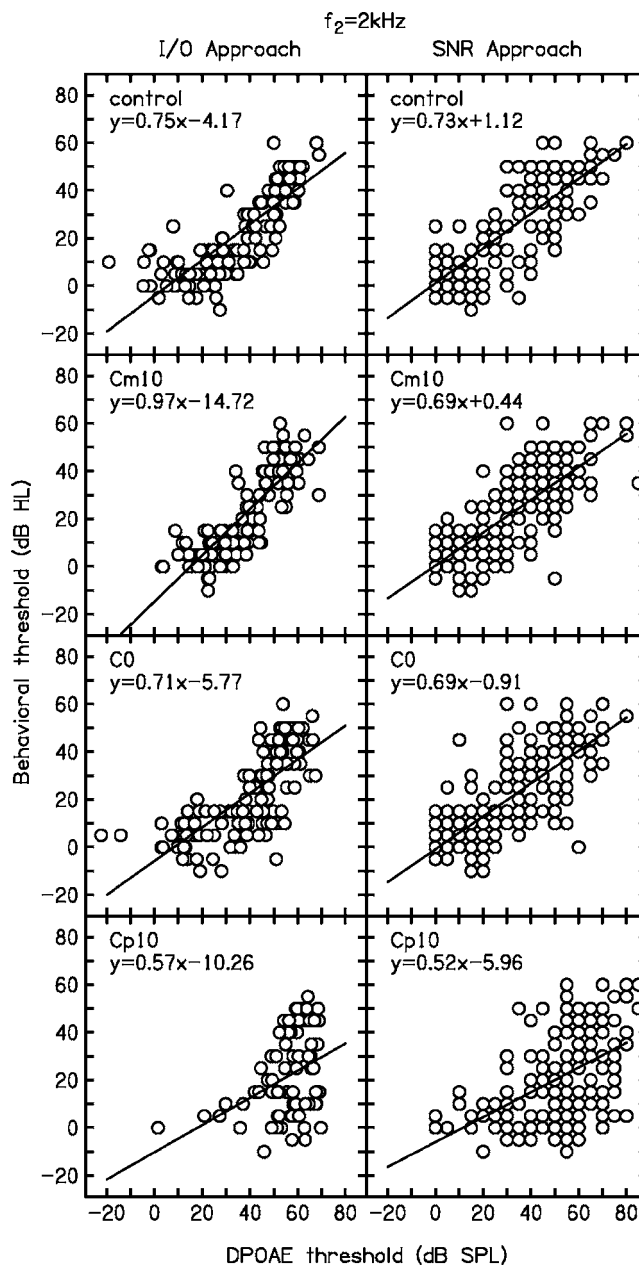


FIG. 5. Behavioral threshold (dB HL) as a function of DPOAE threshold (dB SPL) for two approaches to estimating DPOAE thresholds when $f_2=2$ kHz. The left column represents DPOAE thresholds obtained from the DPOAE I/O functions using the Gorga *et al.* (2003) modification of the Boege and Janssen (2002) technique (the I/O-function approach). The right column represents comparable data when the 3 dB SNR point is taken as the DPOAE threshold (the SNR approach). Each row represents data for a different suppression condition with suppressor level increasing from the top to the bottom. Solid lines in each panel represent linear fits to the data. The associated linear equations are shown as insets in each panel.

within each panel represent linear fits to the data. Insets in each panel describe the test condition and the equations associated with the linear fits.

Two definitions of DPOAE thresholds were used and subsequently correlated with behavioral thresholds. The DPOAE thresholds plotted in the left columns of Figs. 5 and 6 were obtained by fitting the DPOAE I/O function with a linear equation as was described in Gorga *et al.* (2003), which represents a modification of the technique originally proposed by Boege and Janssen (2002). Here, we will refer

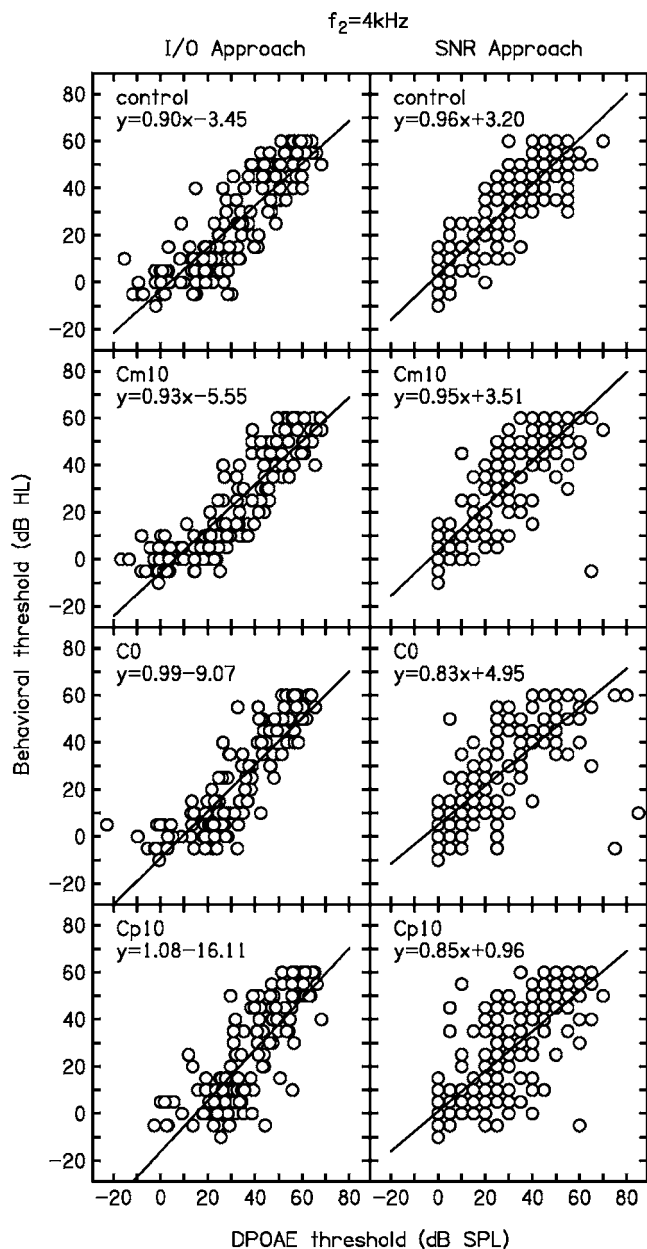


FIG. 6. Behavioral threshold (dB HL) as a function of DPOAE threshold (dB SPL) for the two approaches to estimating DPOAE thresholds when $f_2=4$ kHz. The plotting convention is the same as for Fig. 5.

to this technique as the *I/O*-function approach for estimating DPOAE threshold. Briefly, these DPOAE thresholds were estimated for *I/O* functions meeting several inclusion criteria. First, DPOAE *I/O* functions were required to have a minimum of three points with $\text{SNR} \geq 10$ dB (the SNR inclusion criterion). For those *I/O* functions meeting this requirement, the DPOAE level was converted to pressure (μPa) and then plotted as a function of L_2 (in dB SPL). A linear equation was fit to the resulting semilog (μPa vs. dB SPL) *I/O* functions. To be included in the next stage of analysis, the semilog *I/O* functions had to meet the following additional regression-based inclusion criteria: slope ≥ 0.1 $\mu\text{Pa}/\text{dB}$, correlation coefficient (r) ≥ 0.7 , and standard error ≤ 9 dB. For those *I/O* functions in which these regression-based inclusion criteria were met, DPOAE threshold was computed by solving the linear equation for the L_2 producing a DPOAE of

0 μPa . The relation between these DPOAE thresholds and behavioral threshold is shown in the left column of Figs. 5 and 6.

The DPOAE thresholds plotted in the right column of Figs. 5 and 6 were obtained by taking a simpler approach in which the lowest L_2 for which the SNR was ≥ 3 dB was defined as DPOAE threshold. This alternate approach was evaluated because correlations between behavioral threshold and simple estimates of DPOAE thresholds (such as those based on SNR) have been reported previously (e.g., Dorn *et al.*, 2001; Gorga *et al.*, 1996; Martin *et al.*, 1990; Nelson and Kimberley, 1992). Here, we will refer to this approach to estimating DPOAE threshold as the SNR approach.

Table II describes the number of ears meeting inclusion criteria, the correlation between behavioral and DPOAE thresholds, and the associated standard errors for the two approaches to threshold prediction when $f_2=2$ and 4 kHz. As can be seen in Figs. 5 and 6 and Table II, when predicting behavioral thresholds using the *I/O*-function approach, the lowest suppressor level (Cm10) resulted in the highest correlation between behavioral and DPOAE thresholds, although the differences were small (improvements of 0.02 to 0.03). When using the simpler SNR approach, the correlation was always poorer for the suppression conditions than for the control condition. Visual comparison of the data plotted in the left and right columns suggests more variability between behavioral and DPOAE thresholds when using the approach in which the L_2 at which $\text{SNR} \geq 3$ dB was defined as DPOAE threshold (the SNR approach). However, correlations and associated standard errors were similar for the two methods for obtaining DPOAE thresholds (see Table II), and actually favored the SNR approach in the control condition. The appearance of an increase in variability is most likely a consequence of the observation that the SNR approach could be applied to nearly every ear ($N=204$ or 205), regardless of test condition. In contrast, a number of ears did not meet the inclusion criteria when using the *I/O*-function approach. These data suggest that the simpler SNR approach may be more useful under clinical conditions and can be expected to produce results similar to those for the *I/O*-function approach but for a larger number of ears.

Figure 7 and Table III provide additional information regarding the ears not meeting the inclusion criteria for the *I/O*-function approach (either the SNR or the regression-based inclusion criteria). When compared to similar data reported by Gorga *et al.* (2003), a larger proportion of ears met all inclusion criteria in the control condition for the present study. In the control condition, 57% and 76% of ears for $f_2=2$ and 4 kHz, respectively, met all inclusion criteria as compared to 40% and 55% of ears for $f_2=2$ and 4 kHz, respectively, in the Gorga *et al.* data set. This overall higher success rate may be due, in part, to the fact that the present study restricted hearing-loss severity to no greater than 60 dB HL at 2 and 4 kHz, whereas ears with more severe hearing losses were included in Gorga *et al.* It also may relate to differences in the stimulus conditions used in the present study as compared to those used in Gorga *et al.* The present study used stimulus conditions that previously have been shown to produce larger DPOAEs in normal-hearing ears than those used

TABLE II. Summary of the threshold prediction results for $f_2=2$ and 4 kHz. Number of ears meeting the inclusion criteria for two approaches to threshold prediction along with correlations and associated standard errors for the four test conditions.

Condition	No. of ears meeting inclusion criteria		Correlation		Standard error (dB)	
	I/O approach	SNR approach	I/O approach	SNR approach	I/O approach	SNR approach
$f_2=2$ kHz						
Control	117	205	0.81	0.82	10.60	10.52
Cm10	117	205	0.84	0.77	9.87	11.55
C0	131	204	0.74	0.75	12.09	11.96
Cp10	75	204	0.38	0.50	16.29	15.79
$f_2=4$ kHz						
Control	156	204	0.86	0.90	10.87	9.61
Cm10	164	204	0.88	0.86	10.30	11.39
C0	147	205	0.86	0.78	10.97	13.76
Cp10	145	204	0.80	0.76	12.81	14.31

by Gorga *et al.* (Neely *et al.*, 2005; Johnson *et al.*, 2006a). These conditions might be expected to result in more ears meeting the inclusion criteria for the I/O-function approach.

As shown in Fig. 7 and Table III, when using the I/O-function approach in the present study, only a small number of the 205 ears did not meet the SNR inclusion criterion in the control condition [10/205 (5%) when $f_2=2$ kHz and 4/205 (2%) when $f_2=4$ kHz]. This represents a smaller proportion of subjects not meeting the SNR inclusion criterion than those reported by Gorga *et al.* for the same f_2 frequen-

cies (45% and 38% for $f_2=2$ and 4 kHz, respectively), which again was probably a consequence of differences in hearing-loss severity among the subjects included in that study and in the present one, but also may relate to the stimulus differ-

TABLE III. These data refer to the I/O-function approach to estimating DPOAE threshold. Shown here are the number of ears that did not meet the SNR and regression-based inclusion criteria, the percentage of these ears that have behavioral thresholds ≤ 30 dB HL, and the mean thresholds and standard deviations for these ears when $f_2=2$ and 4 kHz.

Condition	$f_2=2$ kHz		Mean threshold	Standard deviation
	No. not meeting SNR criterion	% with thresholds ≤ 30 dB HL		
Control	10	10	49.5	9.6
Cm10	10	30	47	13.8
C0	10	10	46	13.3
Cp10	56	46	31.6	17.8
Condition	$f_2=2$ kHz		Mean threshold	Standard deviation
	No. not meeting Regression criteria	% with thresholds ≤ 30 dB HL		
Control	78	88	13.8	14.5
Cm10	78	86	14.4	14.8
C0	64	88	14.5	16.3
Cp10	74	91	11.3	14.1
Condition	$f_2=4$ kHz		Mean threshold	Standard deviation
	No. not meeting SNR criterion	% with thresholds ≤ 30 dB HL		
Control	4	25	52.5	15
Cm10	5	40	39	27.5
C0	10	20	46	20.8
Cp10	9	11	47.8	10.9
Condition	$f_2=4$ kHz		Mean threshold	Standard deviation
	No. not meeting Regression criteria	% with thresholds ≤ 30 dB HL		
Control	45	53	26.3	23.7
Cm10	36	42	33.6	21.8
C0	48	58	25.4	22.1
Cp10	51	61	24.3	22.9

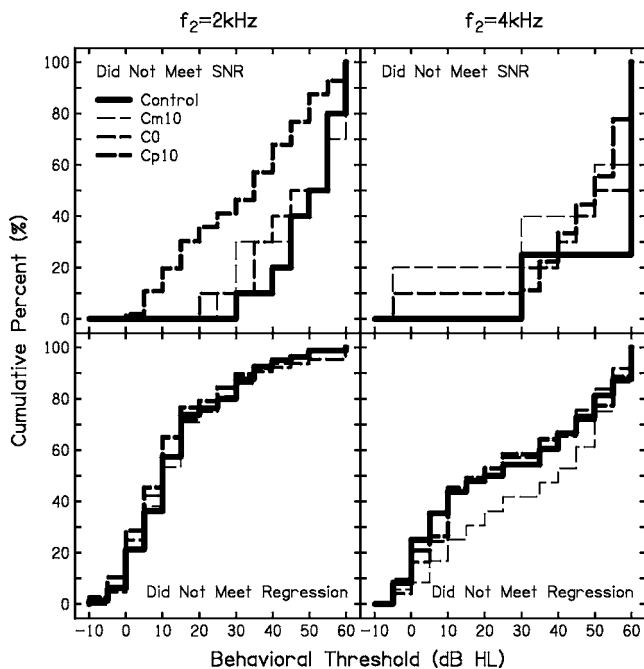


FIG. 7. Cumulative percentage of the behavioral thresholds (dB HL) for those I/O functions not meeting the Gorga *et al.* (2003) SNR inclusion criterion (top row) or the regression inclusion criteria (bottom row) when using the I/O-function approach to estimate DPOAE threshold. Data for the 2 and 4 kHz conditions are plotted in the left and right columns, respectively. The parameter within each panel is test condition. The control condition is represented by the thick, solid line, with increasing suppressor level represented by dashed lines of increasing thickness.

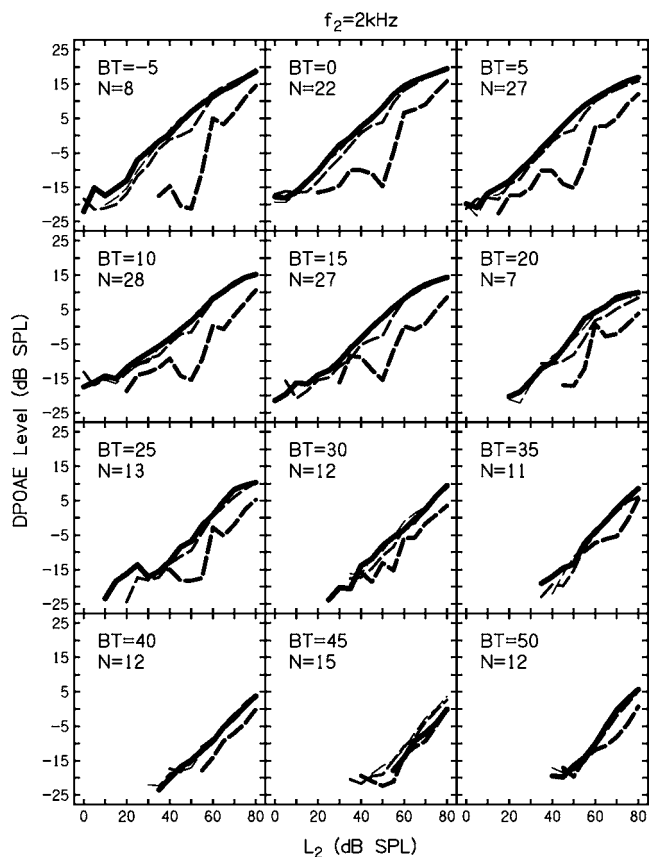


FIG. 8. Average DPOAE level (dB SPL) as a function of L_2 (dB SPL) when $f_2=2$ kHz. Each panel plots the average DPOAE level as a function of L_2 for all ears with the same behavioral threshold, the inset in each panel indicates the number of ears contributing in the control condition. Test condition is indicated by line type and thickness. Data for the control condition are plotted as thick solid lines with increasing suppressor level indicated by dashed lines with increasing thickness.

ences. Of those ears not meeting the SNR inclusion criterion (I/O-function approach), the ear was most likely to have a hearing loss that exceeded 30 dB HL, which is consistent with previous observations (Gorga *et al.*, 2003). With the exception of the highest suppressor level (Cp10) when $f_2=2$ kHz, the suppressors produced little change in the number of cases meeting the SNR inclusion criterion. The highest suppressor level when $f_2=2$ kHz resulted in more cases in which the 10 dB SNR inclusion criterion was not met, especially among ears with less hearing loss, an observation that is consistent with the idea that this suppression condition reduced the overall level of the DPOAE. In the control condition in the present study, more ears were excluded for not meeting the regression inclusion criteria than was reported by Gorga *et al.* For $f_2=2$ kHz, 38% did not meet the regression criteria as compared to 14% in the Gorga *et al.* data set. When $f_2=4$ kHz, 22% and 7% of cases did not meet inclusion criteria for the present study and for Gorga *et al.* study, respectively. While it is not clear why there were differences in the number of ears meeting the regression inclusion criteria for the present study as compared to Gorga *et al.*, the different suppressor levels had little influence on the extent to which regression-based criteria were met.

The data plotted in Figs. 8 ($f_2=2$ kHz) and 9 ($f_2=4$ kHz) represent the average I/O functions recorded for

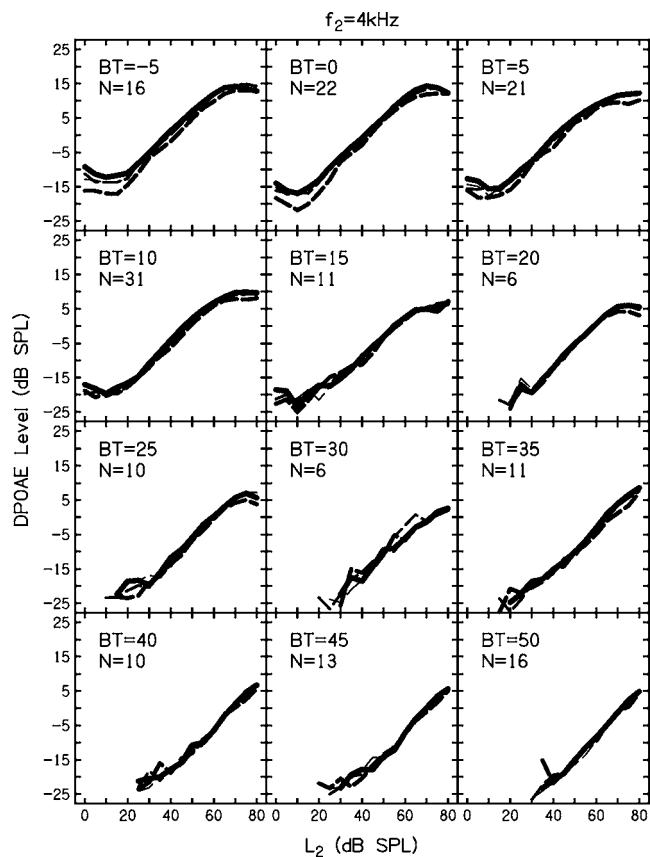


FIG. 9. Average DPOAE level (dB SPL) as a function of L_2 (dB SPL) when $f_2=4$ kHz. The plotting convention is the same as for Fig. 8.

ears with behavioral thresholds ranging from -5 to 50 dB HL. Each panel represents data for ears having different behavioral thresholds, as indicated within each panel; thus, each cell plots the average DPOAE level as a function of L_2 for all ears with the same behavioral threshold. Test condition is represented by line thickness and type (solid versus dashed), with the control condition plotted as a thick, solid line and the suppressor conditions plotted as dashed lines with increasing thickness indicating increasing suppressor level. Although the present study included ears with behavioral thresholds of -10 , 55 , and 60 dB HL, the average I/O functions for these threshold categories are not plotted because fewer than 5 ears were represented at both 2 and 4 kHz. For each of the cells shown in Figs. 8 and 9, the average I/O function was based on data from at least five subjects, with the number of ears contributing data in the control condition identified in each panel. The number of ears for the suppressor conditions in each panel may be less than the control condition due to over suppression that is observed in some ears and particularly for the highest suppressor level (Cp10).

The data plotted in Fig. 8 indicated that the lowest suppressor level (Cm10) had little influence on the shape of the average I/O function regardless of threshold category. The higher suppressor levels (C0 and Cp10) tended to reduce the DPOAE level, particularly for low L_2 levels. The reduction in DPOAE level was most pronounced for the highest suppressor level and ears with better behavioral thresholds. Inspection of the data plotted in Fig. 9 for $f_2=4$ kHz suggests

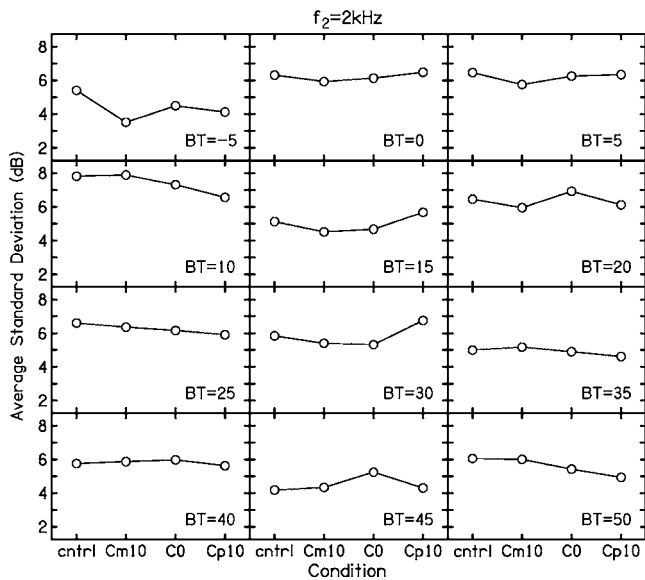


FIG. 10. Average standard deviation (dB) of DPOAE level across the entire I/O function as a function of test condition for each of 12 behavioral-threshold categories when $f_2=2$ kHz. These data represent the standard deviations associated with the I/O functions shown in Fig. 8.

that the suppressors had little effect on the DPOAE I/O functions. All of these observations are consistent with previous results showing that (1) high level suppressors have the largest effect on DPOAE level (e.g., Heitmann *et al.*, 1998; Konrad-Martin *et al.*, 2001; Johnson *et al.*, 2006b), (2) source interactions appear to be more apparent for low-level stimuli, compared to higher primary levels (e.g., Stover *et al.*, 1996b; Konrad-Martin *et al.*, 2001; Mauermann and Kollmeier, 2004), and (3) source interactions (as reflected in fine structure) are less evident at 4 kHz, compared to 2 kHz (e.g., Dhar and Shaffer, 2004; Johnson *et al.*, 2006b).

It is possible that the presence of the suppressor had little influence on the shape of the I/O function, but had an influence on the variability in DPOAE levels among ears with the same behavioral thresholds. If mixing of contributions from the distortion and reflection sources contributes to variability across ears, reducing the contribution from the reflection source might be expected to reduce the variability in DPOAE levels recorded from ears with the same behavioral threshold. Data exploring the influence of the suppressors on between-ear variability are plotted in Figs. 10 ($f_2=2$ kHz) and 11 ($f_2=4$ kHz), where average standard deviations are plotted as a function of stimulus condition for each of 12 behavioral-threshold categories. The standard deviation in DPOAE level for each of the threshold categories was computed for each L_2 . These standard deviations were then averaged for all L_2 levels within a given threshold category to provide the average s.d. for each threshold category and each test condition.

There were individual threshold categories where one or more suppressor levels produced a reduction in the average standard deviation (see, for example, $BT=-5$ when $f_2=2$ kHz and $BT=30$ when $f_2=4$ kHz). However, in general, the effect of the suppressors on the average standard deviation was small and included both reductions and increases in

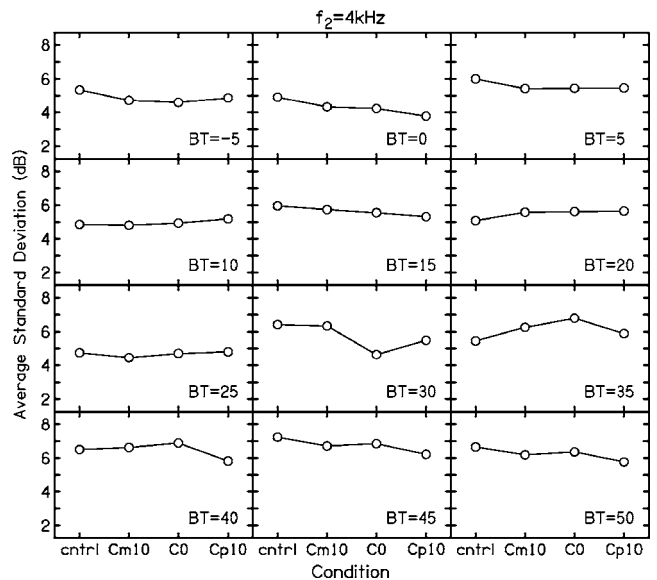


FIG. 11. Average standard deviation (dB) of DPOAE level across the entire I/O function as a function of test condition for each of 12 behavioral-threshold categories when $f_2=4$ kHz. These data represent the standard deviations associated with the I/O functions shown in Fig. 9.

variability across behavioral-threshold category. It, therefore, appears that the suppressors had little influence on between-ear variability in DPOAE level.

IV. DISCUSSION

In summary, the data reported here suggest that using a suppressor to control the cochlear sources contributing to the DPOAE did not improve the accuracy of the dichotomous decision in which auditory status is determined from DPOAE measurements. There was variability in the effectiveness of the suppressors across ears, although (on average) the lowest suppressor level reduced the depth of DPOAE fine structure in a small group of subjects in whom diagnostic errors were most likely to occur. For these subjects, the use of the suppressors also improved the accuracy of the dichotomous decision. Although the analyses for this group showed an improvement in DPOAE test performance when cochlear-source contribution was controlled, the lack of improvement (or, in some cases, the reduction) in the larger, unselected group suggests that the use of a suppressor will not improve test performance during more general clinical applications. Additionally, there was no improvement in the accuracy with which DPOAE measures predict behavioral threshold when using suppressors to control source contribution as compared to the control condition.

A. Dichotomous decision

In contrast to our hypotheses at the outset, the use of a suppressor did not improve test performance at any L_2 for either f_2 in the large group of subjects ($N=205$). If anything, the use of a suppressor resulted in poorer test performance (smaller A_{ROC}) for virtually every stimulus condition. As described earlier, there are several possible reasons for these observations. It may be that uncontrolled cochlear-source interaction and the resulting DPOAE fine structure does not

play a role in limiting DPOAE test performance. It is also possible that the use of predetermined suppressor levels to control cochlear-source contribution is not an effective approach. And, finally, it appears to be the case that suppressors have the unintended effect of reducing DPOAE level in some subjects with normal hearing, with a concomitant negative impact on test performance.

Regardless of the factors that underlie the reasons for the results observed in the present study, the fact remains that the use of suppressors to control source contribution under clinical conditions would be expected to reduce test performance, not improve it. The data collected in the small group of subjects for whom diagnostic errors were most likely were examined in order to further understand the results that were observed in the large group. The data presented in Figs. 2 and 3 suggest that, while there was variability in the effectiveness of the predetermined suppressors, the lowest suppressor level (Cm10) frequently reduced fine structure without affecting overall DPOAE level at the frequency for which fine structure was most evident in the control condition ($f_2=2$ kHz). Additionally, as shown in Fig. 3, the suppressors had little influence on fine-structure depth in the ears with hearing loss when $f_2=2$ kHz or in either group when $f_2=4$ kHz. Although some ears with hearing loss may produce DPOAE fine structure, previous reports indicate that it is only observed when auditory function at the cochlear location from which the reflection source arises is normal or nearly normal (Mauermann *et al.*, 1999a; Stover *et al.*, 1999; Konrad-Martin *et al.*, 2002). Additionally, as observed in the present study, and in previously reported data (e.g., Dhar and Shaffer, 2004; Johnson *et al.*, 2006b; Reuter and Hammershoi, 2006), fine structure is less prevalent in ears with normal hearing for frequencies surrounding 4 kHz. Given these observations, it was unlikely that we would see an effect of suppression in the ears with hearing loss or in ears with normal hearing at 4 kHz. Extending these observations to predictions of results from the small group, for the conditions where fine structure was not likely to be present, no effect of the suppressors was observed. In contrast, for the condition where fine structure was likely (normal ears for frequencies surrounding 2 kHz), the lowest suppressor level was effective at reducing fine-structure depth.

Our rationale for recruiting the normal ears with the smallest DPOAEs and the impaired ears with the largest DPOAEs for additional data collection was based on the idea that these ears were most likely to be misclassified when DPOAE level is used to make dichotomous decisions regarding auditory status. Furthermore, we hypothesized that, if interactions between distortion and reflection sources produce errors in the interpretation of clinical DPOAE measures, these errors would occur because the two sources added together in a destructive manner in normal ears, making them look more like impaired ears, or added in a constructive manner in impaired ears, making them look more like normal ears. The observation of larger A_{ROC} values for the lowest suppression levels and for the DCT condition as compared to the control condition when $f_2=2$ kHz suggests that the interaction of the distortion and reflection sources played a role in producing errors in this restricted subject

population. Furthermore, these data suggest that the use of both the lowest suppressor level and the DCT improved the accuracy of the dichotomous decision. It is important to remember, however, that these results are in conflict with the results observed in the larger, unselected group of subjects.

The most likely explanation for the differences in results between the larger group and the small, selected sample is that the suppressors introduced more errors in the larger, unselected group, than they reduced in the smaller group. To be effective clinically, suppressors must not only reduce fine-structure depth but also must cause little or no change to the overall DPOAE level, which presumably is generated near the f_2 place (i.e., the distortion source). It is for this reason that three suppressor levels were included in the present measurements. In this way, we hoped to include at least one suppressor level for each condition and each subject that was effective without directly determining the effectiveness of each suppressor in every subject, a procedure that cannot be used during clinical assessments. The effect of the suppressors on fine-structure depth and overall level in the larger group of subjects is unknown, but appears to differ from what was observed in the small group, in whom these effects were evaluated in more detail. In the large group, the most likely explanation for the reduction in A_{ROC} for the highest suppressor level is that this suppressor level reduced the overall DPOAE level in ears with normal hearing. The lack of improvement in A_{ROC} for the lower suppressor levels suggests that any normal ears for which the suppressors improved the accuracy of the dichotomous decision (i.e., those normal ears included in the small group) were more than offset by additional normal ears for which the suppressors introduced errors (presumably by suppressing the overall level). The poorer test performance in the larger sample cannot be attributed to influences on DPOAE level in ears with hearing loss because a reduction in DPOAE level in those subjects could only improve test performance (by increasing the hit rate).

The improvement in test performance (i.e., larger A_{ROC} values) observed in the small, selected group of ears for the lowest suppressor levels relative to the no-suppressor conditions suggests that interaction between the distortion and reflection sources may play a role in producing errors in identification of auditory status. However, observations in the large, unselected group of subjects, demonstrate that the suppression approach to controlling source contribution will not improve test performance, most likely because it introduces errors by oversuppressing the distortion source in ears with normal hearing. The DCT (and other frequency-smoothing approaches, e.g., Kalluri and Shera, 2001) can eliminate the reflection-source contribution without affecting the contribution from the distortion source, thereby leaving the overall level of the DPOAE unchanged, and might produce improvements in DPOAE test performance in an unselected group of subjects. However, the time involved in collecting the data with sufficient frequency resolution to compute the DCT is prohibitive in clinical contexts. As an alternative to the time-consuming approach associated with the DCT procedure, Long *et al.* (2004) have suggested that a frequency-sweep paradigm may be used to eliminate fine structure in DPOAE

responses. It remains to be seen if this technique produces improvements in DPOAE test performance for a general unselected group of subjects in a clinically feasible manner.

B. Threshold prediction

In general, the correlations between behavioral thresholds and DPOAE thresholds (control condition) are consistent with other reports in the literature (e.g., [Martin et al., 1990](#); [Dorn et al., 2001](#); [Boege and Janssen, 2002](#); [Gorga et al., 2003](#)). As noted earlier, suppressors did not improve the correlation and, in some cases, reduced the correlation. Of interest is the observation that nearly identical correlations were observed when either the simple SNR approach or the more complicated I/O-function approach were used to predict behavioral thresholds for both the control and suppression conditions (compare correlations for SNR approach and I/O approach in [Table II](#)). The SNR-based approach is conceptually simpler and has the added advantage of being applicable to a larger proportion of ears because threshold is defined as the L_2 for which the SNR=3 dB rather than the more restricted I/O-function approach that requires at least three points with 10 dB SNR. When $f_2=2$ kHz, the SNR-based approach could be used to estimate DPOAE thresholds in 100% (205/205) of ears. In contrast, the I/O-function approach could be used to predict DPOAE thresholds in only 57% (117/205) of ears at the same frequency. When $f_2=4$ kHz, the SNR and I/O-function approaches resulted in predictions of DPOAE thresholds in 99.5% (204/205) and 76% (156/205) of ears, respectively. While the number of cases in which the SNR-based approach could be used to predict DPOAE thresholds decreases when hearing losses exceed 60 dB HL, a similar decrease would be expected in the number of cases for which the I/O-function approach could be used to provide an estimate of DPOAE threshold.

The observation that the highest suppressor level excluded more ears for failing to meet the SNR inclusion criterion (I/O-function approach) is consistent with the idea that oversuppression occurred for this suppressor level. In other words, the highest suppressor level reduced both the reflection- and the distortion-source contributions to the DPOAE. Changes observed in the shape of the mean I/O functions plotted in [Figs. 8 and 9](#) are also consistent with this idea. These data also demonstrated that oversuppression was a more frequent occurrence when $f_2=2$ kHz as compared to 4 kHz.

V. CONCLUSIONS

In summary, the present data suggest that there would be little or no value in adding a suppressor during routine clinical DPOAE measurements, either to predict auditory status or to predict behavioral threshold from DPOAE threshold. In fact, the use of suppressors sometimes decreased test performance. While it was possible to improve diagnostic accuracy in a subset of preselected subjects for whom errors were most likely to occur, the improvement in test performance for these subjects was only achieved at the expense of a decrease in test performance in the larger, unselected sample of subjects. Furthermore, it is not possible to identify, a pri-

ori, those clinical subjects for whom the simultaneous presentation of a suppressor will improve diagnostic accuracy. Given that the unselected sample is representative of the patients one might encounter in the clinic, the present data would argue in favor of clinical measurements of DPOAEs without regard to concerns related to controlling source contribution with a suppressor.

ACKNOWLEDGMENTS

Work supported by the NIH NIDCD F32-DC007536, R01-DC02251, P30-DC04662. The helpful comments of two anonymous reviewers are also gratefully acknowledged.

¹A reduction in the contribution of the reflection source to the DPOAE does not necessarily result in a reduction in the depth of the fine structure. In cases where the reflection source dominates the DPOAE response, reducing the reflection-source contribution to the point where it is equal in level to the distortion-source component will produce more complete cancellation between the two sources. In this case, the depth of the fine structure may increase with reductions in reflection-source contribution ([Talmadge et al., 1999](#); [Dhar and Shaffer, 2004](#)).

- ANSI (1996). "Specifications for audiometers," ANSI 3.6-1996 (American National Standards Institute, New York).
- Boege, P., and Janssen, T. (2002). "Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O-functions in normal and cochlear hearing loss ears," *J. Acoust. Soc. Am.* **111**, 1810–1818.
- Dhar, S., and Shaffer, L. A. (2004). "Effects of a suppressor tone on distortion-product otoacoustic emissions fine structure: Why a universal suppressor level is not a practical solution to obtaining single-generator DP-Grams," *Ear Hear.* **25**, 573–585.
- Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (2001). "Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired ears," *J. Acoust. Soc. Am.* **110**, 3119–3131.
- Dorn, P. A., Piskorski, P., Gorga, M. P., Neely, S. T., and Keefe, D. H. (1999). "Predicting audiometric status from distortion product otoacoustic emissions using multivariate analyses," *Ear Hear.* **20**, 149–163.
- Elliott, E. (1958). "A ripple effect in the audiogram," *Nature (London)* **181**, 1076.
- Furst, M., Reshef, I., and Attias, J. (1992). "Manifestations of intense noise stimulation on spontaneous otoacoustic emission and threshold microstructure: Experiment and model," *J. Acoust. Soc. Am.* **91**, 1003–1014.
- Gorga, M. P., Dierking, D. M., Johnson, T. A., Beauchaine, K. L., Garner, C. A., and Neely, S. T. (2005). "A validation and potential clinical application of multivariate analyses of distortion-product otoacoustic emission data," *Ear Hear.* **26**, 593–607.
- Gorga, M. P., Neely, S. T., Bergman, B., Beauchaine, K. L., Kaminski, J. R., Peters, J., and Jesteadt, W. (1993). "Otoacoustic emissions from normal-hearing and hearing-impaired subjects: Distortion product responses," *J. Acoust. Soc. Am.* **93**, 2050–2060.
- Gorga, M. P., Neely, S. T., and Dorn, P. A. (1999). "Distortion product otoacoustic emission test performance for a priori criteria and for multi-frequency audiometric standards," *Ear Hear.* **20**, 345–362.
- Gorga, M. P., Neely, S. T., Dorn, P. A., and Hoover, B. M. (2003). "Further efforts to predict pure-tone thresholds from distortion product otoacoustic emission input/output functions," *J. Acoust. Soc. Am.* **113**, 3275–3284.
- Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., and Peters, J. (1997). "From laboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss," *Ear Hear.* **18**, 440–455.
- Gorga, M. P., Nelson, K., Davis, T., Dorn, P. A., and Neely, S. T. (2000). "Distortion product otoacoustic emission test performance when both $2f_1-f_2$ and $2f_2-f_1$ are used to predict auditory status," *J. Acoust. Soc. Am.* **107**, 2128–2135.
- Gorga, M. P., Stover, L., Neely, S. T., and Montoya, D. (1996). "The use of cumulative distributions to determine critical values and levels of confidence for clinical distortion product otoacoustic emission measurements," *J. Acoust. Soc. Am.* **100**, 968–977.
- He, N.-j., and Schmiedt, R. A. (1993). "Fine structure of the $2f_1-f_2$ acoustic

- distortion product: Changes with primary level," *J. Acoust. Soc. Am.* **94**, 2659–2669.
- He, N.-j., and Schmiedt, R. A. (1997). "Fine structure of the $2f_1-f_2$ acoustic distortion products: Effects of primary level and frequency ratios," *J. Acoust. Soc. Am.* **101**, 3554–3565.
- Heitmann, J., Waldmann, B., and Plinkert, P. K. (1996). "Limitations in the use of distortion product otoacoustic emissions in objective audiometry as the result of fine structure," *J. Acoust. Soc. Am.* **253**, 167–171.
- Heitmann, J., Waldmann, B., Schnitzler, H.-U., Plinkert, P. K., and Zenner, H.-P. (1998). "Suppression of distortion product otoacoustic emissions (DPOAE) near $2f_1-f_2$ removes DP-gram fine structure - Evidence for a secondary generator," *J. Acoust. Soc. Am.* **103**, 1527–1531.
- Horst, J. W., Wit, H. P., and Albers, F. W. J. (2002). "Quantification of audiogram fine-structure as a function of hearing threshold," *Hear. Res.* **176**, 105–112.
- Johnson, T. A., Neely, S. T., Garner, C. A., and Gorga, M. P. (2006a). "Influence of primary-level and primary-frequency ratio on human distortion product otoacoustic emissions," *J. Acoust. Soc. Am.* **119**, 418–428.
- Johnson, T. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (2006b). "Reducing reflected contributions to ear-canal distortion product otoacoustic emissions in humans," *J. Acoust. Soc. Am.* **119**, 3896–3907.
- Kalluri, R., and Shera, C. A. (2001). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," *J. Acoust. Soc. Am.* **109**, 622–637.
- Kapadia, S., and Lutman, M. E. (1999). "Reduced 'audiogram ripple' in normally-hearing subjects with weak otoacoustic emissions," *Audiology* **38**, 257–261.
- Kim, D. O., Paparello, J., Jung, M. D., Smurzynski, J., and Sun, X. (1996). "Distortion product otoacoustic emission test of sensorineural hearing loss: Performance regarding sensitivity, specificity, and receiver operating characteristics," *Acta Oto-Laryngol.* **116**, 3–11.
- Knight, R. D., and Kemp, D. T. (2001). "Wave and place fixed DPOAE maps of the human ear," *J. Acoust. Soc. Am.* **109**, 1513–1525.
- Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., Cyr, E., and Gorga, M. P. (2002). "Sources of DPOAEs revealed by suppression experiments, inverse fast Fourier transforms, and SFOAEs in impaired ears," *J. Acoust. Soc. Am.* **111**, 1800–1809.
- Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (2001). "Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears," *J. Acoust. Soc. Am.* **109**, 2862–2879.
- Long, G. R. (1984). "The microstructure of quiet and masked thresholds," *Hear. Res.* **15**, 73–87.
- Long, G. R., Talmadge, C. L., and Lee, J. (2004). "Using sweeping tones to evaluate DPOAE fine structure," presented at the 27th Annual Midwinter Research Meeting of the Association for Research in Otolaryngology, Daytona Beach, FL.
- Long, G. R., and Tubis, A. (1988a). "Investigations into the nature of the association between threshold microstructure and otoacoustic emissions," *Hear. Res.* **36**, 125–138.
- Long, G. R., and Tubis, A. (1988b). "Modification of spontaneous and evoked otoacoustic emissions and associated psychoacoustic microstructure by aspirin consumption," *J. Acoust. Soc. Am.* **84**, 1343–1353.
- Lutman, M. E., and Deeks, J. (1999). "Correspondence amongst microstructure patterns observed in otoacoustic emissions and Bekesy audiometry," *Audiology* **38**, 263–266.
- Martin, G. K., Ohlms, L. A., Franklin, D. J., Harris, F. P., and Lonsbury-Martin, B. L. (1990). "Distortion product emissions in humans. III. Influence of sensorineural hearing loss," *Ann. Otol. Rhinol. Laryngol. Suppl.* **147**, 30–42.
- Mauermann, M., and Kollmeier, B. (2004). "Distortion product otoacoustic emission (DPOAE) input/output functions and the influence of the second DPOAE source," *J. Acoust. Soc. Am.* **116**, 2199–2212.
- Mauermann, M., Long, G. R., and Kollmeier, B. (2004). "Fine structure of hearing threshold and loudness perception," *J. Acoust. Soc. Am.* **116**, 1066–1080.
- Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (1999a). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio f_2/f_1 ," *J. Acoust. Soc. Am.* **106**, 3473–3483.
- Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (1999b). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss," *J. Acoust. Soc. Am.* **106**, 3484–3491.
- Neely, S. T., and Gorga, M. P. (1998). "Comparison between intensity and pressure as measures of sound level in the ear canal," *J. Acoust. Soc. Am.* **104**, 2925–2934.
- Neely, S. T., Johnson, T. A., and Gorga, M. P. (2005). "Distortion-product otoacoustic emission measured with continuously varying stimulus level," *J. Acoust. Soc. Am.* **117**, 1248–1259.
- Neely, S. T., and Liu, Z. (1993). "EMAV: Otoacoustic emission averager," Technical Memorandum 17, Boys Town National Research Hospital, Omaha, NE.
- Nelson, D. A., and Kimberley, B. P. (1992). "Distortion-product emissions and auditory sensitivity in human ears with normal hearing and cochlear hearing loss," *J. Speech Hear. Res.* **35**, 1142–1159.
- Oswald, J. A., and Janssen, T. (2003). "Weighted DPOAE input/output functions: A tool for automatic assessment of hearing loss in clinical application," *Z. Med. Phys.* **13**, 93–98.
- Rao, K. R., and Yip, P. (1990). *Discrete Cosine Transform: Algorithms, Advantages, Applications* (Academic, Boston).
- Reuter, K., and Hammershoi, D. (2006). "Distortion product otoacoustic emission fine structure analysis of 50 normal-hearing humans," *J. Acoust. Soc. Am.* **120**, 270–279.
- Shaffer, L. A., and Dhar, S. (2006). "DPOAE component estimates and their relationship to hearing threshold," *J. Am. Acad. Audiol.* **17**, 279–292.
- Shaffer, L. A., Withnell, R. H., Dhar, S., Lilly, D. J., Goodman, S. S., and Harmon, K. M. (2003). "Sources and mechanisms of DPOAE generation: Implications for the prediction of auditory sensitivity," *Ear Hear.* **24**, 367–379.
- Shera, C. A. (2004). "Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions," *Ear Hear.* **25**, 86–97.
- Shera, C. A., and Guinan, J. J., Jr. (1999). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," *J. Acoust. Soc. Am.* **105**, 782–798.
- Siegel, J. H. (1994). "Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes," *J. Acoust. Soc. Am.* **95**, 2589–2597.
- Siegel, J. H., and Hirohata, E. T. (1994). "Sound calibration and distortion product otoacoustic emissions at high frequencies," *Hear. Res.* **80**, 146–152.
- Stover, L. J., Gorga, M. P., Neely, S. T., and Montoya, D. (1996a). "Toward optimizing the clinical utility of distortion product otoacoustic emission measurements," *J. Acoust. Soc. Am.* **100**, 956–967.
- Stover, L. J., Neely, S. T., and Gorga, M. P. (1996b). "Latency and multiple sources of distortion product otoacoustic emissions," *J. Acoust. Soc. Am.* **99**, 1016–1024.
- Stover, L. J., Neely, S. T., and Gorga, M. P. (1999). "Cochlear generation of intermodulation distortion revealed by DPOAE frequency functions in normal and impaired ears," *J. Acoust. Soc. Am.* **106**, 2669–2678.
- Swets, J. A. (1988). "Measuring the accuracy of diagnostic systems," *Science* **240**, 1285–1293.
- Swets, J. A., and Pickett, R. M. (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection* (Academic, New York).
- Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (1999). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," *J. Acoust. Soc. Am.* **105**, 275–292.
- 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.
- Wilson, H. K., and Lutman, M. E. (2006). "Mechanisms of generation of the $2f_2-f_1$ distortion product otoacoustic emission in humans," *J. Acoust. Soc. Am.* **120**, 2108–2115.
- Zweig, G., and Shera, C. A. (1995). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," *J. Acoust. Soc. Am.* **98**, 2018–2047.
- Zwicker, E. (1986). "Spontaneous oto-acoustic emissions, threshold in quiet, and just noticeable amplitude at low levels," in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York), pp. 49–59.
- Zwicker, E., and Schloth, E. (1984). "Interrelation of different oto-acoustic emissions," *J. Acoust. Soc. Am.* **75**, 1148–1154.

Cochlear reflectivity in transmission-line models and otoacoustic emission characteristic time delays

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

Christopher A. SHERA^{c)}

Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 20 July 2007; revised 18 September 2007; accepted 19 September 2007)

In transmission-line models of cochlear mechanics, predictions about otoacoustic-emission delays depend on the place- or wave-fixed nature of the emission generation mechanism. In this work, transient evoked otoacoustic emissions (TEOAEs), recorded at different stimulus levels in 10 young subjects, were analyzed using wavelet-based time-frequency analysis to determine the latency of each frequency component of the response. The same wave forms were Fourier analyzed to evaluate the phase-gradient delay as a function of frequency. Interpreting the relation between these two characteristic delays using cochlear models shows that most of the TEOAE response can be attributed to place-fixed reflection mechanisms. The causality principle explains observed correlations between fluctuations of the TEOAE amplitude and phase-gradient delay.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799498]

PACS number(s): 43.64.Jb, 43.64.Kc [BLM]

Pages: 3554–3561

I. INTRODUCTION

Otoacoustic emissions (OAEs) are acoustic signals measurable in the ear canal and generated in the cochlea, either spontaneously or as a response to an acoustic stimulus (Probst *et al.*, 1991). OAEs are a by-product of active, nonlinear amplification mediated by the outer hair cells. This nonlinear amplification gives the hearing system an excellent frequency discrimination capability, a high sensitivity to quiet sounds, dynamic range compression, and a protective saturation mechanism at high intensities.

The OAE classification depends on the evoking technique: Spontaneous OAEs are recorded without any stimulus, transient evoked OAEs (TEOAEs) are measured using a broadband, short duration, acoustic transient stimulus, while stimulus frequency OAEs (SFOAEs) are evoked by a pure tone. When the stimulus consists of two tones, with frequencies in a particular ratio (usually, $f_2/f_1 \approx 1.22$), distortion product OAEs (DPOAE) can be recorded at the cubic tone $2f_1-f_2$ (apical DPOAEs) or $2f_2-f_1$ (basal DPOAE).

Cochlear models (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). Due to the tonotopicity of the BM each Fourier component of the stimulus propagates

until it reaches its tonotopic place, where it produces the maximum transversal deformation of the basilar membrane (maximum of the activity pattern), then it is, almost totally, locally absorbed. In the above-mentioned models, OAEs are generated near the resonance place and propagate toward the base along the BM as a backward TW. The existence of such backward TWs has recently been questioned by He *et al.* (2007) for DPOAEs, on the basis that the phase of the BM vibration at the DPOAE frequency at a basal location leads that measured at a more apical location. Although the quality of their measurements is high, the interpretation of the He *et al.* data (2007) is not that straightforward, due to the complexity of the DP generation mechanisms. Thus, in this study we will assume that the OAEs propagate along the BM as backward TWs, keeping in mind the open discussion about this important issue.

Two main mechanisms are usually considered for the OAE generation: Nonlinear distortion and linear reflection (SHERA and Guinan, 1999). Although the cochlea behaves approximately linearly at low excitation levels, the cochlear response nonlinearity generates distortions at higher intensities. When the cochlea is simultaneously excited by two different frequencies, the system nonlinearity generates responses at frequencies that are different from the stimulus frequencies, as in the case of DPOAEs.

Nonlinear distortion occurs in a region about the tonotopic place of the stimulus frequency or, as in the case of the

^{a)}Electronic-mail: r.sisto@dil.ispesl.it

^{b)}Electronic-mail: arturo.moleti@roma2.infn.it

^{c)}Electronic-mail: shera@epl.meei.harvard.edu

DPOAEs, about a cochlear place that depends only on the frequencies of the stimulus, so, in these cases, a wave-fixed generation mechanism is thought to occur.

Another generation mechanism is associated with the reflection of a fraction of the TW in proximity of the tonotopic place. Models postulate the presence of micromechanical irregularities of the cochlear structure, randomly distributed along the basilar membrane to explain the OAE generation (Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera *et al.* 2005). These impedance irregularities act as backscattering sources for the forward wave. In this case, the mechanism would be place fixed, because the backscattering sources are localized at fixed positions. As the TW reaches appreciable amplitude only near the resonant place, the position of the OAE source is in any case coincident with the tonotopic region for each stimulus frequency.

Recently a generation mechanism of wave-fixed type was hypothesized, in which the reflection of a fraction of the incoming wave is due to the sharp transversal impedance variation in the maximum activity pattern region (Sisto and Moleti, 2005). Classical analyses of passive cochlear models (de Boer and MacKay, 1980), and perturbative analyses based on the osculating parameters formalism applied to the Wentzel-Kramers-Brillouin (WKB) solution in a one-dimensional (1D) transmission-line model (Shera and Zweig, 1991) suggest that any such wave-fixed mechanism will not produce appreciable reflections, due to the resonant absorption in the region where reflections originate. It is not fully clear if these last calculation techniques can be confidently used in the limit of a high quality factor in the resonance region, because there the wave vector rapidly changes and the group velocity inverts itself, violating the conditions required for the validity of the WKB approximation (Talmadge *et al.*, 2000). Therefore, the discussion remains open and needs experimental support.

In this complex theoretical-phenomenological framework, it is interesting to study the characteristic times of the OAE response. They provide, in fact, important information, not only regarding the signal propagation along the BM, but also regarding the OAE generation mechanisms. In the case of the click evoked TEOAEs (click duration $\approx 100 \mu\text{s}$) the 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. For this analysis different time-frequency techniques may be used, such as the short time Fourier transform, the wavelet transform (WT) (e.g., Tognola *et al.*, 1997; Sisto and Moleti, 2002), or matching pursuit algorithms (Jedrzejczak *et al.*, 2004). In our theoretical scheme, in which OAEs propagate backwards along the BM, this time delay is necessary for each frequency component of the TW to reach its tonotopic place and travel back to the base of the cochlea. This delay is a function of the mechanical characteristics of the BM, and, particularly, of the frequency selectivity of the active cochlear filter, which introduces a “build-up” time delay that is longer as the quality factor Q increases.

In the case of SFOAEs, evoked by a pure tone (continuous sinusoidal signal) it is possible to associate with each

frequency f another characteristic time, the phase-gradient delay, defined as the derivative of the phase with respect to the angular frequency, measured by varying the stimulus frequency in a short interval around f . For DPOAEs, the same technique can be used, but in that case the stimulus is composed by two tones at frequencies f_1 and f_2 , so it is possible to vary only f_1 , only f_2 or both in a constant ratio, defining three different characteristic times associated with DPOAEs.

In the case of TEOAEs, it is possible to measure the phase-gradient delay by analyzing the Fourier transform of the pulse response. It is therefore possible to directly compare the two different characteristic delays obtained by analyzing the same data. Time-frequency analysis of the TEOAE wave forms provides a direct estimate of the physical delay between the transient stimulus and the OAE response. In fact, the square of the wavelet coefficient amplitude at a given time and frequency measures the energy density within the corresponding “tile” in the time-frequency plane, and the time of the maximum of the wavelet coefficient of a given frequency is therefore associated with the physical delay of the OAE response at that frequency, provided that the contributions from noise and linear artifact are sufficiently small. Instead, the characteristic times associated with the phase gradient depend not only on the transmission properties of the BM, but also on the nature of the OAE generation mechanisms. Place-fixed and wave-fixed mechanisms, in fact, produce utterly different predictions about the frequency dependence of the phase of the response. Consequently, the comparison between the phase-gradient delay and the physical delay of a wave packet, measurable in the time domain using a transient stimulus, provides important information about OAE generation mechanisms. As different generation mechanism may have different dependence on the stimulus intensity, it is important to study this phenomenology at different stimulus levels.

In this work, the results of a comparison between the characteristic times of human TEOAEs, obtained by the WT technique and by the evaluation of the phase-gradient delay, are shown. The interpretation of the results in terms of transmission-line cochlear models permits us to test for the presence of different OAE generation mechanisms.

II. METHODS

TEOAE signals were measured in a group of young subjects (23–30 years, 6 males and 4 females) at different click stimulus levels, ranging from 60 to 90 dB peak sound pressure level (dB pSPL). The ILO Echoport 292 (Otodynamics, Ltd.) system was used in the nonlinear acquisition mode. The latency of each OAE component was evaluated by a time-frequency technique based on the wavelet transform, as described in Moleti *et al.* (2005). The wavelet transform is a linear time-frequency tool (Mallat, 1989), in which the wave form $s(t)$ is decomposed into a linear superposition of wavelets:

$$WT(\alpha, \tau) = \int s(t)h_{\alpha, \tau}(t)dt. \quad (1)$$

All these wavelets are scaled and time-shifted copies of a “mother” wavelet $h_0(t)$, an oscillating function with zero mean:

$$h_{\alpha, \tau}(t) = \frac{1}{\sqrt{|\alpha|}}h_0\left(\frac{t - \tau}{\alpha}\right). \quad (2)$$

In this study, we have used the mother wavelet (Tognola *et al.*, 1997):

$$h_0(t) = \frac{\cos(20t)}{(1 + t^4)}, \quad (3)$$

with t in milliseconds. Different choices of the mother wavelet do not significantly modify the results of the TEOAE time-frequency analysis, and it has also been demonstrated that even different time-frequency analysis methods give very similar latency estimates (Jedrzejczak *et al.*, 2004). Each wavelet is localized both in time, by the parameter τ , and in frequency, by the value of its scale parameter α . In the case of Eq. (3):

$$f(\alpha) = \frac{10}{\pi\alpha} \text{ kHz}. \quad (4)$$

As is apparent from Eq. (2), α is directly proportional to the time duration of the wavelet, and inversely proportional to its central frequency. Therefore the central frequency and the bandwidth of each wavelet are proportional to each other. At fixed α , the wavelet coefficient $WT(\alpha, \tau)$ is a function of time that represents the time evolution of the signal amplitude within the corresponding frequency band.

The continuous wavelet transform was integrated over ten 500 Hz bandwidths, centered at frequencies f_i from 0.5 to 5 kHz. The wavelet latency in the i th frequency band, $\tau_w(f_i)$, is defined as the time when the band-averaged wavelet coefficient reaches its maximum absolute value. This time is that associated with the return time of the most intense energy pulse within the i th frequency band.

The phase gradient delay was evaluated by differentiating the FT phase of the evoked response. Due to the limited acquisition time (20 ms), this operation provides an estimate mod(20 ms) of the phase gradient delay, i.e., with a possible shift of ± 20 ms. This ambiguity was arbitrarily removed by translating the obtained delays in a 20 ms range. As shown in Fig. 1, for each 500 Hz band a different range was chosen, using as a reference a typical power law fit to the experimental TEOAE latency (e.g., Sisto and Moleti, 2002). Of course, this choice is arbitrary, and, in Sec. IV, it will be discussed if it could be partly responsible for the agreement between phase-gradient delays and wavelet latencies. On the other hand, any other choice is equally arbitrary, and this one seemed to have at least an empirical justification based on previous knowledge.

III. THEORETICAL MODELS

Sisto and Moleti (2005) suggested that the rapid variation of the BM impedance experienced by each frequency

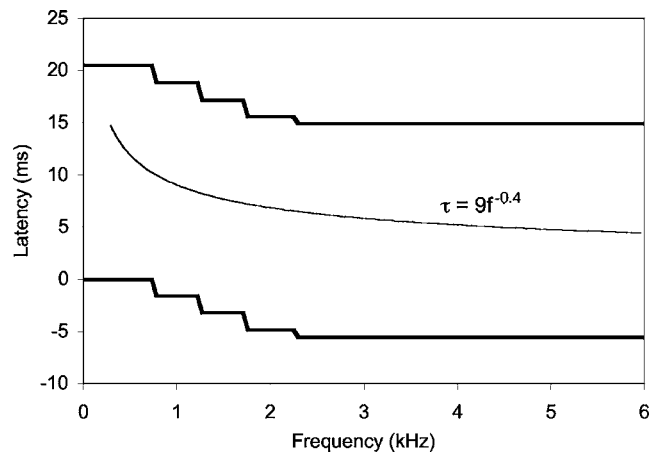


FIG. 1. The mod(20 ms) ambiguity in the evaluation of the phase-gradient delay has been arbitrarily solved by choosing a frequency-dependent acceptance window, limited by the two thick lines. The limits of the window have been chosen using as a reference the typical power law fit function (e.g., Sisto and Moleti, 2002) to the wavelet TEOAE latency (thin line).

component of the TW as it approaches its cochlear place can produce significant wave-fixed reflection, and thereby contribute to the generation of OAEs. The reflection mechanism due to rapid variation of the wave number is the analogue of the well-known reflection of a quantum particle from a potential well (Sisto and Moleti, 2005), which has a simple heuristic interpretation, being associated with the fact that a wave packet is partially reflected when it approaches a region in which the potential varies abruptly within a distance on the order of a wavelength.

Unfortunately, the osculating parameters method applied to the WKB solution of a 1D transmission line model (Shera and Zweig, 1991) suggests that the wave-fixed reflection due to this variation in BM impedance is negligible. This prediction can be obtained perturbatively: The local difference between the full equation and the equation exactly solved by the WKB solution becomes a source of wave reflection, which is distributed along the BM. The linearized equations in the frequency domain for a 1D transmission line with constant series impedance (box model) can be written:

$$\frac{\partial^2 P_d(x, \omega)}{\partial x^2} + k^2 P_d(x, \omega) = 0 \quad (5)$$

where ω is the angular OAE frequency, k is the wave vector, P_d is the differential pressure applied to the BM, and x is the longitudinal coordinate measured on the BM from the base. The use of this frequency domain representation is justified by the experimental observation that the TEOAE and SFOAE responses are quite similar to each other (Kalluri and Shera, 2007), suggesting that, even if the response at each frequency is definitely a nonlinear function of the stimulus, it may be approximately considered independent from the cochlear excitation at frequencies outside the critical bandwidth.

The resonance frequency is related to the cochlear position x along the BM by the Greenwood map (Greenwood, 1990):

$$\omega_0(x) = \omega_1 + \omega_{\max} e^{-k_\omega x}, \quad (6)$$

where $\omega_{\max} = (2\pi)20\,655$ rad/s, $\omega_1 = (2\pi)145$ rad/s, and $k_\omega = 1.382$ cm⁻¹.

The wave vector is tonotopically resonant and it may also include delayed-stiffness terms (Zweig 1991; Talmadge *et al.*, 1998):

$$k^2 = \frac{k_0^2 \omega^2}{\Delta(x, \omega)}, \quad (7)$$

$$\Delta(x, \omega) = \omega_0^2(x) - \omega^2 + i\omega\Gamma(x) - \rho_f \omega_0^2(x) e^{i\psi_f/\omega_0(x)} - \rho_s \omega_0^2(x) e^{i\psi_s/\omega_0(x)}. \quad (8)$$

The numerical values of the time-delayed stiffness parameters can be found in Talmadge *et al.* (1998). These slow and fast time-delayed stiffness terms effectively act as additional damping and antidamping terms, and they are very important for predicting the correct BM transfer function amplitude and phase, as well as a tall and broad activity pattern extending to a region basal to the resonant place. This is not a main point in this study, whose results apply to a generic transmission line cochlear model, assuming transmission of forward and backward waves along the BM and a tonotopically resonant cochlear response.

For a tonotopically resonant 1D transmission line cochlear model, the roundtrip delay of a particular Fourier component of the OAE response to a pulse stimulus is given by (Sisto and Moleti, 2002):

$$\tau_{rt}(\omega) = 2 \int_0^{\hat{x}} \frac{\partial \text{Re}(k)}{\partial \omega} dx. \quad (9)$$

This is the physical delay associated with the roundtrip path of a traveling wave packet centered at frequency $f = \omega/2\pi$, propagating along the BM at its local group velocity, from the base to its resonance place and back.

In this study the assumption is made that the wavelet latency $\tau_W(f)$ defined in the previous section is a direct measure of this physical delay [$\tau_W(f) = \tau_{rt}(f)$], because $\tau_W(f)$ measures the time interval between the click stimulus and the arrival of the most intense OAE wave packet at frequency f .

In the WKB approximation, the solution has the form of a superposition of basis functions, which are progressive and regressive waves. The WKB solution is the exact solution of the equation (Shera and Zweig, 1991):

$$\frac{\partial^2 \beta_\pm}{\partial x^2} + k^2(1 + \varepsilon_\pm) \beta_\pm(x, \omega) = 0. \quad (10)$$

If the function ε is sufficiently small, the perturbative regime holds and the basis functions are good approximate solutions of Eq. (5). The exact solution can be developed on the basis of the WKB functions:

$$P_d = P_+ + P_- = \psi_+ \beta_+ + \psi_- \beta_-, \quad (11)$$

where P_+ represents a wave that is traveling forward to the apex, whereas P_- is a wave traveling back to the base. The coefficients of the basis functions, called osculating parameters, can be interpreted as amplitudes for the forward and backward traveling waves. The amplitude of the backward

traveling wave generated by the partial reflection of a forward traveling wave of unitary amplitude (reflection coefficient) is

$$\psi_-(x) \cong \int_x^\infty dx' \beta_+ \sigma_+ \beta_+, \quad (12)$$

where σ_+ can be defined as a perturbative potential, which is a local source of reflection, given by

$$\sigma_+ = k^2 \gamma \varepsilon_+. \quad (13)$$

The function γ is the inverse of the Wronskian determinant of the system of second-order differential equations (10). The Wronskian is the determinant of a matrix whose first line contains the basis functions and whose second line contains their first derivatives, and it is therefore dependent on the choice of the basis functions. Using the basis functions of the type:

$$\beta_\pm^W(x, \omega) \approx \frac{1}{\sqrt{k}} e^{\mp i \int_0^x k(x', \omega) dx'}, \quad (14)$$

which are usually those defined as WKB basis functions, yields the following equations:

$$\varepsilon_\pm^W = -\frac{1}{k^2} \sqrt{k} \frac{\partial^2}{\partial x^2} \frac{1}{\sqrt{k}}; \quad \gamma^W = \frac{1}{2i}; \quad \sigma_\pm^W = -\frac{i}{2} k^2 \varepsilon_\pm^W. \quad (15)$$

By studying the behavior of the integral along the BM given by Eq. (12), which is a perturbative estimate of the cochlear reflection, it can be seen that the integral, after having reached a significant amplitude at a cochlear position close to the tonotopic place, falls nearly exactly to zero at the tonotopic place, independent of the quality factor of the resonance itself. In the case of high quality factor resonances, it is not clear if this perturbative technique can be fully trusted in the resonance region, where the WKB approximation becomes inadequate due to the rapid variation of the wave vector and the inversion of the group velocity. Therefore, it is difficult to exclude that a small but measurable contribution to cochlear reflectivity, and so to OAEs generation, could come from a mechanism of the type proposed by Sisto and Moleti (2005). This mechanism would be intrinsically wave fixed, because it would be localized to the resonant place position, which varies with the frequency of the considered stimulus component.

In the perturbative WKB calculation, the phase of the OAE response is equal to the phase of ψ_- evaluated at $x=0$. Due to the resonant shape of the potential σ_+ , the integral of Eq. (12) is dominated by the contribution from the resonance place. Therefore, its phase is given, approximately, by

$$\phi(\omega) \approx -2 \int_0^{\hat{x}} \text{Re}(k) dx, \quad (16)$$

an expression that ignores small contributions from the external factor, $k^{-1/2}$, and any ‘‘random’’ contributions from the perturbative potential, σ_+ .

If the OAE generation mechanism is place fixed, the mean phase-gradient delay at angular frequency ω (i.e., the nonrandom component of the delay) is given by

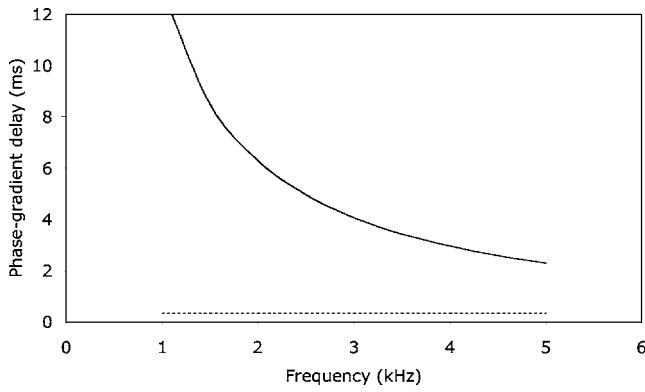


FIG. 2. Mean component of the phase-gradient delay predicted for place-fixed (solid line) and wave-fixed (dotted line) OAE generation mechanisms, using a very simple 1D transmission-line cochlear model (Sisto and Moleti, 2007), with $Q=10$.

$$\tau_{pg}(\omega) = -\frac{d\phi}{d\omega} = 2 \int_0^{\hat{x}} \frac{\partial \text{Re}(k)}{\partial \omega} dx = \tau_{rt}(\omega), \quad (17)$$

whereas, if the OAEs generation mechanism were wave-fixed, the approximated equation would be valid:

$$\begin{aligned} \tau_{pg}(\omega) = -\frac{\partial \phi}{\partial \omega} - \frac{\partial \phi}{\partial \hat{x}} \frac{\partial \hat{x}}{\partial \omega} &= 2 \int_0^{\hat{x}} \frac{\partial \text{Re}(k)}{\partial \omega} dx \\ &+ 2 \text{Re}(k(\hat{x})) \frac{\partial \hat{x}}{\partial \omega} \approx \tau_{rt}(\omega) - \frac{2 \text{Re}(k(\hat{x}))}{\omega k_\omega} \end{aligned} \quad (18)$$

where the Greenwood map [Eq. (5)] has been used in a simplified form, neglecting the constant term ω_1 , and where we explicitly reminded that the phase-gradient delay is operationally defined as the total derivative of the phase with respect to the frequency and so, in the case of the wave-fixed generation model, also the variation of the upper limit of the integration path has to be considered. Although the last term in Eq. (18) is model dependent, it is typically of the same order of τ_{rt} . In scale-invariant cochlear models, the wave-fixed prediction is that the phase-gradient delay will be approximately zero. If the scale-invariance symmetry is broken, as happens in the real cochlea, the wave-fixed phase-gradient delay predicted by Eq. (18) can be either positive or negative but in either case it is typically much smaller than τ_{rt} . Using, for example, a very simple linearized cochlear model (Sisto and Moleti, 2007), we get the prediction:

$$\tau_{rt}(\omega) = \frac{k_0}{\omega k_\omega} \left(\beta \sqrt{Q} - \frac{2\omega}{\omega_{\max}} \right) = \tau_{pg}(\text{place fixed}) \quad (19)$$

$$\tau_{pg}(\text{wave fixed}) \cong \frac{2k_0}{\omega_{\max} k_\omega} \quad (20)$$

where β is a model-dependent dimensional constant of order unity, the quality factor Q is an effective one, including the effect of the delayed stiffness terms. In Fig. 2 the two functions are plotted, for a scale-invariant case, with $Q=10$. More accurate models would give different quantitative results, but the main point is that wave-fixed OAE generation implies much shorter phase-gradient delays.

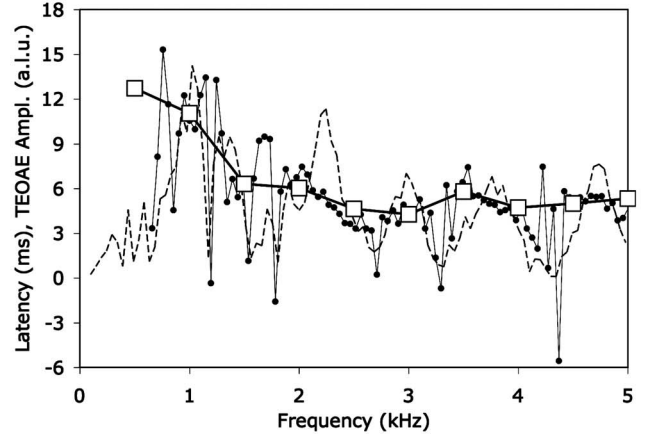
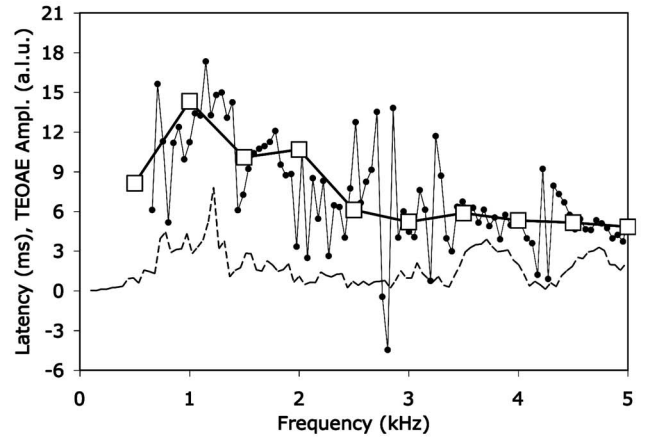


FIG. 3. Comparison between phase-gradient delay (circles) and wavelet latency (squares) in the right ear of subject AM, for stimulus levels of 60 and 90 dB pSPL. The TEOAEs spectrum is also shown (dashed line) in arbitrary linear units.

IV. RESULTS AND DISCUSSION

Figure 3 shows the phase-gradient delays (circles) and wavelet delays (squares) versus frequency, in the case of the ear AM-R, for two stimulus levels, 60 (top) and 90 dB pSPL (bottom). At lower stimulus levels the signal-to-noise ratio (SNR) is lower, and, consequently, the phase and its gradient show a more irregular behavior. In general, the overall correspondence between the two delays is quite evident, even at the lowest stimulus levels. At high stimulus levels the SNR is higher, and the noise-induced fluctuations of the phase-gradient delay are much smaller. We note that there are still a few points for which the phase-gradient delay is around zero. These points typically correspond to frequencies at which the TEOAE amplitude (dashed line) is very low, as it is evident from the bottom panel of Fig. 3.

The correlation between the two different latency measurements, on the whole set of ears and range of stimulus levels, is shown in Fig. 4 for a frequency of 1.5 kHz. The wide range of latencies is due to the variation of the stimulus level (Sisto and Moleti, 2007). The correlation is generally good, except for a subset of data for which the phase-gradient delay is significantly shorter than the wavelet latency. A warning should be given here, because this asymmetry could partly be dependent on the choice of the phase-gradient latency “window” shown in Fig. 1. These data could

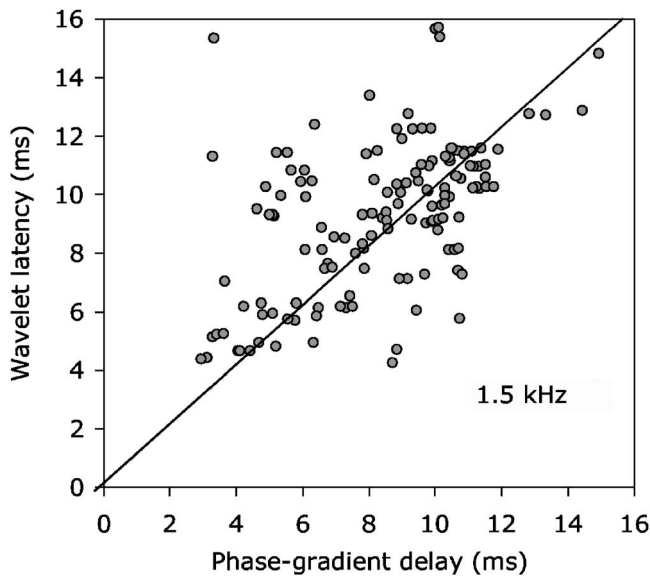


FIG. 4. Correlation between phase-gradient delay and wavelet latency (obtained by the WT technique) in all the subjects, at 1.5 kHz, with stimulus varying from 60 to 90 dB pSPL. The phase-gradient delay has been averaged over three spectral points centered at 1.5 kHz to reduce the size of the fluctuations due to noise.

be interpreted as coming from cochlear regions where some wave-fixed mechanism (e.g., reflection and/or distortion) is dominant. However, different explanations are also possible: phase-gradient delays close to zero can arise due to fluctuations in TEOAE amplitude, as discussed later, and/or due to errors caused by low SNR. In any case, the interpretation of the results shown above suggests that the main contribution to the TEOAE signal comes from place-fixed mechanisms, as predicted by the coherent-reflection (CRF) theory (Shera *et al.*, 2005; Talmadge *et al.*, 1998), but it does not exclude the presence of a small wave-fixed contribution, whose nature deserves further investigation.

It is possible to observe correlated spectral fluctuations of phase-gradient delay and TEOAE amplitude in the single TEOAE response. In some cases, these correlated fluctuations are quite evident (Fig. 5, top panel) in the frequency bands characterized by both high SNR and strong TEOAE level fluctuations. If the SNR is low, large fluctuations of τ_{pg} due to noise may completely mask this effect. Selecting only the frequency bands with SNR higher than a given threshold level, is it possible to give a different graphical representation of this correlation for each ear by plotting the fluctuation of τ_{pg} with respect to the band wavelet delay τ_w , against the corresponding fluctuation of the TEOAE amplitude with respect to its mean amplitude value in the corresponding frequency band. This is shown, for four ears, in the bottom panels of Fig. 5, for SNR > 3. In some cases, the correlation is significant. Similar correlations between OAE amplitude and group-delay features were observed in chinchilla SFOAEs by Siegel *et al.* (2005). Correlated fluctuations between TEOAE magnitude and phase-gradient delay are expected because of constraints imposed by causality (i.e., by the physical fact that TEOAEs cannot occur before their evoking stimulus).

If $p_{\text{oe}}(t)$ is the TEOAE pressure evoked by an acoustic impulse applied at $t=0$, then causality requires that $p_{\text{oe}}(t < 0) = 0$. In the frequency domain, causality requires that the real and imaginary parts of $P_{\text{oe}}(\omega)$, the TEOAE spectrum, be Hilbert transforms of one another (e.g., Bode, 1945). To see how correlated fluctuations in TEOAE magnitude and delay can arise, we consider an idealized model for the TEOAE spectrum.

Since causal functions can be written as the product of an all-pass and a minimum-phase component (Papoulis, 1962), we let:

$$P_{\text{oe}}(\omega) = P_{\text{mp}}(\omega)e^{-i\omega\tau}, \quad (21)$$

where $P_{\text{mp}}(\omega)$ is minimum-phase and the pure delay $e^{-i\omega\tau}$ is the all-pass component. We know empirically that TEOAE magnitudes fluctuate up and down with frequency. (According to the coherent reflection model, magnitude fluctuations reflect the spatial irregularity of the perturbations that “scatter” the forward-traveling wave. Magnitude fluctuations can also arise due to interference between place- and wave-fixed components.) To explore how TEOAE magnitude fluctuations affect the phase-gradient delay, we model variations in $|P_{\text{mp}}(\omega)|$ using a simple sinusoidal ripple. In other words, we take

$$|P_{\text{mp}}(\omega)| = P_0(1 + r \cos(\omega\tau_r)), \quad (22)$$

where P_0 is the mean emission magnitude and the parameters r and τ_r determine the amplitude and spectral period of the ripple. The phase of $P_{\text{mp}}(\omega)$ is then determined by the requirement that $P_{\text{mp}}(\omega)$ be minimum-phase (i.e., that the real and imaginary parts of $\log P_{\text{mp}}(\omega)$ be Hilbert transforms of one another). When the ripples are small ($r \ll 1$) this yields a total TEOAE of the form

$$P_{\text{oe}}(\omega) = P_0(1 + r \cos(\omega\tau_r))e^{-i(\omega\tau + r \sin(\omega\tau_r))}. \quad (23)$$

Computing the log-magnitude and phase-gradient delay of $P_{\text{oe}}(\omega)$ yields

$$\log|P_{\text{oe}}(\omega)| \approx \log P_0 + r \cos(\omega\tau_r), \quad (24)$$

and

$$\tau_{pg}(\omega) \approx \tau + r\tau_r \cos(\omega\tau_r). \quad (25)$$

Note that the ripples in the phase-gradient delay have the same frequency dependence and phase as the ripples in the magnitude. Thus, when the TEOAE magnitude is at a minimum, the phase-gradient delay will also be minimum. If the magnitude fluctuations are large or rapid enough (i.e., $r\tau_r > \tau$) it is possible to have phase-gradient delays close to or even less than zero. These correlations are independent of how the TEOAEs are generated; they result simply from the constraints imposed by causality.

Although our simple “ripple model” for TEOAE magnitude fluctuations provides only an extremely idealized representation of actual TEOAEs, it does predict correlations seen in the data.

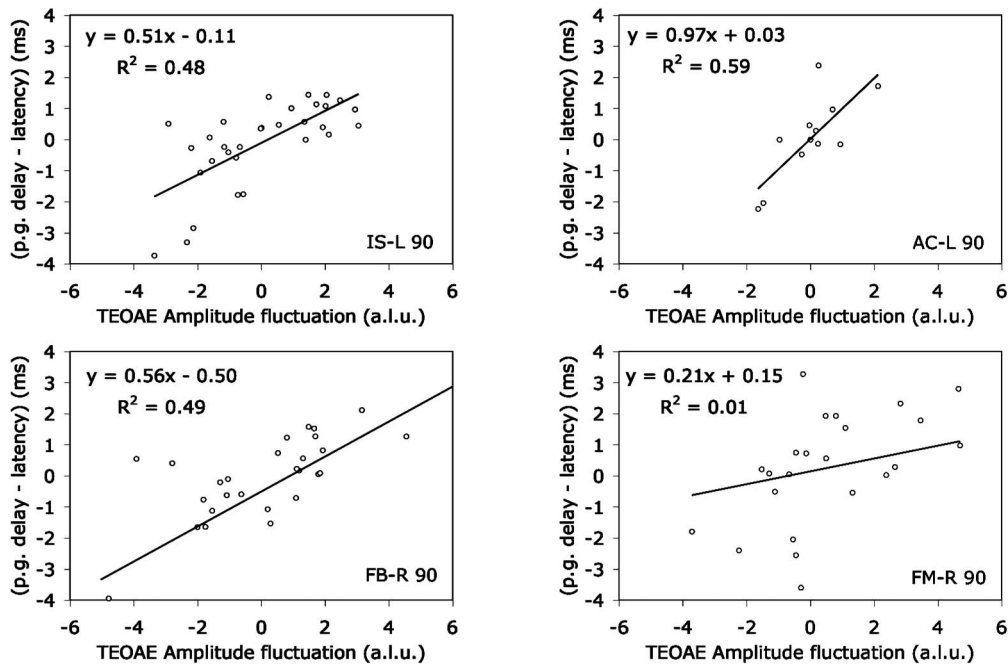
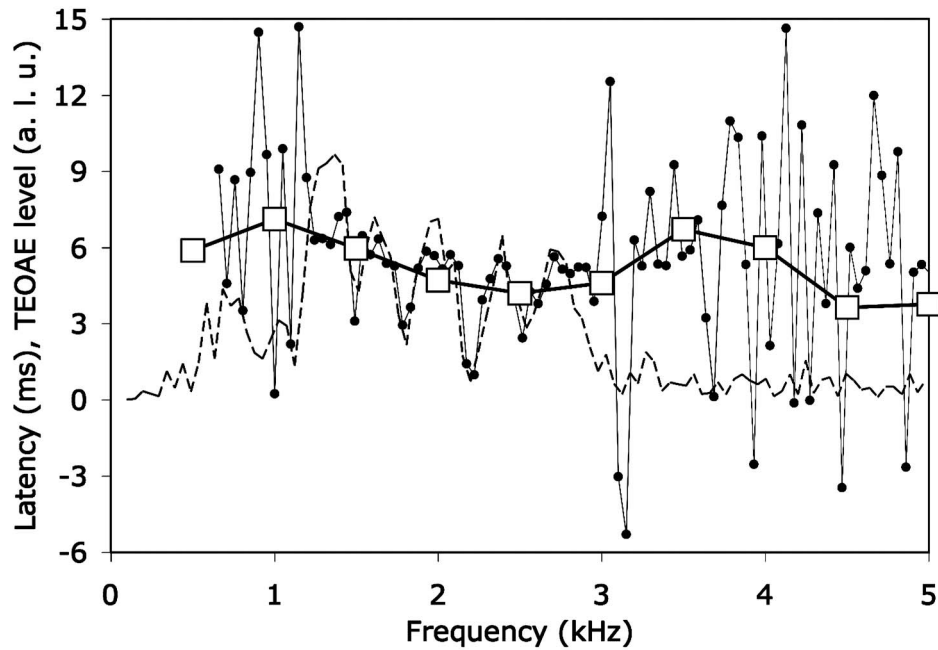


FIG. 5. Correlation between the fluctuations of TEOAE amplitude and phase-gradient delay. In the high-SNR part of the TEOAE spectrum, oscillations of the TEOAE amplitude (dashed line), correlated with those of the phase-gradient delay (circles) around the wavelet latency (squares), are clearly visible in the left ear of subject IS. This correlation can be shown also (bottom, for four ears) by plotting the difference between the phase-gradient delay and the wavelet band latency against the fluctuations of the TEOAE amplitude, having selected only the frequency bands in which TEOAE SNR > 3.

V. CONCLUSIONS

The results of this study generally support place-fixed generation models of OAEs, showing agreement between the TEOAEs latency estimates based on phase-gradient measurement and those obtained by applying a wavelet algorithm. The data may also contain small contributions from OAEs with phase-gradient delays much shorter than the wavelet latency, perhaps as a result of some wave-fixed generation mechanism. Further study is needed to establish whether

these secondary components originate via partial reflection of each stimulus frequency component near its tonotopic place. Correlated spectral fluctuations of phase-gradient delay and TEOAE amplitude have been observed in single TEOAE responses, as predicted by basic causality requirements.

ACKNOWLEDGMENTS

The authors thank the two anonymous reviewers for helpful comments on the manuscript. This work was partially

supported by Grant No. R01 DC003687 (C.A.S.) from the NIDCD, National Institutes of Health.

- Bode, H. (1945). *Network Analysis and Feedback Amplifier Design* (Van Nostrand Reinhold, Princeton, N.J.).
- de Boer, E., and MacKay, R. (1980). "Reflections on reflections," *J. Acoust. Soc. Am.* **67**, 882–890.
- 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.
- He, W., Nuttall, A. L., and Ren, T. (2007). "Two-tone distortion at different longitudinal locations on the basilar membrane," *Hear. Res.* **228**, 112–122.
- 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.
- Mallat, S. G. (1989). "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 674–693.
- 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.
- Papoulis, A. (1962). *The Fourier Integral and its Applications* (McGraw-Hill, New York).
- Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (1991). "A review of otoacoustic emissions," *J. Acoust. Soc. Am.* **89**, 2027–2067.
- 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., and Zweig, G. (1991). "Reflection of retrograde waves within the cochlea and at the stapes," *J. Acoust. Soc. Am.* **89**, 1290–1305.
- Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (2005). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," *J. Acoust. Soc. Am.* **118**, 2434–2443.
- Sisto, R., and Moleti, A. (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. (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.
- 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.
- 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.

Comparing stimulus-frequency otoacoustic emissions measured by compression, suppression, and spectral smoothing

Radha Kalluri

Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye & Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02138

Christopher A. Shera^{a)}

Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye & Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114, Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02138 and Department of Otology & Laryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 25 June 2007; revised 28 August 2007; accepted 29 August 2007)

Stimulus-frequency otoacoustic emissions (SFOAEs) have been measured in several different ways, including (1) nonlinear compression, (2) two-tone suppression, and (3) spectral smoothing. Each of the three methods exploits a different cochlear phenomenon or signal-processing technique to extract the emission. The compression method makes use of the compressive growth of emission amplitude relative to the linear growth of the stimulus. The emission is defined as the complex difference between ear-canal pressure measured at one intensity and the rescaled pressure measured at a higher intensity for which the emission is presumed negligible. The suppression method defines the SFOAE as the complex difference between the ear-canal pressure measured with and without a suppressor tone at a nearby frequency. The suppressor tone is presumed to substantially reduce or eliminate the emission. The spectral smoothing method involves convolving the complex ear-canal pressure spectrum with a smoothing function. The analysis exploits the differing latencies of stimulus and emission and is equivalent to windowing in the corresponding latency domain. Although the three methods are generally assumed to yield identical emissions, no equivalence has ever been established. This paper compares human SFOAEs measured with the three methods using procedures that control for temporal drifts, contamination of the calibration by evoked emissions, and other potential confounds. At low stimulus intensities, SFOAEs measured using all three methods are nearly identical. At higher intensities, limitations of the procedures contribute to small differences, although the general spectral shape and phase of the three SFOAEs remain similar. The near equivalence of SFOAEs measured by compression, suppression, and spectral smoothing indicates that SFOAE characteristics are not mere artifacts of measurement methodology. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2793604]

PACS number(s): 43.64.Jb, 43.64.Kc, 43.64.Bt, 43.58.Ry [BLM]

Pages: 3562–3575

I. INTRODUCTION

Although the sounds evoked from the ear by a single pure tone were the first otoacoustic emissions (OAEs) to be discovered (reviewed in Kemp, 1998), they are, ironically, also the most difficult to measure. The difficulty arises because unlike other OAEs, stimulus-frequency emissions (SFOAEs) overlap with the evoking stimulus in both time and frequency. Unlike emissions evoked by transient stimuli, such as clicks, SFOAEs occur simultaneously with the evoking tone, and they therefore cannot be separated from the stimulus simply by time windowing. And unlike distortion-product emissions (DPOAEs), SFOAEs occur at the same frequency as the stimulus, and they therefore cannot be disentangled from stimulus components of the pressure using Fourier analysis.

Seemingly inseparable from their evoking wave forms, SFOAEs just do not appear as physically compelling as other emissions. Whereas transient-evoked emissions present an intuitive, echo-like quality clearly evident in the response wave form, and whereas DPOAEs occur in concert with aural combination tones that can readily be heard, SFOAEs, by contrast, lack these reassuringly tangible features and have been relegated in reviews to little more than insubstantial “errors of extrapolation” (e.g., Patuzzi, 1996). The absence of standardized and commercially available procedures for measuring SFOAEs only multiplies these spectral misgivings. Since different studies use different techniques and results are rarely compared, skeptics might be forgiven for wondering whether SFOAE characteristics are more artifacts of methodology than the robust and independent properties of actual sounds emitted by the ear.

^{a)}Electronic mail: shera@epl.meei.harvard.edu

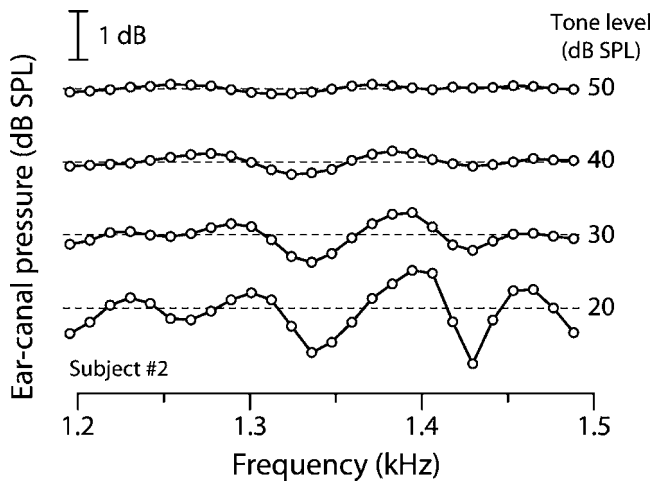


FIG. 1. Two components in the ear-canal pressure produced by a pure tone. The magnitude of the ear-canal pressure vs tone frequency at four different intensities is shown. Sound produced by the ear creates an intensity-dependent oscillatory fine structure that appears superposed on the constant stimulus background indicated by the dashed lines.

A. One stimulus tone, two response components

The classic evidence for the existence of SFOAEs comes from sweeping the frequency of a nominally constant-amplitude pure tone. In normal-hearing subjects one finds that the ear-canal pressure has the qualitative form illustrated in Fig. 1. Rather than the constant pressure anticipated on the basis of the high-level *in situ* calibration, two pressure components appear in the ear canal, distinguished by their behavior with frequency and intensity (e.g., Zwicker and Schloth, 1984; Shera and Zweig, 1993). At the highest levels (50 dB SPL in this example), the magnitude of the ear-canal pressure is, as expected, a smooth and almost constant function of frequency. But as the stimulus level is reduced, an oscillatory component, here with a period of roughly 100 Hz, appears superimposed on the constant “background.” Similar oscillatory patterns appear in the phase, albeit shifted relative to those in the magnitude by roughly 90°.

The measurements suggest that at fixed stimulus amplitude the total ear-canal pressure can be regarded as the sum of two complex components: A constant background component representing the nominal stimulus pressure (P_0) and a frequency-dependent component identified as the “stimulus-frequency emission” (SFE). The total ear-canal pressure thus has the form:

$$P_{\text{tot}}(f; P_0) = P_0 + \text{SFE}(f; P_0). \quad (1)$$

Phases are measured relative to the stimulus pressure, P_0 , which is defined to be real (the subscript zero is a mnemonic for “no emission”). The oscillatory structure apparent in Fig. 1 represents an interference pattern created as the relative phase of the two pressure components—stimulus and emission—rotates with frequency; spectral peaks and valleys occur at frequencies of maximal constructive and destructive interference. Physiologically labile and susceptible to suppression, the SFE component of the pressure originates within the cochlea through electrohydraulic processes collectively dubbed “stimulus reemission” (e.g., Kemp and Chum, 1980; Guinan, 1990).

B. Three measurement methods

According to Eq. (1), measuring the stimulus-frequency emission pressure $\text{SFE}(f; P_0)$ requires determining both the total pressure, $P_{\text{tot}}(f; P_0)$, and the stimulus, P_0 , at the probe frequency, f . The three measurement methods alluded to in the section heading refer to three different ways—compression, suppression, and spectral smoothing—of determining the stimulus pressure, P_0 . Measurement of the total pressure is the same in all cases. If we let P_0^m denote the stimulus pressure measured using method m , then the corresponding SFOAE is obtained by complex (or vector) subtraction:

$$\text{SFE}_m = P_{\text{tot}} - P_0^m. \quad (2)$$

In sub- and superscripts the three methods are abbreviated “com,” “sup,” and “ss,” respectively. Each of the three methods determines the stimulus pressure by exploiting a different cochlear phenomenon or signal-processing technique. The following paragraphs introduce each method in turn.

1. Nonlinear compression

The nonlinear-compression method (e.g., Kemp and Chum, 1980) exploits the presumed compressive growth of SFOAE amplitude. Because the emission grows compressively while the stimulus grows linearly, the ratio of the two approaches zero at high stimulus levels. Thus, increasing the stimulus amplitude decreases the SFOAE contribution to the total pressure. An estimate, P_0^{com} , of the stimulus pressure at low levels suitable for use in Eq. (2) can then be obtained by linearly scaling down the total ear-canal pressure measured at high levels.

2. Two-tone suppression

The suppression method estimates the stimulus pressure by using cochlear two-tone suppression to remove the emission contribution to the total pressure. In this method, the stimulus pressure P_0^{sup} is obtained by measuring the total ear-canal pressure at the probe frequency in the presence of an additional “suppressor” tone, usually at a nearby frequency (e.g., Kemp and Chum, 1980; Guinan, 1990; Kemp et al., 1990; Shera and Guinan, 1999; Schairer et al., 2003; Neely et al., 2005; Siegel et al., 2005). The suppressor tone is assumed to substantially reduce or eliminate the emission evoked by the probe.

3. Spectral smoothing

The spectral smoothing method estimates the stimulus pressure using signal-processing techniques. In particular, the estimate P_0^{ss} is obtained by smoothing the total pressure to remove spectral oscillations resulting from interference between the stimulus and the emission. When the stimulus amplitude is constant and the calibration extremely stable, smoothing amounts to little more than drawing a horizontal line at the nominal stimulus amplitude (e.g., Fig. 1), and similarly for the phase. More generally, smoothing is accomplished by convolving the total pressure $P_{\text{tot}}(f; P_0)$ with a smoothing function (Shera and Zweig, 1993; Kalluri and Shera, 2001).

C. Overview

Although it is generally assumed that the three SFOAE measurement methods outlined earlier yield identical emissions, no equivalence has ever been established. As a result, it remains unclear whether the emission properties depend in important ways on the methods used to measure them. If, for example, SFOAEs measured using suppression were largely the result of nonlinear interactions between the probe and suppressor—interactions not present when other measurement methods are employed—then so-called “SFOAEs” might not even exist as a well-defined and unitary phenomena. To address these issues, this paper reports and compares human SFOAEs measured using the three different methods. In a nutshell, we find that the SFOAEs measured using the different methods are nearly identical.

II. METHODS

A. Stimulus delivery and acquisition

Signals were delivered and recorded in one ear each of four normal-hearing human subjects while they were comfortably seated in a sound-isolated chamber (Ver *et al.*, 1975). All procedures were approved by human studies committees at the Massachusetts Eye and Ear Infirmary and the Massachusetts Institute of Technology. Stimulus wave forms were generated and responses acquired and averaged digitally using a custom data-acquisition system. The system consists of a National Instruments 4461 data-acquisition board controlled by custom software written in LABVIEW and MATLAB, two Shure E2c earphones, an Etymotic Research ER10c preamplifier, a Knowles EK3103 microphone, and a custom-built sound-delivery apparatus. We used Shure E2c insert earphones because, unlike ER10c earphones (cf. Schairer *et al.*, 2003), their response at the fundamental frequency grows nearly linearly at levels below 75 dB SPL. Additionally, their compact design made them easy to adapt for insertion into the human ear canal.

All stimuli were generated digitally using a fixed sampling rate of 48 kHz and discrete Fourier transform lengths of $N=4096$ points (85.33 ms), resulting in a digital frequency resolution of approximately 12 Hz. Stimulus wave forms were ramped on and off with 5 ms half-Blackman windows. After digital-to-analog conversion, all stimuli were high-pass filtered with a cutoff frequency of 0.15 kHz before presentation to the earphone. Acquired wave forms were filtered using a bandpass filter with cutoff frequencies of 0.15 and 15 kHz. Potential noise artifacts in the filtered wave forms were detected by comparing the rms difference between a stored artifact-free reference wave form and the current data buffer against a subject-dependent threshold. When the difference exceeded the rejection threshold, the response wave form in the newly acquired data buffer was discarded; otherwise the wave form was added to the averaging buffer. Ongoing replacement of the reference buffer minimized the effects of slowly varying drifts in the baseline signal.

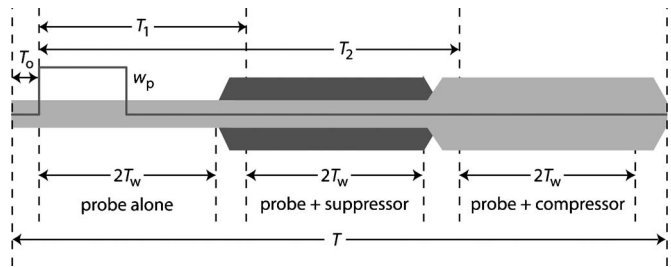


FIG. 2. The interleaved three-interval paradigm. The probe-alone, probe + suppressor, and probe + compressor segments each have duration $2T_w$. Intervening segments provide time for ramping the stimuli on and off and allowing the response to settle. The entire three-interval stimulus repeats with period $T=7T_w$. Wave forms are extracted using rectangular windows; only the window $w_p^{(0)}(t)$ is shown.

B. The interleaved three-interval paradigm

Except when noted, the stimuli needed to obtain data for all three measurement methods were presented in three interleaved segments, as illustrated in Fig. 2. Interleaving, previously employed for measuring SFOAEs using suppression (e.g., Shera and Guinan, 1999; Kalluri and Shera, 2001), minimizes the effects of time-dependent drifts. The three intervals are defined as follows: (1) The first (or “probe-alone”) segment contains a single tone (the probe) at frequency f_p and level L_p . During this segment the ear-canal pressure at the probe frequency contains both the stimulus and the emission. We used probe levels ranging from 20 to 50 dB SPL. (2) The second (or “probe+suppressor”) segment contains both the ongoing probe tone and an additional suppressor tone at a slightly lower frequency ($f_s \cong f_p - 46.9$ Hz). We typically used suppressor levels L_s between 55 and 60 dB SPL. The suppressor is assumed to substantially reduce or eliminate the emission at the probe frequency. To minimize the effect of earphone nonlinearities we presented the probe and suppressor tones using different sound sources. (3) The third (or “probe+compressor”) segment contains the probe and a compressor tone of level L_c at the same frequency as the probe. Probe and compressor have the same phase and were generated as a single tone by one transducer; the level of the two tones together is L_{pc} . Because the OAE grows compressively, the amplitude of the low-level stimulus can be estimated by scaling the pressure measured during this third segment.

The probe-alone, the probe+suppressor, and probe + compressor wave forms [denoted $p_p(t)$, $p_{ps}(t)$, and $p_{pc}(t)$, respectively] were obtained from the measured ear-canal pressure, $p(t)$, by averaging the two N -point subbuffers within each of the three segments. The two sub-buffers (indexed $n=0,1$) were extracted from $p(t)$ using windows $w_p^{(n)}(t)$, $w_{ps}^{(n)}(t)$, and $w_{pc}^{(n)}(t)$. For example, the probe-alone pressure was computed as

$$p_p(t) = \frac{1}{2} \sum_{n=0}^1 w_p^{(n)}(t) p(t). \quad (3)$$

Analogous equations were used for the probe+suppressor and probe+compressor wave forms. The windows are N -point rectangular boxcars of duration T_w :

$$w_p^{(n)}(t) = w_{\text{box}}(t - T_o - nT_w; T_w), \quad (4)$$

$$w_{\text{ps}}^{(n)}(t) = w_{\text{box}}(t - T_o - T_1 - nT_w; T_w), \quad (5)$$

and

$$w_{\text{pc}}^{(n)}(t) = w_{\text{box}}(t - T_o - T_2 - nT_w; T_w), \quad (6)$$

where the boxcar window is defined by

$$w_{\text{box}}(t; T_w) = \begin{cases} 1 & \text{for } 0 \leq t \leq T_w \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

The window offsets T_o , T_1 , and T_2 were chosen to allow time for the system to return to steady state after ramping the suppressor and compressor tones on or off using 5 ms half-Blackman windows. With our parameters, settling times were at least 16 ms, or somewhat longer than the maximum emission latencies for the frequencies we tested (e.g., [Shera and Guinan, 2003](#)). The window duration $T_w = N\Delta t$, where N is the buffer size and Δt is the sampling period, was set to match the length of the Fourier analysis buffer. Stimulus frequencies were chosen so that the analysis buffer always contained an integral number of cycles of both probe and suppressor. The entire three-interval stimulus was repeated continuously with period T until the desired number of artifact-free responses had been obtained (typically, at least 32, yielding a measurement noise floor of approximately -25 dB SPL). For the measurements reported here, the set of parameters $\{T_o, T_1, T_2, T\}$ had the values $\{\frac{1}{4}, \frac{9}{4}, \frac{9}{2}, 7\} \times T_w$.

At any given frequency, the interleaved three-interval paradigm collects the data necessary for all three SFOAE measurement method quasismultaneously. From the three averaged wave forms, $p_p(t)$, $p_{\text{ps}}(t)$, and $p_{\text{pc}}(t)$, we obtain both the total complex pressure,

$$P_{\text{tot}}(f) = P_p(f) \equiv F\{p_p(t)\}, \quad (8)$$

and the three estimates of the stimulus pressure:

$$P_0^{\text{sup}}(f) = F\{p_{\text{ps}}(t)\}e^{i2\pi fT_1}, \quad (9)$$

$$P_0^{\text{com}}(f) = F\{p_{\text{pc}}(t)\}10^{(L_p - L_{\text{pc}})/20}e^{i2\pi fT_2}, \quad (10)$$

$$P_0^{\text{ss}}(f) = P_{\text{tot}}(f) * S(f). \quad (11)$$

All expressions are evaluated at the probe frequency ($f = f_p$). In these equations $F\{\cdot\}$ denotes the N -point discrete Fourier transform, an asterisk (*) indicates convolution, and $S(f)$ is the smoothing filter. The complex phasors compensate for the ongoing phase shift of the probe stimulus, which was always adjusted to cosine phase in the ear canal at time $t = T_o$. The total pressure and the three estimates of the stimulus pressure were used in Eq. (2) to compute the corresponding SFOAEs.

Although most of the data reported here were collected using the interleaved three-interval paradigm, when comparing suppression and smoothing we sometimes expedited the measurements by eliminating the compressor segment of the stimulus, thereby adopting an abbreviated two-interval paradigm. We observed no systematic differences between the emissions extracted using the two- and three-interval paradigms.

C. Measurement details

1. Nonlinear compression

The compression technique relies on the linearity of the measurement system. For example, if the earphone output were to depend nonlinearly on the driving voltage, then linearly rescaling the ear-canal pressure measured at high levels would not accurately approximate the stimulus pressure at low levels. The resulting errors of extrapolation would manifest themselves as spurious otoacoustic responses. To circumvent the earphone nonlinearities that plague the Etymotic ER10c,¹ we designed a sound-delivery system using Shure E2c sources. The Shure earphones have an acoustic output at the fundamental frequency that grows quite linearly with driving voltage at ear-canal stimulus levels below 75 dB SPL. We evaluated the linearity of our measurement system by measuring SFE_{com} in a cavity with acoustic impedance comparable to that of the ear canal. Artifactual “emissions” measured in the cavity were at or below the noise floor for all the measurements we report.

2. Two-tone suppression

Because maximal suppression in humans occurs at suppressor frequencies close to the probe (e.g., [Brass and Kemp, 1993, 1991](#); [Backus, 2005](#)), we fixed f_s at about 47 Hz below f_p (a low-side suppressor). At the close spacing and relatively high suppressor levels used here, differences between high- and low-side suppressors are typically small. Measurements we made in two subjects at suppressor levels of 60 dB SPL showed that decreasing the frequency separation between the probe and suppressor produced no significant difference in SFOAE level. These results are consistent with those of [Backus \(2005\)](#), who also measured the frequency dependence of suppression in several subjects.

We used suppressor levels chosen to remove all or most of the SFOAE. Figure 3 shows how the magnitudes of SFE_{sup} and SFE_{com} measured at various probe levels depend on the level of the suppressor and compressor. The data indicate that emission levels plateau at suppressor/compressor levels roughly 20 dB greater than the probe, indicating that such levels are generally sufficient to remove most of the evoked emission. In the majority of our measurements we used 55 or 60 dB SPL suppressors and compressors. Although 60 dB SPL suppressors and compressors evidently eliminate most of the SFOAE at probe levels of 40 dB SPL and below, they are not as effective with probe levels of 50 dB SPL. Although more intense suppressors would remove more of the emission at the higher probe levels, they are also more likely to evoke unwanted efferent responses ([Guinan et al., 2003](#)).

3. Spectral smoothing

To implement spectral smoothing we applied the convolution theorem (e.g., [Papoulis, 1962](#)) to convert the convolution in Eq. (11) into an equivalent multiplicative windowing in the time domain.² In particular, we computed $P_0^{\text{ss}}(f)$ using

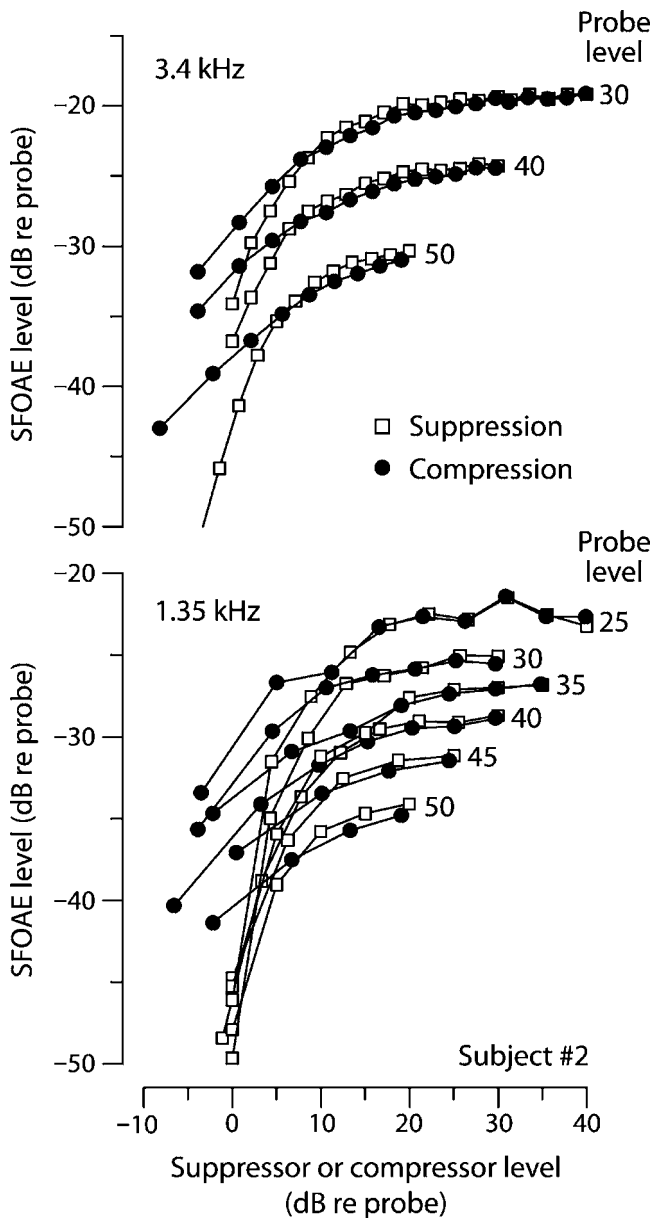


FIG. 3. Dependence of SFOAE magnitude on suppressor and compressor levels. Magnitudes of SFE_{sup} vs suppressor level L_s (open squares) and SFE_{com} vs compressor level L_c (closed circles) are shown. The values along both axes are expressed in decibels (dB) relative to the probe, whose level is varied parametrically and given in dB SPL adjacent to each curve. The top and bottom panels show results at two different probe frequencies (3.4 and 1.35 kHz) in the same subject. At suppressor and compressor levels comparable to or less than L_p , SFOAE magnitudes depend strongly on L_s and L_c and are larger when measured using compression. At levels L_s and L_c slightly above L_p , the SFE_{sup} and SFE_{com} curves cross into a regime in which the larger SFOAEs are obtained using suppression. At suppressor and compressor levels much greater than the probe level, relative emission magnitudes plateau, and the two methods yield nearly identical results.

$$P_0^{SS}(f) = \mathcal{F}^{-1}\{\mathcal{F}\{P_{tot}(f)\}\hat{S}(\tau)\}, \quad (12)$$

where the window $\hat{S}(\tau)$ is defined by $\hat{S}(\tau) \equiv \mathcal{F}\{S(f)\}$. In this equation $\mathcal{F}\{\cdot\}$ represents Fourier transformation with respect to a dimensionless log-frequency coordinate (Zweig and Shera, 1995; Knight and Kemp, 2000; Kalluri and Shera, 2001). (Note that the transform $\mathcal{F}\{\cdot\}$ differs from the discrete time Fourier transform, $F\{\cdot\}$, used to compute the complex

pressure from the averaged time wave forms.) The resulting Fourier-conjugate variable, τ , represents the emission latency in periods of the stimulus frequency. Fourier transformation with respect to the logarithmic frequency coordinate results in narrower, more well-defined peaks that prove more amenable to separation by windowing in the latency domain. To provide sharp window boundaries while avoiding excessive ringing in the frequency response we used a tenth-order recursive-exponential window for $\hat{S}(\tau)$ (Shera and Zweig, 1993). As detailed in footnote 10 of Kalluri and Shera (2001), the recursive-exponential window depends on the parameter τ_{cut} , which controls the window duration (or, equivalently, the band-width of the smoothing function). In the measurements reported here we took $\tau_{cut}=5$ stimulus periods, a duration sufficient to separate stimulus and emission at all stimulus levels used.

In practice, measurements are only available over a finite frequency range, and the smoothing operation is therefore complicated by end effects. To minimize these problems the analyzed frequency range was chosen to include an approximately integral number of spectral fine-structure cycles, and smoothing was performed using periodic boundary conditions (i.e., the complex pressure data were effectively wrapped around a cylinder by using the Fourier transform $\mathcal{F}\{\cdot\}$). When necessary, linear ramps were subtracted and subsequently restored after smoothing, to remove discontinuities in the real and imaginary parts of the pressure at the “seam.” Our need to subtract linear ramps was substantially reduced by using the repeated-calibration procedure (described in the following) to remove time-dependent drifts in the background pressure.

D. Minimizing spurious spectral structure

1. The repeated-calibration procedure

In previous work (e.g., Kalluri and Shera, 2001, 2007), we calibrated the insert earphones *in situ* immediately before and after each measurement series, defined as a collection of averaged measurements at multiple frequencies and/or levels. Comparisons of the pre- and postseries calibrations indicate that the in-the-ear calibrations are generally quite stable, at least in the absence of obvious movement of the earphone within the canal. Although this “single-calibration” procedure works well for measurements employing the suppression and compression methods, it proved inadequate for the careful cross-frequency comparisons we wished to make using spectral smoothing. The problem is that the spectral smoothing method assumes that all measured spectral structure removed by the smoothing filter represents an actual SFOAE. This assumption about the origin of the point-to-point spectral structure works well when the emission is large compared to artifactual changes in stimulus pressure (e.g., due to small movements of the probe, fluctuations in middle-ear cavity pressure and impedance, etc.). But when the emissions are relatively small (e.g., at higher stimulus levels) or the artifacts unusually large (e.g., in restless subjects who jostle the probe), the assumption often breaks down. When the assumption fails, emissions obtained by smoothing contain spurious components.

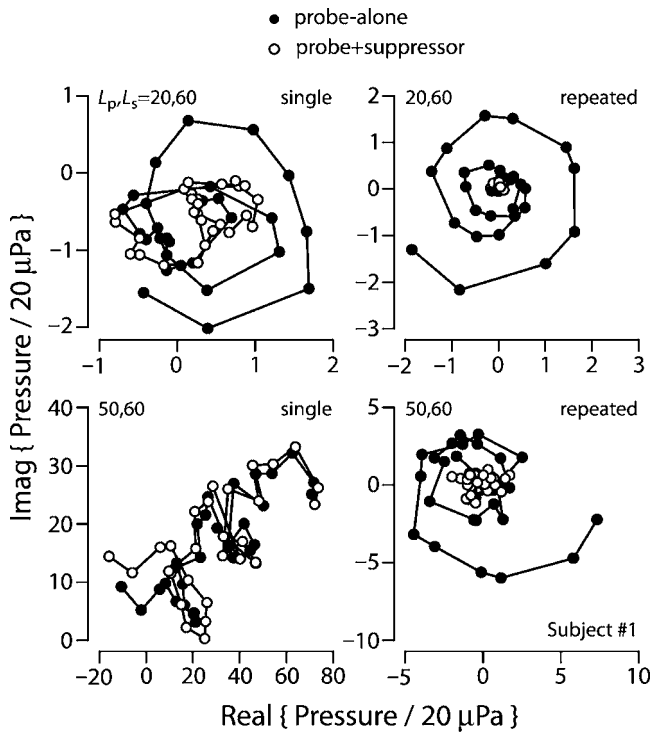


FIG. 4. Reducing spurious spectral structure by repeated calibration. The left-hand panels show polar plots of the probe-alone (closed dots) and probe+suppressor pressures (open dots) measured using the single-calibration procedure at two different probe levels (20 and 50 dB SPL, top and bottom, respectively). The right-hand panels show the same quantities measured using the repeated-calibration procedure. Pressures are shown in units of $20 \mu\text{Pa}$. The nominal stimulus pressures were subtracted off to center the data near the origin. Suppressor and calibration levels were fixed at 60 dB SPL. The data are from subject 1 at probe frequencies ranging from 3 to 3.5 kHz.

To reduce artifactual contamination of the SFOAEs measured using spectral smoothing, we adopted a “repeated calibration” procedure in which we recalibrated the probe earphones immediately before making measurements at each new frequency or level. Recalibrations were based on the ear-canal pressures produced by simultaneous presentation of high-level probe and suppressor tones (usually at levels near 60 dB SPL). When implemented using repeated calibrations, a measurement series consisting of ten probe frequencies involves eleven calibrations: one initial calibration performed using broad-band chirps, and ten repeated calibrations, one before each measurement, performed using pure tones.

Figure 4 illustrates the benefits of repeated calibration using data collected at two different probe levels. The left-hand panels show the probe-alone and probe+suppressor pressures $[P_p(f)$ and $P_{ps}(f)]$ measured when using the single-calibration procedure; the right-hand panels show the same pressures measured using repeated calibration. The data are shown as a function of frequency on polar plots with real and imaginary axes (closed dots show P_p ; open dots show P_{ps}). Interpreted using Eq. (1), previous SFOAE measurements (e.g., Shera and Guinan, 1999; see also Fig. 1) suggest that the probe-alone pressure, $P_p(f)$, which contains both stimulus and emission, generally rotates clockwise around the almost constant $P_{ps}(f)$. (We subtracted the nominal stimulus pressure from each data set so that the trajectories center

close to the origin.) The data collected using the carefully controlled repeated-calibration procedure show precisely this pattern.

In the single-calibration data, however, the pattern is obscured, especially at the higher probe level (lower left), where relative emission magnitudes are small. Rather than clustering at or rotating around the origin, the high-level, single-calibration data form a complex pattern more reminiscent of a random walk than a spiral. Although the $P_p(f)$ and $P_{ps}(f)$ trajectories still spiral about one another, their relative motion is masked by an irregular shared translational component. The additional structure in the data arises from small variations in the stimulus pressure, presumably due to intermittent shifts in probe placement and middle-ear impedance. Although negligible relative to the stimulus, the resulting pressure variations can be significant compared to the emission one is trying to estimate. As illustrated in the right-hand panels of Fig. 4, recalibrating the earphones before each measured frequency point all but eliminates these spurious sources of variation.

2. Calibration contamination by emissions

Although repeated calibrations largely eliminate one source of artifactual spectral structure, the calibration procedure itself can introduce another. The problem is that the calibration is necessarily done *in the ear*, so that the pressure measured during the calibration always contains both stimulus and emission. Since the emission component varies with frequency, it introduces spectral structure into the calibration that then contaminates the estimate of SFE_{ss} (and, to a much lesser extent, the emissions measured by compression and suppression).

We make these comments more precise by noting that the calibration procedure determines the earphone driving voltage needed to produce a constant pressure. The procedure assumes that the earphone is the only sound source in the system. Although this is true when calibrating in a test cavity, it is not true when calibrating in the ear. Because the calibration stimulus evokes an emission, the ear-canal pressure measured during the calibration contains both stimulus and emission pressures. In our procedure, the calibration is performed by presenting the probe and suppressor tones simultaneously. At the probe frequency, the total ear-canal pressure measured during the calibration is

$$P_{\text{tot}} = P_0^{\text{cal}} + SFE(P_0^{\text{cal}}; P_s^{\text{cal}}), \quad (13)$$

where superscripts “cal” identify probe and suppressor pressures used during the calibration. (To simplify the notation, we have left the dependence on probe frequency implicit.) The first term on the right is simply the probe pressure P_0^{cal} ; the second term is the SFOAE evoked by the probe in the presence of the suppressor, P_s^{cal} .

At each frequency, the value of P_{tot} measured during the calibration is used in subsequent measurements to determine the earphone driving voltage needed to produce a constant stimulus pressure. But because the calibration data are contaminated by SFOAEs, the stimulus pressure actually produced by the earphone will mistakenly be adjusted to com-

compensate for the sound produced by the ear. Thus, when the calibration data are used to produce a nominal pressure \bar{P}_0 , the actual stimulus pressure will be

$$P_0 \cong \bar{P}_0 - \frac{\bar{P}_0}{\bar{P}_0^{\text{cal}}} \text{SFE}(\bar{P}_0^{\text{cal}}; \bar{P}_s^{\text{cal}}), \quad (14)$$

where diacritical bars denote nominal (or target) pressures, and we have simplified by ignoring higher-order terms [e.g., we have replaced $\text{SFE}(P_0^{\text{cal}}; P_s^{\text{cal}})$ with $\text{SFE}(\bar{P}_0^{\text{cal}}; \bar{P}_s^{\text{cal}})$; the two pressures may differ very slightly if the earphone has shifted in the interval since the previous calibration]. In other words, the actual probe pressure produced by the earphone will differ from the target value by an amount proportional to the SFOAE evoked by the calibration stimulus.

When the stimulus pressure produced during a subsequent emission measurement is given by Eq. (14), the total ear-canal pressure measured during the probe-alone segment becomes

$$P_p = P_0 + \text{SFE}(P_0; 0) \cong \bar{P}_0 [1 - \text{SFE}(\bar{P}_0^{\text{cal}}; \bar{P}_s^{\text{cal}}) / \bar{P}_0^{\text{cal}}] + \text{SFE}(\bar{P}_0; 0). \quad (15)$$

To estimate the SFOAE obtained by smoothing, we simulate the smoothing operation by subtracting off the constant nominal stimulus pressure, \bar{P}_0 . The result is

$$\text{SFE}_{\text{ss}}(\bar{P}_0) \cong \text{SFE}(\bar{P}_0; 0) - \frac{\bar{P}_0}{\bar{P}_0^{\text{cal}}} \text{SFE}(\bar{P}_0^{\text{cal}}; \bar{P}_s^{\text{cal}}). \quad (16)$$

Because SFOAEs contaminate the calibration, the emission $\text{SFE}_{\text{ss}}(\bar{P}_0)$ obtained by smoothing differs from the true value [i.e., $\text{SFE}(\bar{P}_0; 0)$] by an amount that depends on the emission evoked by the calibration stimulus scaled by the ratio of probe and calibration pressures.

In contrast to the emissions obtained by smoothing, emissions obtained using the suppression method are, to first order, immune to calibration contamination. To see this, note that the pressure measured during the probe+suppressor segment is

$$P_{\text{ps}} = P_0 + \text{SFE}(P_0; P_s) \cong \bar{P}_0 [1 - \text{SFE}(\bar{P}_0^{\text{cal}}; \bar{P}_s^{\text{cal}}) / \bar{P}_0^{\text{cal}}] + \text{SFE}(\bar{P}_0; \bar{P}_s). \quad (17)$$

Computing SFE_{sup} by subtracting Eqs. (15) and (17) yields

$$\text{SFE}_{\text{sup}}(\bar{P}_0) \cong \text{SFE}(\bar{P}_0; 0) - \text{SFE}(\bar{P}_0; \bar{P}_s), \quad (18)$$

which is independent of the emission evoked by the calibration stimulus. Although emissions evoked during the calibration contaminate both P_p and P_{ps} , they do so symmetrically and cancel out in the difference.

Figure 5 corroborates our analysis by illustrating the effect of calibration contamination on measurements of SFE_{ss} and SFE_{sup} . The figure shows measurements performed in the three different calibration regimes suggested by Eq. (16):

- (1) Low-level calibration ($\bar{P}_0^{\text{cal}} \ll \bar{P}_0$). In this regime (top row), the emission evoked during the calibration procedure [i.e., the second term in Eq. (16)] generally domi-

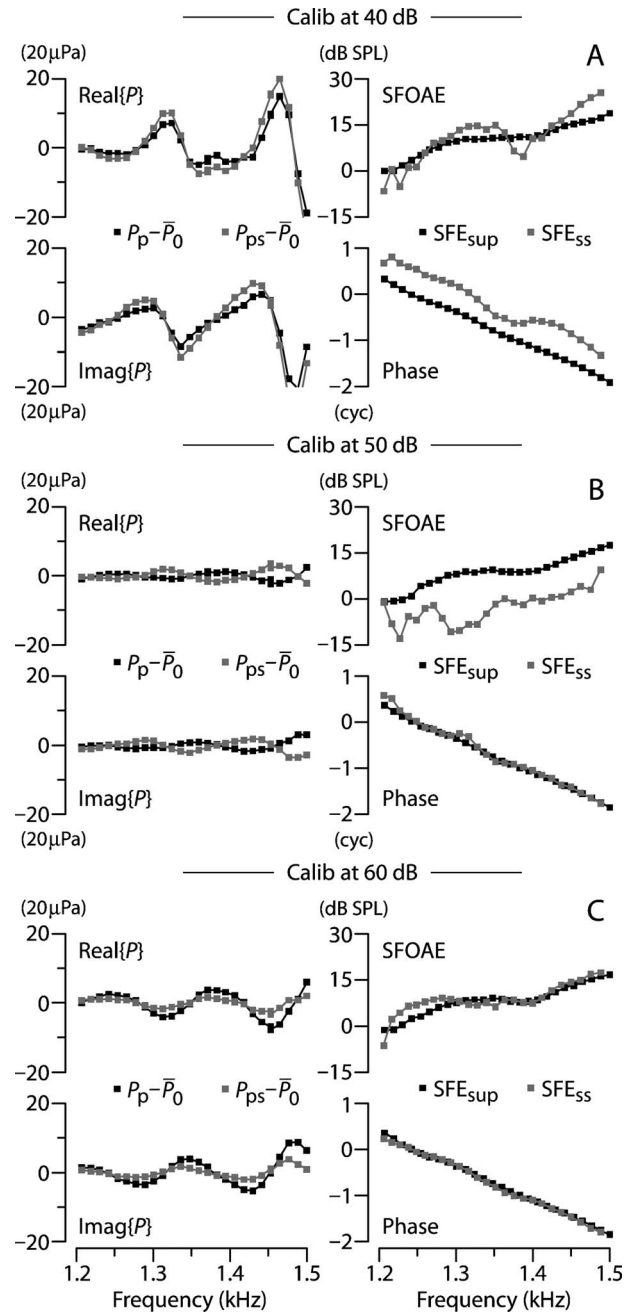


FIG. 5. Effects of calibration contamination by emissions. The three panels show measurements made at nominal stimulus intensities $\{L_p, L_s\} = \{50, 55\}$ dB SPL as determined by performing in-the-ear calibrations using equal-level probe and suppressor pressures as described in the text. The calibration levels used in panels (A), (B), and (C) were $L_{\text{cal}} = \{40, 50, 60\}$ dB SPL, respectively. In each panel, the left-hand column shows the real and imaginary parts of the probe-alone (P_p , black squares) and probe+suppressor pressures (P_{ps} , gray squares). To simplify the plotting, the nominal probe stimulus pressure (\bar{P}_0 , equivalent to 50 dB SPL) was subtracted from each pressure. The right-hand column shows the magnitude and phase of SFE_{sup} (black squares) and SFE_{ss} (gray squares) obtained from the data in the left column. Although the values of SFE_{sup} are nearly independent of calibration level, the values of SFE_{ss} vary systematically with L_{cal} and approximate SFE_{sup} only when calibrations are performed at levels much greater than the probe [e.g., panel (C)]. The data are from subject 2.

nates the value of SFE_{ss} . Because the calibration component contributes with an overall minus sign, the phase of SFE_{ss} appears shifted by roughly 180° from the SFOAE phase obtained using suppression.

- (2) Probe-level calibration ($\bar{P}_0^{\text{cal}} \cong \bar{P}_0$). In this regime (middle row), the two terms in Eq. (16) can have similar magnitudes but opposite signs, resulting in near cancellation. Cancellation is especially likely when the suppressor tone used during the calibration fails fully to suppress the SFOAE evoked by the probe [e.g., when $\bar{P}_s^{\text{cal}} \leq \bar{P}_0^{\text{cal}}$; see Fig. 3]. Our repeated-calibration procedure used equal level probes and suppressors and therefore falls in this category. Because of the cancellation, the emissions obtained using smoothing are substantially smaller than those obtained using suppression (for $\bar{P}_s \gg \bar{P}_0$).
- (3) High-level calibration ($\bar{P}_0^{\text{cal}} \gg \bar{P}_0$). In this regime (bottom row), contamination by the calibration term in Eq. (16) is small; smoothing and suppression therefore give similar results (for $\bar{P}_s \gg \bar{P}_0$).

Note that unlike the emissions obtained using smoothing, the emissions obtained using suppression remain largely insensitive to the relative intensity of the calibration stimulus, as predicted.

In summary, repeated calibration using high-level probe and suppressor tones minimizes spurious spectral structure that contaminates SFOAEs measured using the smoothing method. Unless otherwise noted, we obtained all SFE_{ss} measurements reported here by employing the repeated-calibration procedure implemented using equal-intensity high-level (60 dB SPL) probe and suppressor tones.³

III. RESULTS

Figure 6 illustrates our main result: When appropriate care is taken for the comparison, SFOAEs measured using compression, suppression, and spectral smoothing are almost identical. The two top panels show magnitudes and phases of $\text{SFE}_{\text{com}}(f)$, $\text{SFE}_{\text{sup}}(f)$, and $\text{SFE}_{\text{ss}}(f)$ measured at 30 dB SPL in subject 2. Closed and open symbols identify measurements made on different days using the standard three-interval paradigm and abbreviated two-interval paradigm, respectively. To reveal finer details of the phase obscured by the steep slope, the bottom panel replots the phase data after removing the overall trend. Except within the spectral notches, day-to-day variations found using the suppression method are on the order of 1 dB; they must to some extent arise from variations in the probe fit within the ear canal. Figure 6 demonstrates that SFOAEs measured using different methods interleaved in time during a single session are generally more similar to one another than SFOAEs measured using the same method (suppression) on different days. In all cases, however, the differences are minor.

Figures 7 and 8 corroborate these results and extend the comparison of compression, suppression, and spectral smoothing to different stimulus intensities and frequency ranges in two subjects. The top four rows in Fig. 7 show SFOAE measurements in subjects 1 and 2 at probe levels of 20, 30, 40, and 50 dB SPL, respectively. The bottom row shows detrended phases at 30 dB SPL. Figure 8 compares the SFOAEs measured in a higher frequency region in the

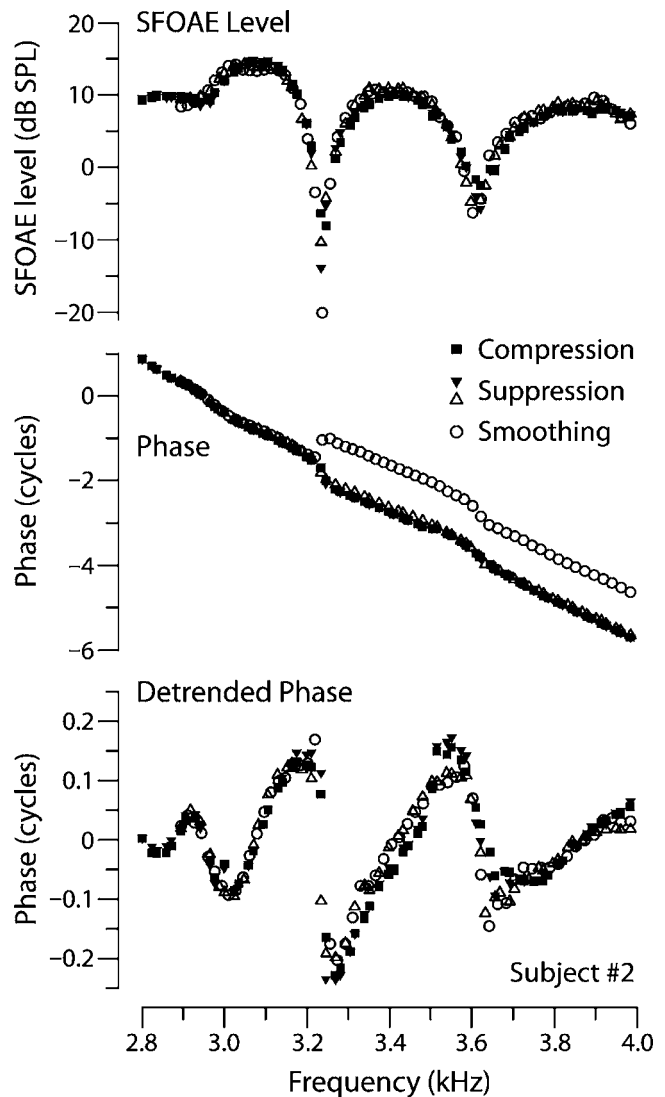


FIG. 6. Near-equivalence of SFOAEs measured using compression, suppression, and spectral-smoothing. The magnitudes (top), phases (middle), and detrended phases (bottom) of SFOAEs measured by compression (SFE_{com} , squares), suppression (SFE_{sup} , triangles), and spectral smoothing (SFE_{ss} , circles) in one subject are shown. Closed and open symbols identify measurements made on different days using the standard three-interval and the abbreviated two-interval paradigms, respectively. Above 3.2 kHz the phase obtained by smoothing differs from the others by approximately one full cycle because of variable phase unwrapping in the deep spectral notch. Detrended phases were computed from the unwrapped phases shown in the middle panel by subtracting a smooth trend that captures the large variation in phase. (Because of the variation in phase unwrapping, the SFE_{ss} trend was computed separately.) The data are from subject 2 at stimulus levels $\{L_p, L_s, L_{pc}\} = \{30, 60, 60\}$ dB SPL.

same two subjects. In all cases, differences between the methods are small, especially at the lower probe levels, and are generally smaller than the variations observed between sessions. Session-to-session differences can be especially large near spectral notches (e.g., see the detrended phase in Fig. 7, lower left); if the notches arise as the result of cancellation, as predicted (e.g., [Zweig and Shera, 1995](#)), then small changes can be expected to produce big effects.

Although the three measurement methods yield nearly equivalent results at low probe levels, systematic differences sometimes emerge at higher levels. Note, for example, the

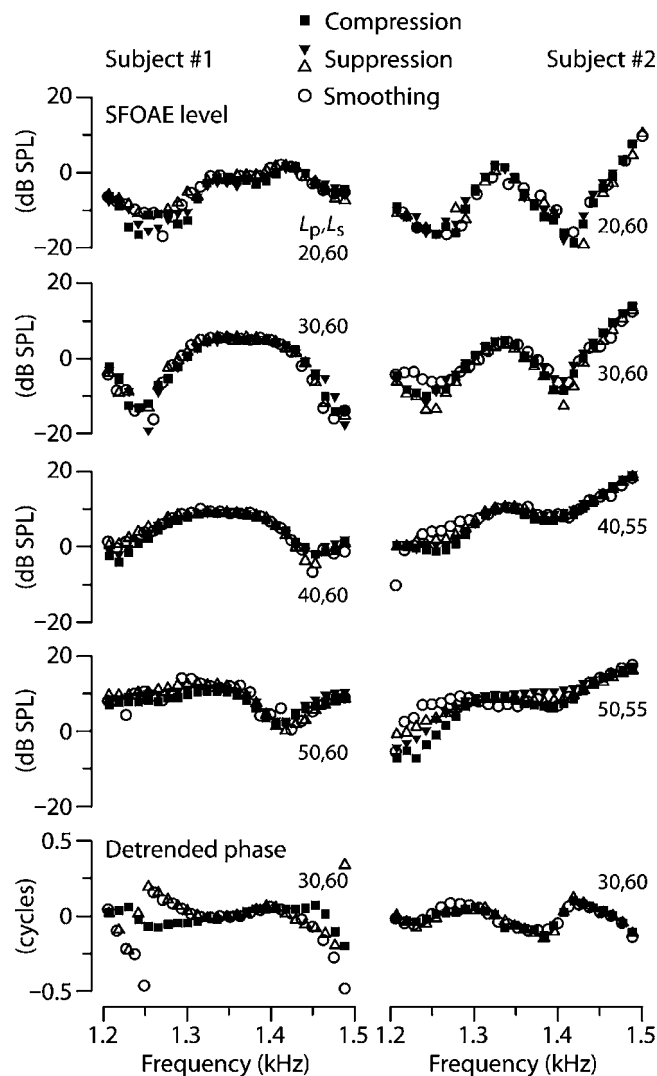


FIG. 7. SFOAEs measured using compression, suppression, and spectral smoothing at different probe levels. The magnitudes (top four rows) and detrended phases (bottom row) of SFOAEs measured in two subjects (left and right columns) using compression (SFE_{com} , squares), suppression (SFE_{sup} , triangles), and spectral smoothing (SFE_{ss} , circles) are shown. Open and closed symbols identify measurements made on different days using the standard three-interval and the abbreviated two-interval paradigms, respectively. Detrended phases, plotted modulo one cycle, were computed from the unwrapped phases by subtracting a smooth overall trend. The data are from subject 2 at levels L_p , L_s , and L_{pc} indicated adjacent to each curve (with $L_{pc}=L_s$).

difference between SFE_{com} and SFE_{sup} evident at 50 dB SPL in subject 2 (Figs. 7 and 8, right column, fourth row). At this probe level, the magnitude of SFE_{com} (closed squares) is systematically lower than that of SFE_{sup} (closed triangles). We suggest these high-level differences are entirely methodological: They arise because we used suppressor and compressor tones whose levels (55 or 60 dB SPL) were too low to remove the emission evoked by the 50 dB SPL probe. In Sec. IV we show that compressive nonlinearities driven by relatively weak stimuli produce systematic differences between suppression and compression similar to those seen here (see also Fig. 3).

Although somewhat noisier than the data in Figs. 7 and 8, SFOAE measurements performed in our two other sub-

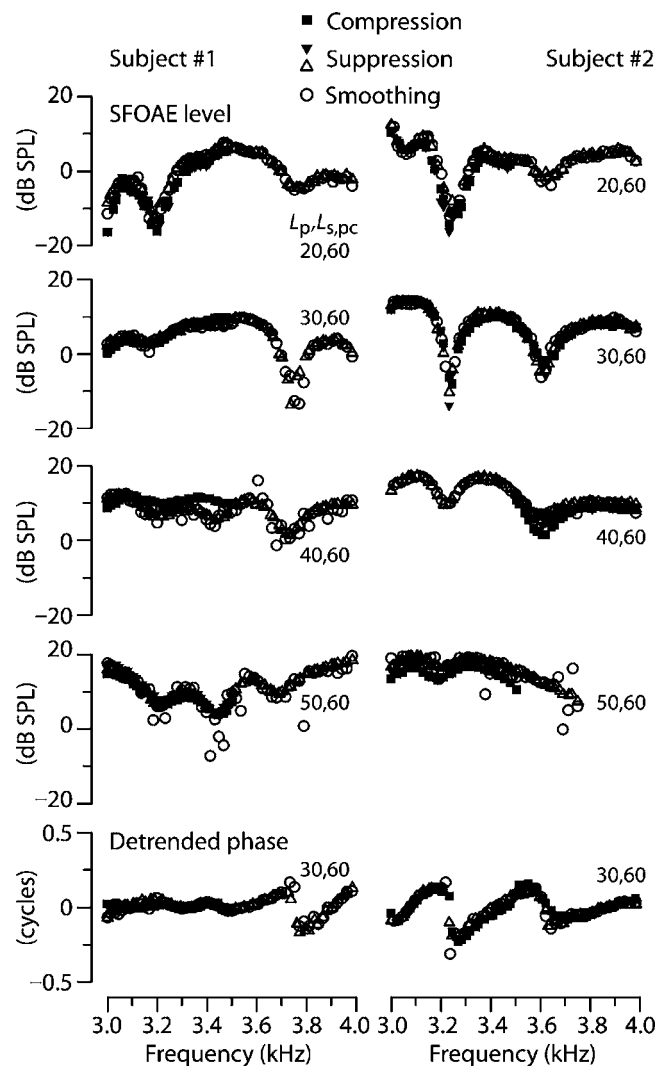


FIG. 8. SFOAEs measured using compression, suppression, and spectral smoothing at higher frequencies. The magnitudes (top four rows) and detrended phases (bottom row) of SFOAEs measured in a higher frequency region of the same subjects as Fig. 7 (left and right columns) are shown. The measurements were made using compression (SFE_{com} , squares), suppression (SFE_{sup} , triangles), and spectral smoothing (SFE_{ss} , circles). Open and closed symbols identify measurements made on different days using the standard three-interval and the abbreviated two-interval paradigms, respectively. The stimulus levels L_p , L_s , and L_{pc} are indicated adjacent to each curve (with $L_{pc}=L_s$). Note that not all measurements were made at all frequencies.

jects using the interleaved three-interval paradigm and repeated calibration are consistent with the results shown here.

IV. DISCUSSION

Our data demonstrate that human stimulus-frequency otoacoustic emissions measured using three different methods—compression, suppression, and spectral smoothing—are all nearly identical. At low and moderate levels, SFOAE magnitude and phase versus frequency functions obtained using the three methods almost overlie one another. When the measurements are made quasisimultaneously using the interleaved paradigm, differences between the SFOAEs obtained using the three methods are smaller than differences found using a single method in multiple sessions. Our results thus indicate that the properties of human

SFOAEs are independent of the methods used to measure them; SFOAEs are not artifacts of methodology.

Each of the three methods we employed exploits a different cochlear phenomenon or signal-processing technique to extract the emission: (1) The compression method makes use of the compressive growth of emission amplitude relative to the linear growth of the stimulus; (2) the suppression method uses cochlear two-tone suppression to reduce or eliminate the emission evoked by the probe; and (3) the smoothing method extracts the emission directly from the probe pressure using signal processing without the need to present additional stimuli. The only stimulus common to all three methods is the low-level probe. The equivalence we report therefore demonstrates that SFOAEs are due almost entirely to the *probe alone*; they are not created through some nonlinear interaction between the probe and the suppressor (or compressor).

We emphasize that SFOAEs measured using the three methods are not automatically equivalent. In order to convince ourselves that any differences we might find were real, we took considerable care controlling for potential confounds (e.g., by using interleaved measurements and repeated calibrations to minimize drifts, by choosing suppressor and compressor levels to remove most if not all of the emission, and by calibrating at levels sufficient to reduce emission contamination). Relaxing these procedures can readily introduce artifacts into the measurements that may masquerade as real differences between the methods.

A. Possible limits of validity

Although our conclusions may apply more widely, we summarize in the following the known limitations of our study. (1) Our comparisons required special care to avoid artifacts and were therefore performed in a relatively small number of subjects ($n=4$). Reassuringly, however, we found similar results in all ears. (2) All of our subjects had normal hearing. Although we hypothesize that similar results will be found whenever emissions remain measurable, additional studies are needed to determine whether our findings generalize to impaired ears. (3) We used low to moderate probe levels (20–50 dB SPL), and the three methods may yield different results at higher levels. Although evidence for differences at higher levels is apparent in our data (e.g., at 50 dB SPL), we believe these differences reflect a methodological limitation of our study, namely the use of suppressors and compressors limited to 60 dB SPL, an intensity too low to remove the entire emission at probe levels above about 40 dB SPL. (4) We used only near-probe suppressors (i.e., suppressors close to the probe frequency). Considerable evidence indicates that suppressors further than roughly half an octave from the probe yield very different results (Siegel *et al.*, 2003, 2004; Shera *et al.*, 2004; Backus, 2005). Differences can arise because more distant suppressors produce incomplete suppression and/or induce additional SFOAE sources. (5) Our comparisons are limited to the frequency range of 1–4 kHz. In particular, we did not explore the behavior in more apical regions of the cochlea, where both emission mechanisms and mechanical suppression and com-

pression may differ considerably from those in the base (Cooper and Rhode, 1995; Rhode and Cooper, 1996; Shera and Guinan, 2003; Siegel *et al.*, 2005; Shera *et al.*, 2007). (6) Finally, our measurements are in humans, a species whose OAE characteristics differ in some respects from those of many laboratory animals (e.g., humans have longer OAE latencies and smaller distortion-source emissions). The near-equivalence we find between SFOAE methods remains to be examined in other species.

B. Differences between compression and suppression

Figure 3 shows that SFOAE levels plateau at suppressor and compressor levels roughly 20 dB greater than the probe. At plateau levels, SFOAEs measured by suppression are nearly indistinguishable from those measured using compression. At lower suppressor and compressor levels, however, the measured behavior is more complicated. For example, SFOAEs obtained using suppression are generally larger than those obtained using compression ($|\text{SFE}_{\text{sup}}| > |\text{SFE}_{\text{com}}|$) at levels just below the plateau region. Since the suppression method produces the larger residual from Eq. (2), a nearby suppressor is evidently more effective at reducing emission amplitude than an on-frequency compressor of the same intensity. Suppressors remain more effective until their level becomes comparable to or lower than the probe, at which point the curves cross one another and the relative amplitudes of SFE_{sup} and SFE_{com} are reversed ($|\text{SFE}_{\text{com}}| > |\text{SFE}_{\text{sup}}|$).

This general pattern of the relative efficacy of suppressor and compressors can be modeled using a simple, saturating nonlinearity.⁴ Figure 9 shows results obtained by using the hyperbolic tangent to simulate the SFOAE measurements shown in Fig. 3. The open squares show the results obtained using the suppression method. The data points give the amplitude of the probe-frequency Fourier component of the function

$$\text{SFE}_{\text{sup}}(t) = T_0 p_{\text{ref}} \{ \tanh[p_p(t)/p_{\text{ref}}] - \tanh[p_{\text{ps}}(t)/p_{\text{ref}}] \}, \quad (19)$$

where $p_p(t) = P_p \cos(2\pi f_p t)$ is the probe tone, $p_s(t) = P_s \cos(2\pi f_s t)$ is the suppressor, and $p_{\text{ps}}(t) = p_p(t) + p_s(t)$. The probe-frequency component of the first term models the SFOAE evoked by the probe tone alone; the same component of the second term models the SFOAE modified by the suppressor. (The probe stimulus itself contributes equally to the total pressure in both cases and therefore cancels in the difference.) The constants T_0 and p_{ref} were chosen to roughly approximate the overall magnitude and saturation characteristics apparent in the data. For comparison, the closed circles show results obtained using the compression method. The data points give the amplitude of the probe-frequency component of the function

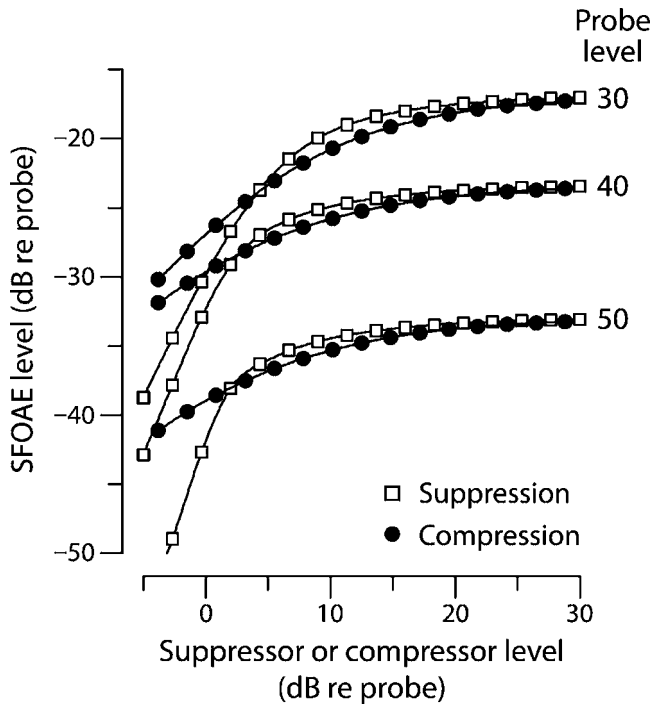


FIG. 9. Simulated SFOAEs produced by a simple, saturating nonlinearity. The dependence of SFOAE magnitude on suppressor and compressor levels is shown in a format similar to Fig. 3. The open squares give magnitudes of SFE_{sup} vs suppressor level L_s computed from Eq. (19); the closed circles give SFE_{com} vs compressor level L_c computed from Eq. (20). All emission and stimulus levels are expressed in decibels relative to the probe amplitude, P_p . The model parameters are $p_{ref} = \text{adB}(30)$ and $T_0 = \text{adB}(-15)$, where the function $\text{adB}(x) = 10^{x/20}$.

$$SFE_{com}(t) = T_0 p_{ref} \left\{ \tanh[p_p(t)/p_{ref}] - \frac{P_p}{P_{pc}} \tanh[p_{pc}(t)/p_{ref}] \right\}, \quad (20)$$

where $p_{pc}(t) = (P_p + P_c) \cos(2\pi f_p t)$ is the sum of probe and compressor.

The simulated SFOAEs computed from Eqs. (19) and (20) manifest the same qualitative behavior found in the data (Fig. 3), including SFE_{sup} and SFE_{com} curves that cross at levels L_s and L_c slightly greater than L_p before asymptoting to the same plateau value at higher levels. The larger emissions obtained using compressors at low levels can be understood by noting that because the compressor is always in-phase with the probe, it produces a greater overall reduction in probe amplitude than does an off-frequency suppressor of the same intensity. Because probe and suppressor are at different frequencies, their relative phase varies over time. Whereas the compressor produces a maximal reduction in the probe twice per cycle, the suppressor has an equivalent effect only on those (relatively rare) occasions when probe and suppressor reach their extrema simultaneously. Of course, the same reasoning implies that $|SFE_{com}|$ should always exceed $|SFE_{sup}|$ when using equal-level compressors and suppressors. To account for the crossover into a regime where suppressors yield the greater residual, we hypothesize that at higher levels, where phase effects are less important and intermodulation components are larger, the suppressor

produces a greater reduction in probe amplitude by diverting probe energy into distortion components (e.g., $2f_s - f_p$) that are not produced when using on-frequency compressors.

C. Pros and cons of the three methods

Each of the three SFOAE measurement methods has advantages and disadvantages. Although the compression method (Kemp and Chum, 1980) has become the de facto “gold-standard,” the method requires an effectively linear sound generation and recording system, requirements not often met by commercial OAE recording equipment. The suppression method is substantially more forgiving in this regard, at least when the probe and suppressor are generated using separate earphones. An advantage of the spectral-smoothing method is that it requires measurement of only a single quantity (namely, the probe-alone pressure) whereas the other methods require an additional measurement (i.e., the response to either the compressor or suppressor). Unlike the compression and suppression methods, the smoothing method therefore allows each probe measurement to serve as its own control against possible systematic changes (e.g., variations in overall emission level due to efferent effects) that may occur during the course of the measurement. In the studies reported here, we sought to minimize these potential problems by interleaving the probe, suppressor, and compressor in time using the three-interval paradigm. Although the spectral smoothing method depends on only the probe-alone pressure, it requires measurements at multiple frequencies. Indeed, the method works best if applied to measurements that span a relatively wide frequency range (i.e., many periods of the interference microstructure) with good frequency resolution (i.e., many points per period). In addition, because of uncertainties introduced near the end points due to incomplete knowledge outside the measured interval, the smoothing method ideally requires measurements over an interval slightly larger than the desired frequency range. The compression and suppression methods, by contrast, impose no such constraints; they require measurements only at the actual frequency (or frequencies) of interest. Finally, the compression and suppression methods are more robust to shifts in probe placement, calibration contamination, and drift, all of which can introduce spurious spectral structure detrimental to the smoothing method, especially when emission amplitudes are relatively small (e.g., at higher probe levels).

D. Other methods of measuring SFOAEs

The three methods compared here are not the only ways to measure SFOAEs. Emissions obtained using methods based on other cochlear phenomena have been directly compared with SFOAEs obtained using suppression and/or compression. These methods involve:

- (1) Measuring the reflection-source component of DPOAEs. In previous work undertaken to test the “two-mechanism” model of DPOAE generation, we extracted the reflection-source (or “DP-place”) component of $2f_1 - f_2$ DPOAEs using both suppression- and smoothing-based methods (Kalluri and Shera, 2001). As predicted

by the model, the extracted DPOAE component matched the SFOAE measured at the same frequency. The match was best when the SFOAE was measured in the presence of an “ f_1 -primary mimicker” (i.e., an additional tone presented at the frequency and level of the f_1 primary used to evoke the DPOAEs). The correspondence between the two emissions suggests that although they may be partially suppressed by the primary tones, reflection-source components of DPOAEs are otherwise largely equivalent to SFOAEs.

- (2) Measuring click-evoked OAEs (CEOAEs). In order to understand the influence of stimulus bandwidth on OAE generation, we compared click-evoked and stimulus-frequency OAEs measured in the same human subjects as a function of stimulus frequency and intensity (Kalluri and Shera, 2007). We found that CEOAEs and SFOAEs are nearly identical when compared at “bandwidth-compensated” sound-pressure levels. These results demonstrate the two emission “types” are generated by the same mechanism; at low and moderate stimulus intensities CEOAEs and SFOAEs are really the same emission evoked in different ways. Measuring CEOAEs therefore provides an alternative, and generally more efficient, means of measuring SFOAEs.
- (3) Killing the subject. This method estimates the stimulus pressure P_0 in Eq. (1) by exploiting the physiological vulnerability of SFOAE generation. Although unforgiving and often prohibited, the method has been attempted in chinchilla, where it yields SFOAEs in good quantitative agreement with those measured using suppression (Siegel, 2004).

Although future work may reveal important quantitative differences, several other methods also yield emissions that appear at least qualitatively similar to the SFOAEs obtained here.⁵ These methods include:

- (1) Measuring upper-sideband DPOAEs (e.g., $2f_2 - f_1$). Knight and Kemp (1999) report strong similarities between CEOAEs and upper-sideband DPOAEs, as well as lower-sideband DPOAEs measured at f_2/f_1 ratios close to 1. The near-equivalence between CEOAEs and SFOAEs (Kalluri and Shera, 2007) suggests a similar correspondence between SFOAEs and these special classes of DPOAEs.
- (2) Using efferent suppression. This method estimates the stimulus pressure P_0 in Eq. (1) by using olivo-cochlear efferent suppression to reduce the amplitude of the emission evoked by the probe (e.g., Guinan, 1990; Guinan et al., 2003). SFOAEs obtained in this way share many qualitative features with SFOAEs obtained using acoustic suppression (e.g., rapidly rotating phases). However, the two remain to be carefully compared and details may depend on how the efferent activity is elicited (e.g., Backus and Guinan, 2005).

E. Are SFOAEs really emissions?

Doubts about the ontological status of SFOAEs—Are they just artifacts of measurement methodology?—have but-

tressed the view, most clearly articulated by Patuzzi (1996), that SFOAEs really ought not be considered “emissions” at all. In this view, SFOAEs are better understood not as sounds produced by the cochlea but as the inevitable “error of extrapolation” that arises because the cochlear input impedance depends on sound intensity. Analogous errors occur, for example, when one linearly extrapolates the output of garden-variety earphones from one intensity to another based on a mistaken presumption of linearity. When the extrapolation fails, no one writes *JASA* papers describing how the earphone is “emitting an SFOAE;” they just say that the earphone has a nonlinear source impedance. So why treat the ear any differently from the earphone?

Although we hope the measurements reported here will allay outstanding methodological doubts about SFOAEs, the most compelling reasons why these otoacoustic “errors of extrapolation” are best regarded as actual emissions come from the characteristics of the SFOAEs themselves. Because SFOAEs appear in the ear canal via their effects on the cochlear input impedance transmitted through the middle ear, the two complementary views of SFOAEs outlined earlier are, in fact, mathematically equivalent. Physically, however, the components of the cochlear input impedance ascribed to SFOAEs really do have features characteristic of emitted sounds: In particular, they look just like reflected waves. For example, at constant intensity one can write the (nonlinear) cochlear input impedance in the form⁶

$$Z_C(f; P_0) = Z_0(f) \frac{1 + R(f; P_0)}{1 - R(f; P_0)}, \quad (21)$$

where $Z_0(f)$ is the “error-free” input impedance (e.g., the input impedance measured at high stimulus levels) and $R(f; P_0)$ parametrizes the intensity-dependent “error” or “emission.”

Although there is nothing mathematically unique about this representation of the input impedance—one could write $Z_C(f; P_0)$ in a myriad other forms—the formulation in Eq. (21) is *natural* because the complex function $R(f; P_0)$ is empirically *simple*: To first approximation, $R(f; P_0)$ is a circle in the complex plane (Shera and Zweig, 1993). This simple form for $R(f; P_0)$ —a slowly varying amplitude and a rapidly rotating phase indicating a substantial delay—has a straightforward physical interpretation in terms of wave reflection within the cochlea (e.g., Zweig and Shera, 1995; Talmadge et al., 1998). Indeed, Eq. (21) with $R(f; P_0)$ resembling a delay has the same general form as the input impedance of a transmission line with internal reflections. When evoked using pure tones, these internal reflections are known as SFOAEs; when evoked using acoustic clicks, the same reflections are called CEOAEs (Kalluri and Shera, 2007). For these reasons, we refer to $R(f; P_0)$ not as an “error” but as the “cochlear reflectance.”

Furthermore, the physical interpretation as a reflectance has predictive power. For example, viewing SFOAEs as reflected waves (i.e., as emissions) immediately suggests a relationship, verified experimentally, between SFOAE phase and the frequency spacing between spontaneous emissions (Shera, 2003). No matter how one may choose to represent SFOAEs, spontaneous OAEs clearly represent actual “emis-

sions” from the ear. At the mathematical level the distinction between an “error of extrapolation” and an actual “emission”—between a nonlinear impedance and a nonlinear reflectance—is largely semantic; at the physical and physiological levels, however, the distinction is evidently both real and productive.

ACKNOWLEDGMENTS

We thank Christopher Bergevin, John Guinan, Jeffery Lichtenhan, and the anonymous reviewers for helpful comments on the manuscript. This work was supported by Grant No. R01 DC03687 from the NIDCD, National Institutes of Health.

¹Schairer *et al.* (2003) provide measurements that quantify ER10c earphone nonlinearities in the context of measuring SFOAEs.

²Interpreted in the time domain, spectral smoothing separates the stimulus and emission by exploiting the latency difference between the two signals. Although the method derives from linear-systems theory, the procedure makes no a priori assumptions about emission generation mechanisms. Techniques for analyzing OAEs in this way were introduced by Shera and Zweig (1991), who applied them to the study of SFOAEs; similar methods have since been applied to other emissions (e.g., Stover *et al.*, 1996; Brown *et al.*, 1996; Fahey and Allen, 1997; Knight and Kemp, 2000; Ren *et al.*, 2000; Kalluri and Shera, 2001; Goodman *et al.*, 2003).

³Problems with calibration contamination could have been further reduced by employing a higher level suppressor tone during the calibration.

⁴The saturating nonlinearity represented in Eqs. (19) and (20) models the effect of a suppressor or compressor on SFOAEs. The model takes for granted that the ear-canal pressure contains a level-dependent probe frequency emission component and explains neither how that component arises within the cochlea nor how it is transmitted to the ear canal; the model is therefore primarily a description of SFOAE measurement rather than SFOAE generation. Models of SFOAE generation can be found elsewhere (e.g., Zweig and Shera, 1995; Talmadge *et al.*, 1998, 2000).

⁵In addition to the more qualitatively different methods discussed here, variations on a theme within a given measurement category generally yield similar results [e.g., the use of amplitude-modulated suppressors (Neely *et al.*, 2005)].

⁶Equation (21) for the input impedance assumes the tapering symmetry found in the basal turn of the cat cochlea (Shera and Zweig, 1991).

Backus, B. C. (2005). “Using stimulus frequency otoacoustic emissions to study basic properties of medial olivocochlear reflex.” Ph.D. thesis, MIT, Cambridge.

Backus, B. C., and Guinan, J. J. (2005). “Do all normal-hearing humans have a medial-olivocochlear acoustic reflex response?” *Assoc. Res. Otolaryngol. Abstr.* **28**, 915.

Brass, D., and Kemp, D. T. (1991). “Time-domain observation of otoacoustic emissions during constant tone stimulation,” *J. Acoust. Soc. Am.* **90**, 2415–2427.

Brass, D., and Kemp, D. T. (1993). “Suppression of stimulus frequency otoacoustic emissions,” *J. Acoust. Soc. Am.* **93**, 920–939.

Brown, A. M., Harris, F. P., and Beveridge, H. A. (1996). “Two sources of acoustic distortion products from the human cochlea,” *J. Acoust. Soc. Am.* **100**, 3260–3267.

Cooper, N. P., and Rhode, W. S. (1995). “Nonlinear mechanics at the apex of the guinea-pig cochlea,” *Hear. Res.* **82**, 225–243.

Fahey, P. F., and Allen, J. B. (1997). “Measurement of distortion product phase in the ear canal of the cat,” *J. Acoust. Soc. Am.* **102**, 2880–2891.

Goodman, S. S., Withnell, R. H., and Shera, C. A. (2003). “The origin of SFOAE microstructure in the guinea pig,” *Hear. Res.* **183**, 1–17.

Guinan, J. J. (1990). “Changes in stimulus frequency otoacoustic emissions produced by two-tone suppression and efferent stimulation in cats,” in *Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer, New York), pp. 170–177.

Guinan, J. J., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (2003). “Medial olivocochlear efferent reflex in humans: Otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency

OAEs,” *J. Assoc. Res. Otolaryngol.* **4**, 521–540.

Kalluri, R., and Shera, C. A. (2001). “Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation,” *J. Acoust. Soc. Am.* **109**, 622–637.

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.

Kemp, D. T. (1998). “Otoacoustic emissions: Distorted echoes of the cochlea’s travelling wave,” in *Otoacoustic Emissions: Basic Science and Clinical Applications*, edited by C. I. Berlin (Singular, San Diego), pp. 1–59.

Kemp, D. T., Brass, D., and Souter, M. (1990). “Observations on simultaneous SFOAE and DPOAE generation and suppression,” in *Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer, New York), pp. 202–209.

Kemp, D. T., and Chum, R. A. (1980). “Observations on the generator mechanism of stimulus frequency acoustic emissions—Two tone suppression,” in *Psychophysical Physiological and Behavioural Studies in Hearing*, edited by G. V. D. Brink and F. A. Bilsen (Delft University Press, Delft), pp. 34–42.

Knight, R. D., and Kemp, D. T. (1999). “Relationships between DPOAE and TEOAE amplitude and phase characteristics,” *J. Acoust. Soc. Am.* **106**, 1420–1435.

Knight, R. D., and Kemp, D. T. (2000). “Indications of different distortion product otoacoustic emission mechanisms from a detailed f_1, f_2 area study,” *J. Acoust. Soc. Am.* **107**, 457–473.

Neely, S. T., Johnson, T. A., Garner, C. A., and Gorga, M. P. (2005). “Stimulus-frequency otoacoustic emissions measured with amplitude-modulated suppressor tones,” *J. Acoust. Soc. Am.* **118**, 2124–2127.

Papoulis, A. (1962). *The Fourier Integral and its Applications* (McGraw-Hill, New York).

Patuzzi, R. (1996). “Cochlear micromechanics and macromechanics,” in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 186–257.

Ren, T., Nuttall, A. L., and Parthasarathi, A. A. (2000). “Quantitative measure of multicomponents of otacoustic emissions,” *J. Neurosci. Methods* **96**, 97–104.

Rhode, W. S., and Cooper, N. P. (1996). “Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*,” *Aud. Neurosci.* **3**, 101–121.

Schairer, K. S., Fitzpatrick, D., and Keefe, D. H. (2003). “Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears,” *J. Acoust. Soc. Am.* **114**, 944–966.

Shera, C. A. (2003). “Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves,” *J. Acoust. Soc. Am.* **114**, 244–262.

Shera, C. A., and Guinan, J. J. (1999). “Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs,” *J. Acoust. Soc. Am.* **105**, 782–798.

Shera, C. A., and Guinan, J. J. (2003). “Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning,” *J. Acoust. Soc. Am.* **113**, 2762–2772.

Shera, C. A., Guinan, J. J., and Oxenham, A. J. (2007). “Otoacoustic estimates of cochlear tuning: Validation in the chinchilla,” *Assoc. Res. Otolaryngol. Abstr.* **30**, 519.

Shera, C. A., Tubis, A., Talmadge, C. L., and Guinan, J. J. (2004). “The dual effect of ‘suppressor’ tones on stimulus-frequency otoacoustic emissions,” *Assoc. Res. Otolaryngol. Abstr.* **27**, 538.

Shera, C. A., and Zweig, G. (1991). “A symmetry suppresses the cochlear catastrophe,” *J. Acoust. Soc. Am.* **89**, 1276–1289.

Shera, C. A., and Zweig, G. (1993). “Noninvasive measurement of the cochlear traveling-wave ratio,” *J. Acoust. Soc. Am.* **93**, 3333–3352.

Siegel, J. H. (2004). “Otoacoustic emissions: Have we been barking up the wrong tree?,” *Assoc. Res. Otolaryngol. Abstr.* **27**, 514.

Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (2005). “Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering,” *J. Acoust. Soc. Am.* **118**, 2434–2443.

Siegel, J. H., Cerka, A. J., Temchin, A. N., and Ruggero, M. A. (2004). “Similar two-tone suppression patterns in SFOAEs and the cochlear microphonics indicate comparable spatial summation of underlying generators,” *Assoc. Res. Otolaryngol. Abstr.* **27**, 539.

Siegel, J. H., Temchin, A. N., and Ruggero, M. A. (2003). “Empirical estimates of the spatial origin of stimulus-frequency otoacoustic emissions,”

- Assoc. Res. Otolaryngol. Abstr. **26**, 679.
- Stover, L. J., Neely, S. T., and Gorga, M. P. (1996). "Latency and multiple sources of distortion product otoacoustic emissions," *J. Acoust. Soc. Am.* **99**, 1016–1024.
- 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.
- Ver, I. L., Brown, R. M., and Kiang, N. Y. S. (1975). "Low-noise chambers for auditory research," *J. Acoust. Soc. Am.* **58**, 392–398.
- Zweig, G., and Shera, C. A. (1995). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," *J. Acoust. Soc. Am.* **98**, 2018–2047.
- Zwicker, E., and Schloth, E. (1984). "Interrelation of different oto-acoustic emissions," *J. Acoust. Soc. Am.* **75**, 1148–1154.

Wavelet and matching pursuit estimates of the transient-evoked otoacoustic emission latency

Giuseppe Notaro,^{a)} Adnan Mohsin Al-Maamury, and Arturo Moleti

Physics Department, University of Roma "Tor Vergata", Via della Ricerca Scientifica, 1, 00133 Roma, Italy

Renata Sisto^{b)}

Occupational Hygiene Department, ISPESL, Via Fontana Candida, 1, 00040 Monte Porzio Catone (Roma), Italy

(Received 20 July 2007; revised 26 September 2007; accepted 26 September 2007)

Different time–frequency techniques may be used to investigate the relation between latency and frequency of transient-evoked otoacoustic emissions. In this work, the optimization of these techniques and the interpretation of the experimental result are discussed. Time–frequency analysis of click-evoked otoacoustic emissions of 42 normal-hearing young subjects has been performed, using both wavelet and matching pursuit algorithms. Wavelet techniques are very effective to provide fast and reliable evaluation of the average latency of large samples of subjects. A major advantage of the matching pursuit technique, as observed by Jędrzejczak *et al.* [J. Acoust. Soc. Am. **115**, 2148–2158 (2004)], is to provide detailed information about the time evolution of the response of single ears at selected frequencies. A hybrid matching pursuit algorithm that includes Fourier spectral information was developed, capable of speeding-up computation times and of identifying “spurious” atoms, whose latency-frequency relation is apparently anomalous. These atoms could be associated with several known phenomena, either intrinsic, such as intermodulation distortion, spontaneous emissions and multiple internal reflections, or extrinsic, such as instrumental noise, linear ringing and the acquisition window onset. A correct interpretation of these phenomena is important to get accurate estimates of the otoacoustic emission latency.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799924]

PACS number(s): 43.64.Jb, 43.64.Yp [BLM]

Pages: 3576–3585

I. INTRODUCTION

Otoacoustic emissions (OAEs) are acoustic signals generated in the cochlea and detected in the ear canal, either spontaneously (SOAEs) or following acoustical stimulation (Probst *et al.*, 1991). The evoked OAE classification is based on the evoking technique: transient evoked OAEs (TEOAEs) are evoked by a broadband transient stimulus, whereas stimulus frequency OAEs (SFOAEs) are evoked by a pure tone. Distortion product OAEs (DPOAEs) are nonlinearly generated at the frequency $2f_1 - f_2$ (apical DPOAE) by the interaction of two primary tones f_1 and f_2 , with frequencies in a particular ratio (usually, $f_2/f_1 \approx 1.22$).

The latency of TEOAEs is a measurable quantity that provides important information about cochlear mechanisms and, in principle, could also be useful for clinical diagnostic purposes. Indeed, the latency of a given frequency component is sensitive to the position of the corresponding OAE source along the cochlear partition and to the speed of the corresponding spectral component of the traveling wave along its round-trip path. The position of the OAE source is given by the tonotopic Greenwood map (Greenwood, 1990). The speed is a model-dependent function of frequency and cochlear position, which, in transmission-line cochlear models, involves the bandwidth of the cochlear filter associated

with each frequency (Moleti and Sisto, 2003). Increased bandwidth of the auditory filters is generally observed in hearing impaired subjects, associated with damage of the outer hair cells.

TEOAE latency estimates have been obtained in several studies, either using time-domain analysis of the TEOAE wave form to identify the onset time of each frequency component or based on measurements of the TEOAE phase-gradient delay. Time-domain analyses include the analysis of the OAE response to tone bursts (Norton and Neely, 1987; Neely *et al.*, 1988). A narrow-band filtering technique was recently applied to the estimate of SFOAE and DPOAE latencies (Konrad-Martin and Keefe, 2005) in order to improve the accuracy of time-domain estimates of the OAE latency as a function of the stimulus level in normal-hearing and hearing-impaired subjects. Similar information had also been obtained from studies of the DPOAE onset time (Whitehead *et al.*, 1996; Martin *et al.*, 1998), directly measured in the time domain, using an ad hoc acquisition technique to cancel the primary contribution without filtering.

More recently, a different time-domain approach to the TEOAE latency estimate has been used, based on the time-frequency analysis of the TEOAE wave forms. Time-frequency wavelet analysis was used for TEOAE latency estimates (Wit *et al.*, 1994; Tognola *et al.*, 1997). The wavelet transform (WT) was also used to look for effects of hearing impairment on the latency in adult subjects exposed to impulsive noise (Sisto and Moleti, 2002). Similar studies on

^{a)}Electronic mail: arturo.moleti@roma2.infn.it

^{b)}Electronic mail: renata.sisto@ispesl.it

neonates showed longer latencies than in adults (Moleti and Sisto, 2003) and even longer in preterm infants (Tognola *et al.*, 2005), suggesting a dependence of latency on cochlear maturation. A very small difference was reported between the TEOAE latency of neonates discriminated by a standard hospital hearing-screening test (Moleti *et al.*, 2005). A different time-frequency approach, based on an optimized bilinear algorithm (a variant of the Wigner-Ville and Choi-Williams algorithms using a kernel function that reduces the “interference” cross terms associated with bilinear methods) was used in DPOAE and SFOAE studies (Konrad-Martin and Keefe, 2005) aimed at testing the predictions of the OAE generation models. Matching pursuit (MP) algorithms (Mallat, 1989) were also considered for estimating the latency and time span of the resonant modes dominating the TEOAE wave form (Jedrzejczak *et al.*, 2004), and to look for systematic latency differences between normal hearing and hearing impaired ears (Jedrzejczak *et al.*, 2005), between left and right ears (Jedrzejczak *et al.*, 2006) and between preterm and full-term neonates (Jedrzejczak *et al.*, 2007).

Phase-gradient delay techniques also provide estimates of the TEOAE latency in humans (Prieve *et al.*, 1996) and guinea pigs (Withnell and McKinley, 2005). They have more frequently been used to get estimates of the SFOAE latency (Shera and Guinan, 2003; Siegel *et al.*, 2005) and of the DPOAE latency (Kimberley *et al.*, 1993; Schoonhoven *et al.*, 2001) in humans and small mammals. Recent studies showed that DPOAE phase-gradient delay is sensitive to aspirin administration, but in the DPOAE case the observed effect may depend on the different effect of aspirin on the wave-fixed and place-fixed DPOAE components (Parazzini *et al.*, 2005). It must be clarified that the phase-gradient and the time-domain latency estimates are not equivalent. In fact, the phase-gradient delay provides an estimate of the cochlear roundtrip travel time if only place-fixed OAE generation mechanisms are considered, as the linear reflection by random irregularities predicted by the coherent reflection filtering theory (Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera and Guinan, 1999), and only in an approximate sense (Shera *et al.*, 2005). The location of a wave-fixed OAE source changes even with very small changes in frequency because the source is fixed to the traveling wave ensemble. Consequently, small changes in primary frequencies are predicted to result in a slowly rotating wave-fixed OAE phase (and therefore, little or no phase-gradient delay) for a scale-invariant cochlea. Both linear reflection and nonlinear distortion are possible sources of TEOAEs, with different weights at different levels of stimulation, and, at a given stimulus level, with different weights at different frequencies, so the phase-gradient delay contains a rather complicate mixture of information. Time-frequency analysis of the TEOAE wave forms provides instead separate estimates of the latencies of multiple-component OAE responses, independently of the place-fixed or wave-fixed nature of the OAE source. This capability is very important to study particular phenomena, such as the multiple intracochlear reflections (Sisto and Moleti, 2002) and the intermodulation distortion components (Yates and Withnell, 1999), observed in the TEOAE response.

From TEOAE latency estimates, objective (even if model-dependent) estimates of cochlear tuning have also been obtained (Sisto and Moleti, 2002; Moleti and Sisto, 2003). This valuable information has not been fully exploited yet, also due to some technical difficulties that limit the precision and the frequency range of applicability of time-domain methods (e.g., Moleti *et al.*, 2005).

In this study we will discuss the effectiveness of two different time-frequency analysis techniques, based on WT and MP algorithms. Advantages and shortcomings of the different approaches will be discussed, with reference to the particular task of evaluating the time delay of the frequency components of typical TEOAE wave forms from normal-hearing subjects. Intrinsic cochlear phenomena will be taken into account, such as intermodulation distortion, long-lasting and SOAEs and multiple intracochlear reflections, as well as extrinsic effects, such as noise, and the wave form deformation due to the acquisition window. The window onset plays indeed an important role in the latency estimates, particularly at high frequencies, because the highest-frequency latency is obviously the shortest.

A MP-based analysis algorithm especially optimized for this task will be presented, and its effectiveness in the identification of some peculiar TEOAE features, such as multiple intra-cochlear reflections, spontaneous and long-lasting OAEs, nonlinear distortion components, etc., will be investigated.

II. METHODS

A. Subjects and data acquisition

The TEOAE wave forms analyzed in this study were measured in both ears of 42 normal-hearing young male subjects (age 18–25 years). This population was chosen in order to get a sample showing those characteristic features of the healthy cochlea (SOAEs, long-lasting OAEs, internal reflections) that could affect the estimate of the average relation between latency and frequency. Conductive problems in the outer and middle ear were ruled out by otoscopy and tympanometry. The audiometric threshold was measured at the standard frequencies 250, 500, 1000, 2000, 3000, 4000, 6000, and 8000 Hz. According to the standard clinical definition, normal-hearing was defined as audiometric thresholds better than or equal to 20 dB HL at all frequencies.

TEOAEs were recorded using the clinical apparatus ILO-96 (Otodynamics, Ltd.), in the standard “derived nonlinear” mode, using 77–80 dB pSPL click stimuli, and recording two alternate averages, synchronized to the stimulus onset, of 260 elementary acquisition periods of 20 ms. The nonlinear mode of acquisition effectively cancels most of the linear part of the response, including the artifact from linear ringing in the ear canal, but, unfortunately, also any linear components eventually present in the cochlear response. The choice of using a nonlinear acquisition method was necessary, because the ringing artifact, particularly at high stimulus levels, may extend up to latencies of 6–8 ms, typically corresponding to those of the TEOAE frequency components between 2 and 3 kHz. For this reason, linear recordings could be effectively used for latency estimates only up to

1.5–2 kHz, unless much lower stimulus levels are used to decrease the relative weight of the linear component of the response.

A window is introduced in the acquisition phase, to suppress the residual linear ringing artifact that was not exactly cancelled by the nonlinear acquisition procedure. This window has linear ramp-up between 2.5 and 5 ms after the click, and ramp-down from 17.5 and 20 ms. The average of the two alternate average wave forms is defined as the signal wave form, whereas the difference between the two provides a rough estimate of the noise level.

B. Wavelet analysis

A standard wavelet analysis was applied to the TEOAE wave forms, as described in [Moleti et al. \(2005\)](#). The wavelet transform is a linear time-frequency tool (e.g., [Mallat, 1989](#)), in which the wave form is decomposed into a linear superposition of wavelets:

$$WT(\alpha, u) = \int x(t)h_{\alpha, u}^*(t)dt, \quad (1)$$

These wavelets are scaled and time-shifted copies of a “mother” wavelet $h_0(t)$, which is an oscillating function with zero mean:

$$h_{\alpha, u}(t) = \frac{1}{\sqrt{|\alpha|}}h_0\left(\frac{t-u}{\alpha}\right). \quad (2)$$

In this study, we have used the mother wavelet ([Tognola et al., 1997](#)):

$$h_0(t) = \frac{\cos(20t)}{(1+t^4)}, \quad (3)$$

with t in millisecond.

Each wavelet is therefore localized both in the time (at time u) and frequency domain, with the scale parameter α defining both its central frequency f and its time duration Δt . The wavelet frequency is proportional to the scale parameter, while its time duration is inversely proportional to it. Therefore the frequency and the time resolution of a given wavelet are proportional, respectively directly and inversely, to its central frequency. The ineludible compromise between time and frequency resolution is realized by wavelets in an optimal way for many biological signals, including OAEs, whose high-frequency components have typically shorter duration and characteristic variation times (therefore requiring smaller temporal uncertainty) than the low-frequency components, whose discrimination requires instead a smaller uncertainty on the frequency.

For each one of the n_k ears, the continuous wavelet transform was integrated over ten 500 Hz bandwidths, centered at frequencies f_i from 0.5 to 5 kHz. For the k th ear, the band-averaged wavelet coefficient $W_k(f_i, t_j)$ represents the time evolution of the signal amplitude in that frequency band. The individual latency of the k th ear in the i th frequency band, $\tau_k(f_i)$, is defined as the time when the absolute value of the individual wavelet coefficient reaches its maximum. Due to the finite extension of the acquisition window (which cuts off the lowest- and highest-frequency TEOAE

response) and to the variations of the stimulus spectral density, only the data between 1 and 4 kHz are fully reliable.

By averaging over all the ears one gets an ear-averaged band-averaged wavelet coefficient:

$$W(f_i, t_j) = \frac{1}{n_k} \sum_k W_k(f_i, t_j) \quad (4)$$

The ear-averaged latency $\tau(f_i)$ may be defined as the time when the ear-averaged wavelet coefficient reaches its maximum amplitude. An alternative definition of ear-averaged latency $\tau^*(f_i)$ is the ear-average of the individual latencies $\tau_k(f_i)$:

$$\tau^*(f_i) = \frac{1}{n_k} \sum_{k=1}^{n_k} \tau_k(f_i), \quad (5)$$

The first definition has been demonstrated to be less sensitive to systematic errors ([Moleti et al., 2005](#)) in the evaluation of ear-averaged TEOAE latencies. In fact, noise and SOAEs yield wavelet coefficients uniformly spread over the whole acquisition interval, which produce a number of unreasonably long individual latency estimates at high frequency. The average in Eq. (5) may be significantly affected by a few long-latency estimates, while the time of the maximum of the ear-averaged wavelet coefficient in Eq. (4), is much more stable.

C. Matching pursuit analysis

A standard MP algorithm was used ([Mallat and Zhang, 1993](#)), which decomposes the wave form into a sum of basic wave forms called “atoms,” using an iterative procedure. The atoms are chosen from a redundant library (dictionary) of functions of the type:

$$g_{u,s,\xi}(t) = g\left(\frac{t-u}{s}\right)e^{-j\xi t} \quad (6)$$

where $g(t)$ is a Gaussian window function, centered at time u , whose time duration depends linearly on the scale parameter s . The central angular frequency of the atom is fixed by the modulation parameter ξ , independent of the time duration of the atom. As a consequence, the bandwidth of each atom is not bound to be inversely proportional to its central frequency, as in the wavelet case. Single atoms may therefore effectively represent the contribution of resonances with different quality factors, including long-lasting high-frequency OAEs, as first demonstrated by [Jedrzejczak et al. \(2004\)](#). The atoms form a highly redundant set of basis functions, which are not orthogonal to each other.

The decomposition procedure starts by finding in the library the atom g_0 that maximizes the inner product with the wave form $x(t)$:

$$G_{u,s,\xi} = \int x(t)g_{u,s,\xi}^*(t)dt \quad (7)$$

The atom g_0 maximizes the “projection” of $x(t)$, which is subtracted from the original wave form to get the first residue:

$$x_1(t) = x(t) - G_{u,s,\xi}x(t) \quad (8)$$

The same procedure is then applied to the first residue, finding a second atom $g_1(t)$, which maximizes the inner product with the first residue. A second residue $x_2(t)$ is obtained, and the procedure is iterated until either a predetermined number of atoms has been found, or the residue amplitude has become smaller than a predetermined error threshold.

For typical TEOAE wave forms, a rather small number of atoms (10–15) is necessary to reconstruct a large part (90%) of the original signal (Jedrzejczak *et al.*, 2006). In our implementation of the standard hierarchical MP algorithm, we have chosen to stop the iteration when either 90% of the energy is reconstructed or the number of atoms exceeds 18. Due to the highly redundant number of atoms in the library, the MP algorithm may be rather time consuming, if a good time and frequency resolution has to be achieved.

D. MP algorithm modification

In the MP studies by Jedrzejczak *et al.* (2004, 2005, 2006), it is evident that most atoms tend to cluster in the time-frequency plane along a line corresponding to a negative power law relation between TEOAE latency and frequency. This result is in good quantitative agreement with wavelet estimates of the latency-frequency relation (Tognola *et al.*, 1997; Sisto and Moleti, 2002; Moleti *et al.*, 2005). Nevertheless, a large spread of the data is generally observed, with several long-latency high-frequency atoms and short-latency low-frequency atoms occur. The latency of these atoms is “anomalous” because it cannot be related in a simple way to the cochlear round trip travel time expected based on the component’s spectral frequency, whereas the general trend suggests that the transmission-line roundtrip delay interpretation is probably correct. When averaging over a sample of subjects, these atoms significantly affect the average latency estimates, particularly in the high-frequency range. It has been observed that similar features arise also by using the wavelet transform technique (Moleti *et al.*, 2005), and that they are more common when the signal to noise ratio (SNR) is low. This is a major problem when comparing the average latency of normal-hearing and hearing-impaired populations, because the latter have systematically lower SNR. This bias could cause the observation of statistically significant differences between the latencies of normal-hearing and impaired populations, not necessarily associated with a reliable diagnostic power in individual hearing-impaired ears.

In addition to noise, other possible sources of atoms with anomalous latency-frequency relation are intermodulation distortion, producing low-frequency short-latency atoms, long-lasting or spontaneous OAEs, producing sometimes subsequent atoms at the same frequency to represent a single sustained long-lasting damped oscillation, or intracochlear reflections between the base and the tonotopic place, producing atoms at the same frequency with multiple latencies. These atoms are not associated with noise, and represent real OAE signals, but they are not representative of the general relation between frequency and latency related to the single roundtrip delay of a single wave packet, and should be

studied separately. For the above reasons, we have developed a hybrid MP-based analysis algorithm, especially dedicated to the analysis of TEOAE wave forms, with the aim of identifying the atoms associated with special cochlear phenomena, selecting them for a separate analysis.

A preliminary Fourier analysis was applied to each wave form. The spectral frequency range was divided into four frequency bands: $0.4 < f < 1$ kHz, $1 \text{ kHz} < f < 1.7$ kHz, $1.7 \text{ kHz} < f < 2.8$ kHz, and $2.8 \text{ kHz} < f < 5$ kHz. The energy content in each band was evaluated. The iterative MP algorithm was modified to search for atoms centered within each band, selecting each time the band with the highest residual energy content. The search in each band was stopped when the addition of another atom would increase the energy of the band by less than 10% of its initial total energy. This condition was meant to accept atoms that increase significantly the energy content of the band, to avoid including atoms that could be probably associated with noise. A possible alternative choice would be to stop the search when the total energy of the band reached some threshold, say 80–90% of the total energy, but this choice would be ineffective for very low SNR bands. In fact, in the limit case of SNR=0, wideband noise would yield a very large number of small atoms at all frequencies and latencies.

The algorithm was designed to recognize as a single prolonged response two or more atoms at the same frequency (within half their bandwidth) if their latencies differed less than their temporal uncertainty. The algorithm detected possible intracochlear reflections by finding atoms with the same frequency at multiple delays. The algorithm identified candidates for intermodulation distortion products as those atoms of frequency f_{DP} with latency shorter than those of a couple of candidate primary atoms of frequencies f_1 and f_2 , with $f_{DP} = 2f_1 - f_2$. Shorter latency of the atoms identified as intermodulation distortion products was assumed based on the predictions of transmission-line cochlear models. If the f_{DP} tone is generated at the f_2 cochlear place, its path is equal to that of the f_2 tone, but its backward travel time is shorter. In fact, the backward travel time of the f_{DP} tone does not include the significant contribution associated with the path near its resonance place because the f_2 place is more basal than the f_{DP} resonant place.

At very low frequencies ($f < 0.5$ kHz), the cochlear round-trip delay τ may be even longer than the acquisition time ($T = 20$ ms). Therefore, low-frequency responses of latency $\tau > 20$ ms could be recorded at a latency $\tau - T$ by the synchronous averaging mechanism. For this reason, we have decided to exclude from the analysis atoms with central frequency lower than 0.4 kHz. This problem could arise also, and even at intermediate frequencies (0.5–1 kHz), for the above mentioned intracochlear reflections. These last observations suggest that using a longer acquisition time in this kind of studies could help preventing ambiguity in the interpretation of the results.

III. RESULTS AND DISCUSSION

TEOAE latencies are made up of a small noncochlear contribution, associated with delays in the middle ear and in

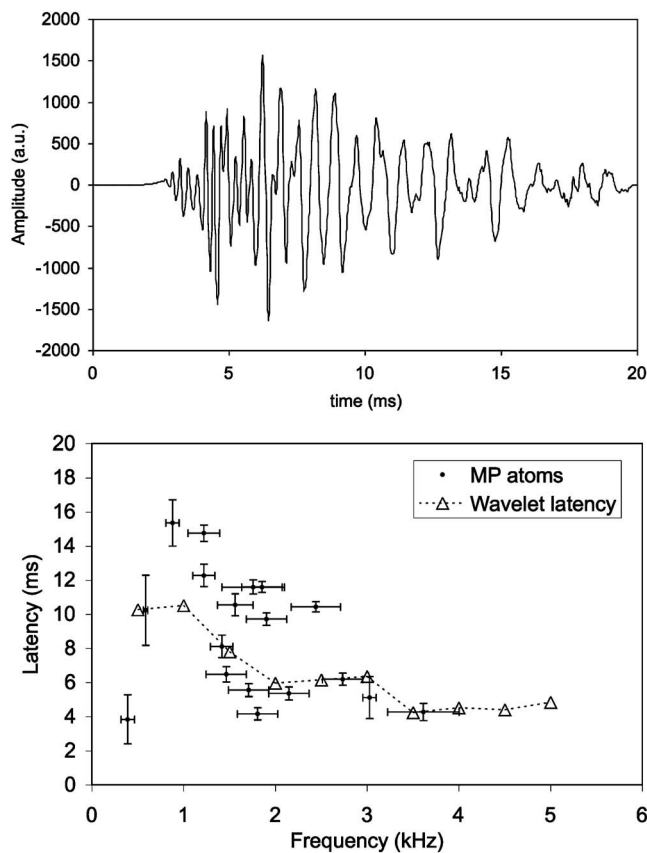


FIG. 1. Typical TEOAE wave form (top panel, ear #7), showing the characteristic dispersion of the frequency response along the time axis, associated with the transmission time along the BM. Time-frequency analysis of the same TEOAE wave form (bottom), using wavelet band-averaged coefficients (white triangles), and MP atoms (black dots). Error bars show the time and frequency width of each MP atom.

the tubing of the insert earphone, and a cochlear roundtrip transmission delay. The noncochlear offset is not subtracted from the latency estimates because in this work we are not interested in advanced model-dependent interpretations of the latency-frequency relation. The offset should be subtracted if the experimental TEOAE latency had to be interpreted as a cochlear roundtrip delay, to allow comparisons with other OAE- or ABR-based estimates of this delay, or to get model-dependent estimates of the cochlear tuning curve (Moleti and Sisto, 2003; Moleti *et al.*, 2005). A typical TEOAE wave form of a healthy ear is shown in the top panel of Fig. 1 for the ear #7 of our sample. The effect of the acquisition window ramp-up is evident between 2.5 and 5 ms, as well as the characteristic TEOAE shape, with the highest frequency components recognizable as fast oscillations in the early phase of the response, and the lower frequencies reaching the detector later, due to their longer roundtrip cochlear path. This monotonic temporal dispersion of the frequency components of the response is generally observed in human TEOAEs, and it is the main reason for using time-frequency techniques in the analysis of these signals. As explained in Sec. II, wavelet analysis projects the wave form on a basis of functions localized in time and frequency that are constant- Q bandpass filters. Instead, the MP technique projects the wave form on a set of atoms,

whose Q is a free parameter. This feature permits to represent the resonant response at a given frequency with a single atom, or with a few atoms of the same frequency.

The individual wavelet latencies of the same ear #7, $\tau_j(f_i)$, are shown in the bottom panel of Fig. 1 (white triangles), along with the atoms obtained from the MP decomposition of the same wave form (black dots). This case shows a general trend in the data: the band-averaged wavelet latency has a rather regular behavior, but much more information is provided by the MP analysis. In particular, MP low-frequency atoms at very long latencies are easily found, while the wavelet technique often fails to identify these late components of the response, below 1 kHz. On the other hand, due to the hierarchical nature of the MP algorithm, few atoms are found in adult subjects above 4 kHz, where the TEOAE energy density is typically low. Wavelet analysis provides a latency estimate for each frequency band, but these estimates become meaningless for the bands with very low SNR. Therefore, a data selection based on band SNR may be necessary, but one should also consider that this kind of data selection might bias the results.

In the top panel of Fig. 2 we show the distribution of all the individual $\tau_j(f_i)$, for all frequency bands and for all the ears. A clear dependence of latency on frequency is visible, even if the vertical dispersion of the data is rather large. A power law fit to all the individual latencies yields: $\tau = 8.6f^{-0.40}$. The normalized, ear-averaged, band-averaged wavelet coefficients for the bands centered at 1, 2, 3, and 4 kHz, are plotted in the middle panel of Fig. 2. Vertical lines indicate the times associated with the amplitude maxima, defined as the ear-averaged latencies $\tau(f_i)$. The few ears with unreasonably long individual high-frequency latency contribute to the characteristic tail of the high-frequency average wavelet coefficients, but they do not significantly affect the position of the peak of the average wavelet coefficient, as long as they represent a small fraction of the sample. In the bottom panel of Fig. 2, the average latencies $\tau(f_i)$ are fitted to a power law, obtaining $\tau = 8.6f^{-0.43}$, very close to what had been obtained by fitting all the individual latencies (top panel). In this case error bars correspond to the width of the average wavelet coefficient at half-maximum.

In Fig. 3, a time-frequency representation is given of all the atoms obtained for all the ears by the standard hierarchical MP algorithm (MP0) and by our modified MP algorithm (MP1). These plots can be compared to the individual wavelet latency estimates shown in the top panel of Fig. 2. As with wavelets, the time-frequency distribution of the atoms clusters along a negative power law relation. The MP latency estimates show a larger dispersion than the wavelet latency estimates. The power law fit yields, for MP0: $\tau = 10.3f^{-0.44}$, and for MP1: $\tau = 10.0f^{-0.46}$. The new MP algorithm performs similarly to the other one, yielding almost the same total number of atoms (an average of 15 atoms per ear for MP0, and 14 for MP1), with a reduction by a factor ten of the computation time. Another small difference is shown in Fig. 4, where the bars represent the fraction of the energy content of each frequency band reconstructed by the two MP algorithms, showing that our modified algorithm (MP1) slightly favors the side bands, in which few atoms are typically found

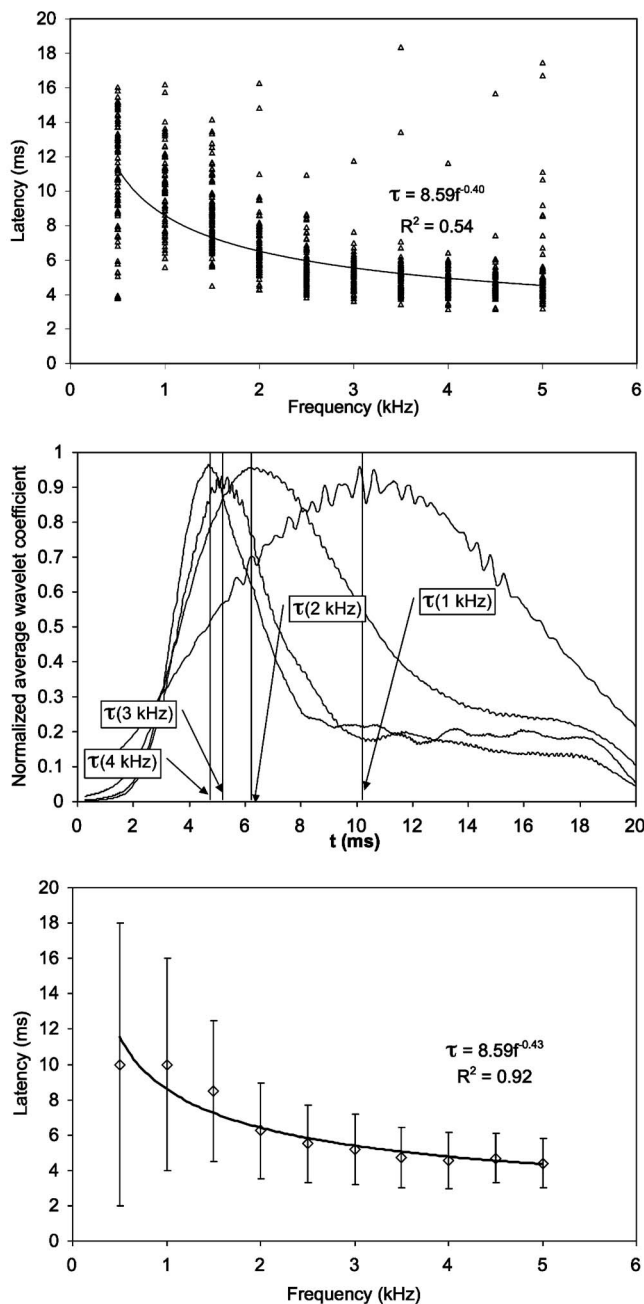


FIG. 2. (top panel) Individual wavelet band coefficients of all 84 ears, fitted to a negative power law; (middle) ear-averaged wavelet coefficient at 1, 2, 3, and 4 kHz. Vertical lines indicate the time associated with the maxima, defined as the ear-average latency τ (bottom) fit of the average latency τ to a power law of frequency.

by the standard hierarchical MP algorithm (MP0). This feature may permit, for a given total number of atoms, to get a slightly better statistical description of the latency/frequency relation over a wider frequency interval by populating the side bands with some additional atoms.

The latency-frequency relations shown in Figs. 2 and 3 are all consistent with previously published results from normal-hearing adults (Sisto and Moleti, 2002; Jedrzejczak et al., 2004). It is interesting to note that the MP estimates give a slightly higher latency and steeper slope of the latency–frequency relation. In Sisto and Moleti (2002), a line identification method applied to the wavelet algorithm also

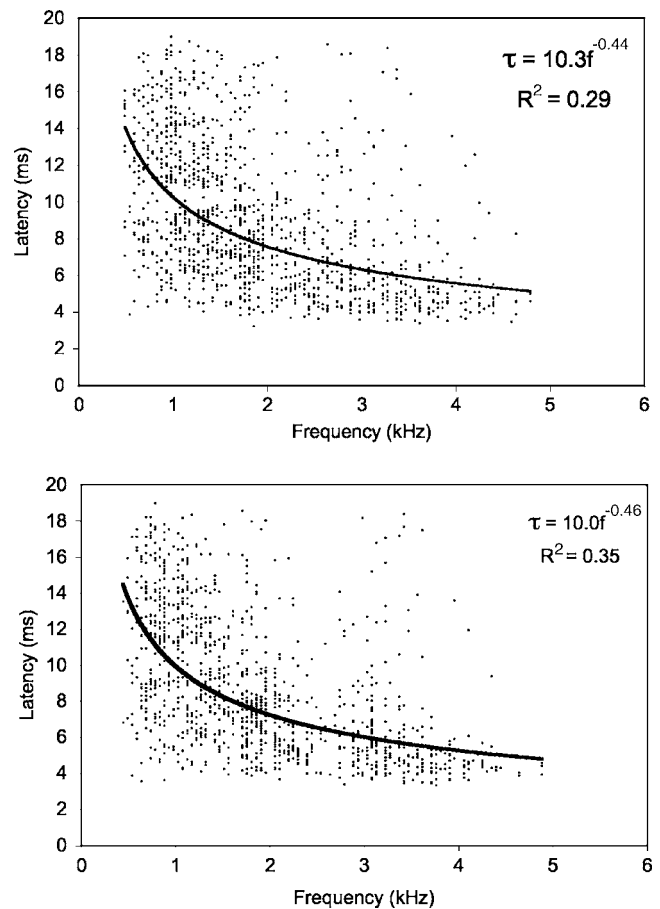


FIG. 3. Atoms obtained from the MP decomposition of all 84 ears, using the standard algorithm MP0 (top) and the modified hybrid algorithm MP1 (bottom). In both cases, the dispersion of the data is quite large.

yielded significantly increased steepness of the latency-frequency relation. The line-identification method is conceptually very similar to the MP approach, because in both cases the delays of the main spectral components of the TEOAE response are measured, associating each delay with the corresponding frequency. Instead, the latency estimated as the time of the maximum of the band-averaged wavelet coefficient is attributed to the central frequency of each band, and

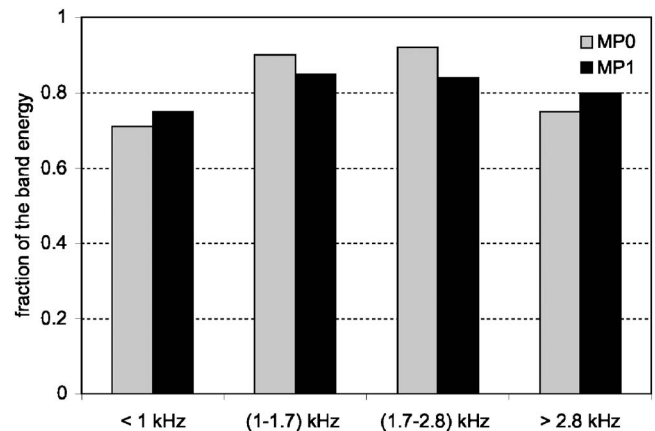


FIG. 4. Fraction of the total energy reconstructed for each frequency band by the two MP algorithms. The new algorithm (MP1) tends to reconstruct slightly better the energy content of the side bands.

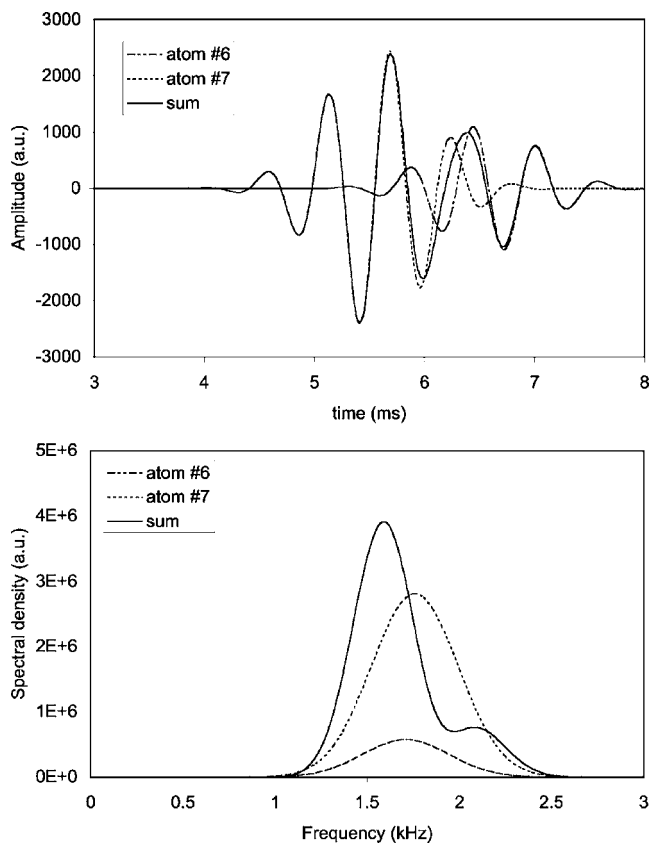


FIG. 5. Two atoms of the same frequency and overlapping time spans are summed by the algorithm to form a single OAE response. In the top panel the wave forms of the two atoms and the resulting composite wave form are shown, whereas the corresponding spectra are shown in the bottom panel.

this fact may cause an underestimate of the slope of the latency-frequency relation. In fact, TEOAE spectra are typically peaked in the midfrequency range (1–2 kHz), and, therefore, the main contribution to low-frequency bands typically comes from their high-frequency borders, whereas the opposite happens for the high-frequency bands. This “dilation” of the frequency scale may cause an underestimate of slope of the latency-frequency relation.

Both the MP and wavelet techniques are effective tools to get information about the average relation between latency and frequency. This information is useful to test cochlear models and to get estimates of physiological quantities, such as tuning, over a large population, but the large spread of the individual wavelet latencies and of the MP atoms at each frequency makes it difficult to use the latency estimates for diagnostic tests on single subjects.

As shown in Fig. 3, MP analysis yields a large number of atoms whose latency appears inconsistent with cochlear traveling wave delay. On the other hand, the accurate time-frequency characterization of the MP atoms permits to explain their origin as due to phenomena predicted by cochlear models. For this purpose, we performed a second level of analysis on the atoms obtained with the modified MPI algorithm, in which, for each ear, the atoms characterized by particular time-frequency relations, which could be signatures of known cochlear phenomena, were identified and grouped together. In Figs. 5–8 we show examples of identi-

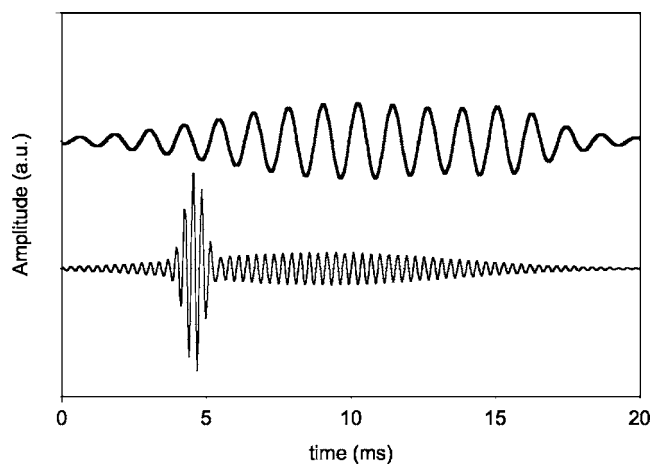


FIG. 6. Long-lasting OAEs, both formed by two atoms. The first case corresponds to a stationary SOAE, the second could be a SOAE excited to a level much higher than its equilibrium level.

fication of these MP atoms with latencies that are not easily explained by a single roundtrip to the emission generating region of the basilar membrane. It would be more difficult and arbitrary to extract the same information using the wavelet technique, because the MP algorithm directly provides a preliminary identification and characterization of a small number of resonant responses at selected frequencies in terms of a few well-defined parameters (amplitude, frequency, latency, and time span of each atom).

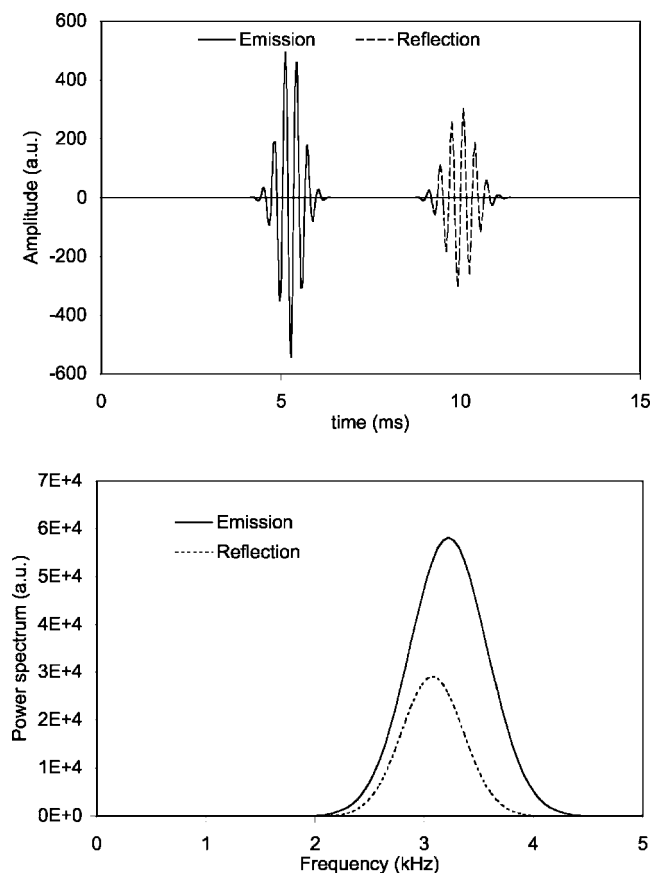


FIG. 7. Two atoms at the same frequency and multiple latency can be interpreted as due to intracochlear reflection.

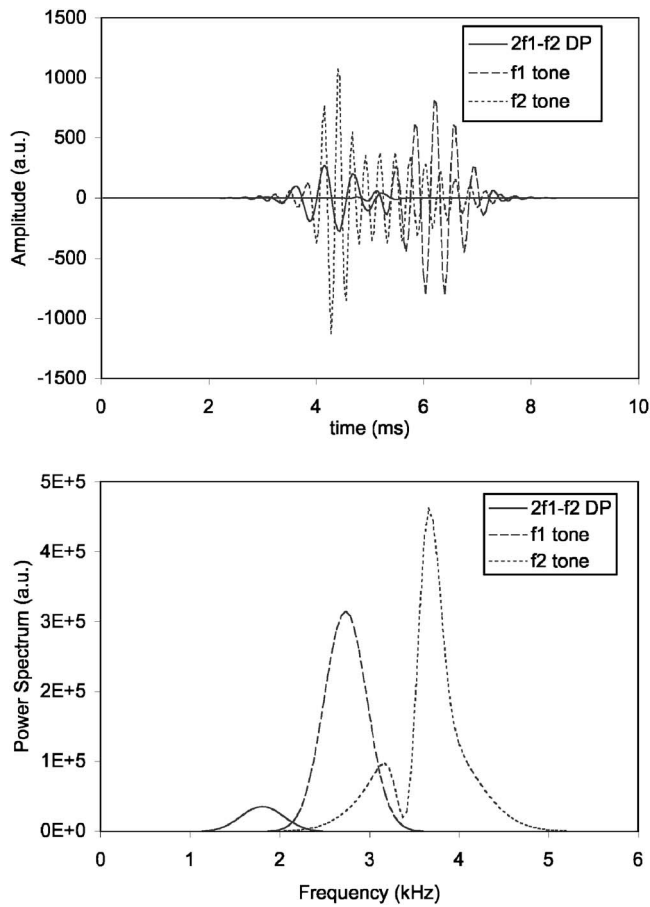


FIG. 8. Three atoms matching the time and frequency criteria to be candidate for intermodulation distortion generation. The $f_{DP}=2f_1-f_2$ atom has a latency shorter than that of the f_2 atom.

In Fig. 5 two atoms of the same ear, with the same frequency and latency (within the limits of measurement accuracy), are shown. In this case a single prolonged OAE response could be associated to these atoms, which are probably separated by the MP algorithm by chance, or to fit the asymmetric non-Gaussian shape of the envelope.

Some OAEs (a total of 23, either corresponding to single atoms or to groups) have been identified as long-lasting OAEs (sometimes also recognizable as SOAEs), if their total time span exceeded an empirical frequency dependent threshold ($\Delta\tau > A/f^{0.5}$, with $A=3$ ms and f expressed in kilohertz). The frequency dependence of the threshold was chosen according to the known frequency dependence of the quality factor of the typical cochlear resonances (Moleti and Sisto, 2003). In Fig. 6 we show two such long-lasting OAEs, each consisting of two atoms. In the top trace, the junction between the two atoms at 13 ms is hardly visible (this is a constant SOAE), whereas in the bottom trace a short strong excitation is visible, followed by a constant amplitude OAE (this behavior is consistent with the synchronous excitation of a SOAE above its equilibrium value).

In Fig. 7 we show an example of atoms with the same frequency and multiple latencies, which can be interpreted as due to multiple intracochlear reflections of the same OAE. These atoms (15 such cases have been found) match the criteria $f_2=f_1$, and $\tau_2=2\tau_1$. Sometimes a different pattern is

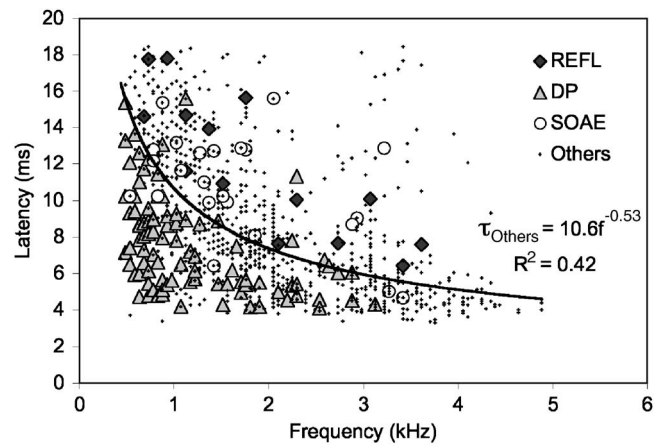


FIG. 9. Time-frequency distribution of all the MP1 atoms. The “special” atoms are identified by different symbols: gray triangles for candidate intermodulation distortion atoms (DP), black diamonds for candidate multiple reflection atoms (REFL), white circles for long-lasting OAEs (SOAE). Each class tends to cluster in a separate region. The fit to the remaining atoms (Others), shown as small dots, is slightly improved with respect to that of Fig. 3.

observed, which could be associated with the same phenomenon: the reflected OAE latency may be shorter than that of the first one, if its true latency τ_2^* is longer than the acquisition time interval (the click repetition time interval $T=20$ ms, in this study). In this case, it would be recorded by the synchronous averaging system at a latency $\tau_2=\tau_2^*-T$. These atoms would match the criteria $f_2=f_1$, and $\tau_2=2\tau_1-T$. In this study, we have not considered this possibility. This issue could be clarified by repeating TEOAE measurements using different click repetition time intervals.

In Fig. 8 we show atoms that are candidates for intermodulation distortion products of other higher-frequency atoms (110 such cases have been found). These atoms match the criteria $f=2f_1-f_2$ and $\tau < \tau_2 < \tau_1$. Yates and Withnell (1999), identified indeed intermodulation distortion components in guinea pig TEOAEs using high-pass filtered clicks. In this study, their signature is associated with the relative position of the three atoms in the joint time-frequency domain. As explained in Sec. II, the ordering of the three latencies $\tau < \tau_2 < \tau_1$, is requested to match the prediction of transmission-line cochlear models, assuming that the $2f_1-f_2$ tone comes from the cochlear place that is resonant at frequency f_2 . The backward path of the distortion product is expected to be faster than that of the primary tone f_2 because it extends over a cochlear region that does not include its own resonance place. We remind here that in transmission-line models the group velocity of each wave packet is minimum near its resonance place.

If we use different symbols to identify all these possible cochlear phenomena, we get the plot shown in Fig. 9, which shows that the DP candidates tend to cluster in the lower left corner of the plot (short latency and low frequency), whereas the multiple reflections candidates and the long-lasting OAEs are found at all frequencies, and contribute to fill the long-latency part of the plot. The fit to the remaining atoms (others) yields: $\tau=10.6f^{-0.53}$ and $R^2=0.42$. Of course the identified atoms are just candidates, and this procedure has a

statistical meaning only. The improved fit obtained after having selected off the special atoms, suggests that a good fraction of them could have been correctly identified.

As already pointed out, anomalous wavelet band latency and MP atoms with anomalous latency/frequency relation are correlated with low TEOAE SNR in the correspondent frequency band. All the special cochlear phenomena considered above are not correlated with low band SNR. Actually, SOAEs, and, to some extent, multiple reflections and nonlinear distortion generation, would rather be correlated with high band SNR. For this reason, these anomalous latency data shown in Fig. 9 could not have been eliminated by a data selection rule based on SNR.

At this stage of the analysis, our sample of MP1 atoms has been reduced from 1171 to 957 (the point-like symbols in Fig. 9). This reduction has been obtained by grouping together atoms of the same frequency and overlapping time span, identifying and removing candidates for long-lasting OAEs, for intermodulation distortion and for multiple reflection. The final number of selected atoms is comparable to the total number of wavelet band coefficients (840) shown in the top panel of Fig. 2.

A SNR-based selection can be applied at this stage, to further improve the fit, after having identified the above mentioned special atoms. To show the effect of noise on both wavelet and MP algorithms, we further selected the wavelet coefficients, reducing them to $N_w=691$ by excluding those coming from bands with TEOAE SNR < 1. Correspondingly, the residual 957 MP atoms were sorted by amplitude, selecting only the N_w largest atoms. The result is shown in Fig. 10, where a rather good fit is obtained with both techniques. Differently from the wavelet case, the fit improvement obtained using an amplitude criterion is not very relevant for MP. The preliminary identification of reflections, distortion products and long lasting emissions was necessary to get this result, because most of the atoms of these categories would have survived a selection based on the amplitude only.

Even when employing the MP algorithm to remove apparently anomalous latency data, there is still considerable spread in the data, as shown in Fig. 10. This variability has to be further reduced, to get latency estimates that could be useful for diagnostic purposes in individual subjects. Further progress in this direction is necessary, which needs both improvement of the experimental setup and full exploitation of the information coming from theoretical cochlear models.

IV. CONCLUSIONS

Both wavelet and matching pursuit algorithms do provide valuable information about the relation between the OAE latency and frequency. The results of this study confirm that these time-frequency techniques are particularly useful to extract information from the TEOAE wave forms. In fact, wavelet and MP techniques permit to accurately study such transient responses. MP algorithms have demonstrated the capability of describing the resonant response at a single frequency with a single or a few atoms. The examination of complex temporal responses associated with individual frequency components of TEOAEs permits careful studies of

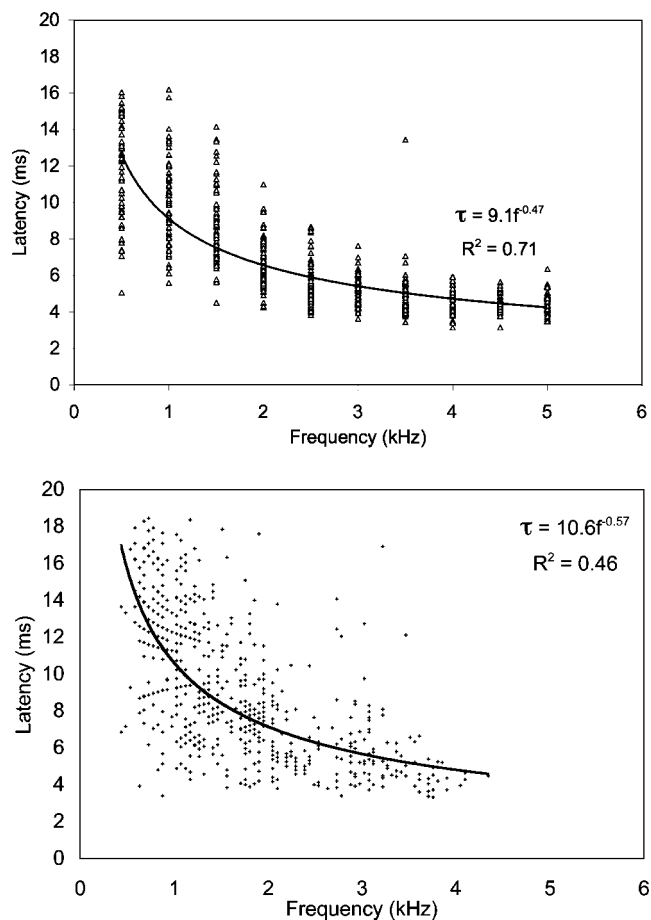


FIG. 10. Selecting only the wavelet latency of bands with SNR > 1 (691 out of 840), most of the unreasonable latency values visible in the top panel of Fig. 2 disappear, and a better fit to a power law is obtained (top). For comparison, we have selected the 691 largest amplitude MP1 atoms from the 957 “other” atoms shown as dots in Fig. 9. A slightly better fit to a power law is obtained (bottom), as shown by the correlation coefficient.

peculiar cochlear phenomena, such as high- Q long-lasting responses, multiple intracochlear reflections, intermodulation distortion components, exploiting the theoretical knowledge of their time-frequency signatures, and opens the possibility of studying quantitatively the variation of these phenomena under different experimental conditions (e.g., different stimulus levels, contralateral stimulation, exposure to ototoxic agents, etc.). On the other hand, it has also been remarked that the representation of the TEOAE wave form as a sum of atoms leads to the inclusion of spurious atoms, which may be associated either with noise or with real cochlear phenomena producing responses with anomalous latency–frequency relation. If a good accuracy is required in the wave form decomposition, the presence of these atoms is necessary, but if, as in this case, one is interested in the average latency–frequency relation, the accuracy of the reconstruction of the original wave form may be partially neglected, to privilege the selection of those atoms whose latency is likely due to roundtrip transmission delay. A step in this direction has been made in this work, showing, on a test population of normal-hearing young subjects, that it is possible: (1) to get faster convergence, by using a hybrid algorithm including Fourier information and (2) to identify atoms or groups of atoms

associated with known cochlear phenomena, whose latency/frequency relation is apparently anomalous, using model-dependent criteria.

- 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., and Konopka, P. J. (2006). "Resonant modes in transiently evoked otoacoustic emissions and asymmetries between left and right ear," *J. Acoust. Soc. Am.* **119**, 2226–2231.
- Jedrzejczak, W. W., Blinowska, K. J., and Konopka, W. (2005). "Time-frequency analysis of transiently evoked otoacoustic emissions of subjects exposed to noise," *Hear. Res.* **205**, 249–255.
- 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.
- 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.
- Kimberley, B. P., Brown, D. K., and Eggermont, J. J. (1993). "Measuring human cochlear traveling wave delay using distortion product emission phase responses," *J. Acoust. Soc. Am.* **94**, 1343–1350.
- Konrad-Martin, D., and Keefe, D. H. (2005). "Transient-evoked stimulus-frequency and distortion-product otoacoustic emissions in normal and impaired ears," *J. Acoust. Soc. Am.* **117**, 3799–3815.
- Mallat, S. G. (1989). "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 674–693.
- Mallat, S. G., and Zhang, Z. (1993). "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Process.* **41**, 3397–3415.
- Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (1998). "Locus of generation for the $2f_1-f_2$ vs $2f_2-f_1$ distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations," *J. Acoust. Soc. Am.* **103**, 1957–1971.
- 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.
- Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (1988). "Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli," *J. Acoust. Soc. Am.* **83**, 652–656.
- Norton, S. J., and Neely, S. T. (1987). "Tone-burst-evoked otoacoustic emissions from normal-hearing subjects," *J. Acoust. Soc. Am.* **81**, 1860–1872.
- Parazzini, M., Hall, A. J., Lutman, M. E., and Kapadia, S. (2005). "Effect of aspirin on phase gradient of $2F_1-F_2$ distortion product otoacoustic emissions," *Hear. Res.* **205**, 44–52.
- Prieve, B. A., Gorga, M. P., and Neely, S. T. (1996). "Click- and tone-burst-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears," *J. Acoust. Soc. Am.* **99**, 3077–3086.
- Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (1991). "A review of otoacoustic emissions," *J. Acoust. Soc. Am.* **89**, 2027–2067.
- 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., 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.
- Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (2005). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," *J. Acoust. Soc. Am.* **118**, 2434–2443.
- Sisto, R., and Moleti, A. (2002). "On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears," *J. Acoust. Soc. Am.* **111**, 297–308.
- 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., Parazzini, M., de Jager, P., Briennes, P., Ravazzani, P., and Grandori, F. (2005). "Cochlear maturation and otoacoustic emissions in preterm infants: A time-frequency approach," *Hear. Res.* **199**, 71–80.
- Tognola, G., Ravazzani, P., and Grandori, F. (1997). "Time-frequency distributions of click-evoked otoacoustic emissions," *Hear. Res.* **106**, 112–122.
- Whitehead, M. L., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (1996). "Visualization of the onset of distortion-product otoacoustic emissions, and measurement of their latency," *J. Acoust. Soc. Am.* **100**, 1663–1679.
- 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.
- Withnell, R. H., and McKinley, S. (2005). "Delay dependence for the origin of nonlinear derived transient-evoked otoacoustic emissions," *J. Acoust. Soc. Am.* **117**, 281–291.
- Yates, G. K., and Withnell, R. H. (1999). "The role of intermodulation distortion in transient-evoked otoacoustic emissions," *Hear. Res.* **136**, 49–64.
- Zweig, G., and Shera, C. A. (1995). "The origin of periodicity in the spectrum of otoacoustic emissions," *J. Acoust. Soc. Am.* **98**, 2018–2047.

The effect of hair on auditory localization cues

Bradley E. Treeby,^{a)} Jie Pan, and Roshun M. Paurobally

*Centre for Acoustics, Dynamics and Vibration, School of Mechanical Engineering,
The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia*

(Received 14 May 2007; revised 15 August 2007; accepted 31 August 2007)

Previous empirical and analytical investigations into human sound localization have illustrated that the head-related transfer function (HRTF) and interaural cues are affected by the acoustic material properties of the head. This study utilizes a recent analytical treatment of the sphere scattering problem (which accounts for a hemispherically divided surface boundary) to investigate the contribution of hair to the auditory cues below 5 kHz. The hair is modeled using a locally reactive equivalent impedance parameter, and cue changes are discussed for several cases of measured hair impedance. The hair is shown to produce asymmetric perturbations to the HRTF and the interaural time and level differences. The changes in the azimuth plane are explicated via analytical examination of the surface pressure variations with source angle. Experimental HRTFs obtained using a sphere with and without a hemispherical covering of synthetic hair show a good agreement with analytical results. Additional experimental and analytical investigations illustrate that the relative contribution of the hair remains robust, regardless of the placement of the pinnae, or inclusion of a cylindrical neck. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2793607]

PACS number(s): 43.66.Ba, 43.66.Pn, 43.20.Fn [KA]

Pages: 3586–3597

I. INTRODUCTION

In a natural listening environment, external sounds are predominantly coupled to the human auditory system via the pinnae and ear canals. For a given source location, the combined diffraction and reflection properties of the external human topography (i.e., the head, torso, and pinnae) create distinct timing and magnitude characteristics in the complex wave forms present at the eardrums. The rudimentary duality connecting the psychophysical interpretation of the salient features of these wave forms with the physical processes that govern their creation is well understood. The head introduces interaural time and level differences for sources offset from the median plane, the torso introduces low frequency reflections for elevated sources, and the folds and cavities of the pinnae create idiosyncratic spectral filtering effects above 3 kHz dependent on the source location (e.g., [Blauert, 1997](#); [Algazi et al., 2001a](#); [Shaw, 1997](#)). Beyond this, however, the contribution of peripheral and detailed features to the auditory percept is not well understood.

In this context, the current study serves to systematically model, describe, and explicate the effect of human terminal scalp hair on the head-related transfer function (HRTF) and related auditory localization cues. The hair is modeled as a hemispherical covering on a spherical head utilizing a recent analytical solution to the corresponding scattering problem by [Treeby et al. \(2007a\)](#). A locally reactive equivalent impedance parameter is used to encapsulate the acoustic properties of human hair as discussed by [Treeby et al. \(2007b\)](#). Cue changes are discussed for several cases of measured hair impedance. The predicted results are experimentally validated using a rigid sphere with a hemispherical covering of

synthetic hair material. The contribution of hair in relation to other peripheral features, such as the neck and pinna offset, is also investigated.

The use of spherical head models to explain features within the human HRTF is commonplace (e.g., [Kuhn, 1977](#); [Duda and Martens, 1998](#); [Brungart and Rabinowitz, 1999](#); [Algazi et al., 2002a](#)). Within certain constraints (e.g., the symmetrical and thus ambiguous directional cues), these models provide a strong theoretical foundation for understanding features within human HRTF. Such models are also utilized to reintroduce the time delays for minimum-phase HRTF reconstructions ([Kulkarni et al., 1999](#)), as the basis of structural or cascaded HRTF formations ([Brown and Duda, 1998](#); [Chan and Chen, 2000](#)), and to augment experimental HRTFs with low frequency information ([Algazi et al., 2002b](#); [Zotkin et al., 2003](#); [2006](#)).

The traditional spherical head model assumes that the scattering surfaces are completely rigid in nature. This is consistent with the general agreement that skin is sufficiently rigid throughout the audible frequency range ([Katz, 2000](#)). However, results from an earlier analytical study by [Treeby et al. \(2007d\)](#) (using a spherical head model with uniformly distributed acoustic properties) suggest that impedance values representative of human hair can introduce noticeable modifications to the interaural azimuth cues. Empirical studies using mannequins ([Riederer, 2005](#)), spheres ([Treeby et al., 2007c](#)), and the boundary element method ([Katz, 2001](#)) have also shown that the addition of hair introduces asymmetrical perturbations to the HRTF in the order of several decibels. The assumption of a uniformly distributed (and rigid) surface boundary in spherical head models (to allow for an analytic solution) precludes their use for investigation or inclusion of these features. The recent analytical development discussed by [Treeby et al. \(2007a\)](#), however, provides a scattering model for a sphere with a hemispherically split

^{a)}Electronic mail: treebs@mech.uwa.edu.au

surface boundary condition. This impedance distribution assimilates the general characteristics of a human head with hair, and thus facilitates an analytical investigation into the effect of hair on the auditory localization cues.

Considering the complexity of everyday auditory environments, there are a multitude of environmental and intrapersonal variations that may cause subtle changes to the encoded auditory cues for a particular source direction. Variations in clothing, headwear, and complex interactions with the architectural surroundings all produce changes to the available spectral features. Nevertheless, this does not necessarily mean that the ability to localize sounds is significantly perturbed. In the absence of methodical and robust psychoacoustic testing, the effect of these variations is best discussed in relation to the magnitude of their effect on the monaural and interaural cues, in combination with the resolution of the auditory system to recognize these changes. The effect of human hair to the auditory percept is discussed here in relation to these thresholds.

Published values for just-noticeable-difference (JND) changes in the interaural time difference (ITD) vary depending on the frequency content of the sound stimulus and the reference ITD used [see [Akeroyd \(2006\)](#) for a recent review]. Additional variations result from slight differences in the definition of the JND threshold (typically 75% correct), the stimulus duration, significant intersubject variability, and the stimulus encoding (whether the stimuli contains onset, offset, or an ongoing ITD). Optimum values of JND can be as low as 10 μ s, although certain subjects may exhibit JNDs well beyond this value. Removing one outlying subject, [Mossop and Culling \(1998\)](#) report broadband ITD JNDs on the order of 19 μ s for a 0 μ s reference ITD (frontal source). This increases to 72 μ s for a 600 μ s reference ITD (source near the interaural axis). Equivalent studies of the interaural level difference (ILD) JND consistently report values on the order of 1 dB (e.g., [Mills, 1958](#); [Hershkowitz and Durlach, 1969](#); [Domnitz and Colburn, 1977](#); [Hartmann and Constan, 2002](#); [Bernstein, 2004](#)). The JND is approximately independent of frequency and has only a weak dependence on the ITD reference. The corresponding minimum audible angle is 1° to 2° under optimum circumstances. This resolution intuitively decreases for source angles away from the median axis, for vertical plane localization, and for localization in the presence of noise ([Stern et al., 1983](#); [Grantham et al., 2003](#)). Detection of the direction of movement generally has a higher threshold than the detection of movement ([Chandler and Grantham, 1992](#)).

II. EXTRACTION OF LOCALIZATION CUES

For a spherical head without pinnas or ear canals, the auditory cues are simply dependent on the surface pressure at the equivalent pinna locations. This pressure is calculated here using the analytical result described by [Treeby et al. \(2007a\)](#) assuming a planar incident wave. The formulation assumes that the surface of the spherical scatterer is divided into two hemispheres, each with a uniformly distributed, locally reacting surface impedance [the validity of a locally reactive surface assumption for human hair is discussed by

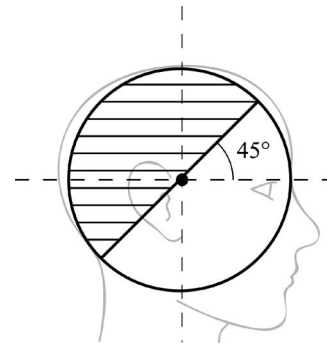


FIG. 1. Sphere scattering model with a hemispherically divided surface covering used to approximate the contribution of hair to the auditory percept. The black dot illustrates the pinna location.

[Treeby et al. \(2007b\)](#)]. If a unity source strength is assumed, at the appropriate surface locations, the complex pressure calculated by this formulation directly corresponds to the HRTF.

The simulated results presented here are based on scattering model parameters representative of human anthropometry. The sphere radius is assumed to be 8.75 cm, the hemispherical impedance boundary (hairline) elevated 45° from the median axis, and the pinnas located in the azimuth plane, offset from the frontal median axis by 90° (see Fig. 1). The symmetric alignment of the hairline and pinnas is chosen so that any asymmetries arising due to the hair covering are easier to distinguish. The effect of the pinna offset in relation to the contribution of hair is discussed in Sec. V A. The lower (facial) hemisphere is assumed to be rigid and the upper hemisphere given several representative values of complex impedance. These regions correspond to the shaded (absorbent) and unshaded (rigid) hemispheres shown in Fig. 1. The calculation of the spherical angles required by the scattering model (which assumes the hemispherical boundary to be coincident with the x - y plane) is facilitated using coordinate system transformation via Euler rotation. This allows the angles for circumferential angular sweeps outside of the axial planes to be easily calculated.

The HRTF typically exhibits idiosyncratic features for source movement throughout three-dimensional space. However, in the absence of more convenient ways to display higher dimensional data, characteristics are normally discussed in relation to source angles within the three elemental planes. For the chosen symmetric hair and pinna alignment, with an appropriate angular remapping the frontal plane is identical to the azimuth plane. Consequently, only changes within the azimuth and median planes will be considered, with the displayed data and discussion corresponding to the head's right ear. These elemental planes are shown in Fig. 2. Whilst the sphere used is pinna-less, the term "ear" is used here to denote the evaluation location in lieu of a more satisfactory description.

The HRTF pairs for left and right pinnas collectively embody the localization cues produced by the spherical head model. However, at lower frequencies it is the interaural cues rather than the monaural spectral details that are of most importance to localization. It is thus appropriate to also examine the effect of hair on these cues. The interaural cues for

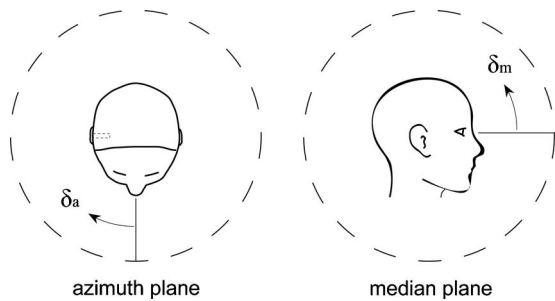


FIG. 2. Spatial reference planes and relative source rotation angles (starting from 0°). For the symmetric hair and pinna alignment used, the frontal plane is identical to the azimuth plane with the appropriate angular remapping.

a particular source location are extracted from the difference in HRTF magnitude and phase information between the two ears. The actual analytical value of ITD differs slightly depending on the method used to extract this disparity (e.g., [Treeby et al., 2007d](#)). The ITD is calculated here using the interaural phase delay. To account for the changes in ITD with frequency, the ITD trends at two characteristic frequencies are discussed (375 and 3000 Hz). These lie within the low and high frequency ITD limits described by [Kuhn \(1977\)](#). Due to the symmetry of the hair covering and pinnas about the median sagittal plane, there is no change in the interaural cues with source movement through the median plane. Changes to the interaural cues are thus only discussed in relation to the azimuth plane.

III. THE EFFECT OF IMPEDANCE ON LOCALIZATION CUES

A. Changes to the head-related transfer function

The changes to the azimuthal and median plane HRTF due to the inclusion of hair are shown in Fig. 3. For convenience, this and all subsequent results discussed in Secs. III A–III C assume a frequency independent hair impedance of $|\zeta|=2$, $\angle\zeta=45^\circ$ (where ζ is the specific acoustic impedance for normal incidence). This corresponds to a relatively dense and absorbent hair surface, and is consistent with the impedance values utilized and discussed by [Treeby et al. \(2007d; a\)](#). Whilst this selection overestimates the magnitude of some changes, it also makes the characteristic trends easier to distinguish. At low frequencies (where hair is much less absorbent), the frequency independent nature of the assumed impedance value produces particularly exaggerated results. Comparative changes and discussion using measured values of hair impedance are provided in Sec. III D. The left panels of Fig. 3 correspond to azimuthal HRTF, and the upper panels to responses for a rigid sphere. As expected, the rigid azimuthal HRTF exhibits a prominent posterior bright spot at 270° and is symmetrical about this angle. Two pronounced ridges of decreased pressure are evident adjacent to this bright spot. As the wavelength is reduced, the in-and-out of phase oscillations that arise due to interactions between symmetric wave paths occur over a shorter spatial distance. Consequently, additional lobes also become apparent at higher frequencies over approximately the same angular re-

gion. For a uniformly distributed surface boundary, moving the source angle around the sphere is equivalent to moving the ear location. The features displayed in the rigid sphere HRTF are thus identical to the stacked surface pressure plot for a rigid sphere discussed by [Treeby et al. \(2007a\)](#).

The central panels of Fig. 3 show the HRTF including hair, and the lower panels the decomposed contribution of the hair covering (calculated by subtracting the rigid response). When the hair is included, the azimuthal HRTF is noticeably perturbed. Significant changes are evident for contralateral angles adjacent to the bright spot. The primary anterior contralateral ridge of reduced pressure ($\sim 280^\circ$) is further decreased (~ 10 dB) relative to the HRTF for a rigid boundary, whilst the posterior ridge ($\sim 260^\circ$) is slightly increased. This results in a significant asymmetry about the contralateral bright spot, with the ridge of decreased pressure noticeably more apparent on the anterior side. The magnitude of the additional anterior contralateral pressure oscillations is also increased, although they oscillate about the same position. In the posterior region, these additional oscillations see a relative positive shift.

The asymmetrical contralateral HRTF changes can be explained by examining the circumferential surface pressure as the source moves around the sphere. Figure 4 shows this pressure magnitude at 2000 Hz for eight angles of source incidence. For contralateral source angles near the interaural axis, the surface pressure is asymmetrical with the primary bright spot lobe favoring the absorbent hemisphere. This is consistent with discussion given by [Treeby et al. \(2007a\)](#). The adjacent pressure nulls are also asymmetrical with the magnitude substantially less on the anterior side, particularly with reference to the equivalent rigid sphere pressure. This explains why the contralateral ridge of decreased pressure evident in the HRTF is more noticeable on the anterior side. The angular locations of the pressure nulls adjacent to the bright spot evident in Fig. 4 for a source angle of 270° are also asymmetric in rotational angle, with the anterior null appearing closer to the interaural axis. This accounts for the relative shift in the contralateral ridges seen in Fig. 3.

In addition to the changes about the contralateral bright spot, HRTF variations are also evident for ipsilateral angles. For the complex impedance phase angle shown in Figs. 3 and 4, this region remains reasonably uniform. However, when the impedance phase angle is decreased, these changes become more perceptible (further discussion on the relationship between hair properties and impedance is given in Sec. III D). For posterior ipsilateral source angles, there is a general decrease in the HRTF magnitude. This decrease is a result of the increased absorption of the frontal surface seen by the source. Again, the changes are asymmetrical about the interaural axis. For certain values of hair impedance, the disparity between anterior and posterior ipsilateral regions can be on the order of several decibels.

HRTF changes in the median plane due to hair are illustrated in the right panels of Fig. 3. For a uniform surface boundary and symmetric pinna alignment, there is no change in the HRTF with source angle. When the hemispherical hair covering is added, asymmetries are created in the scattering surface and HRTF changes are consequently established.

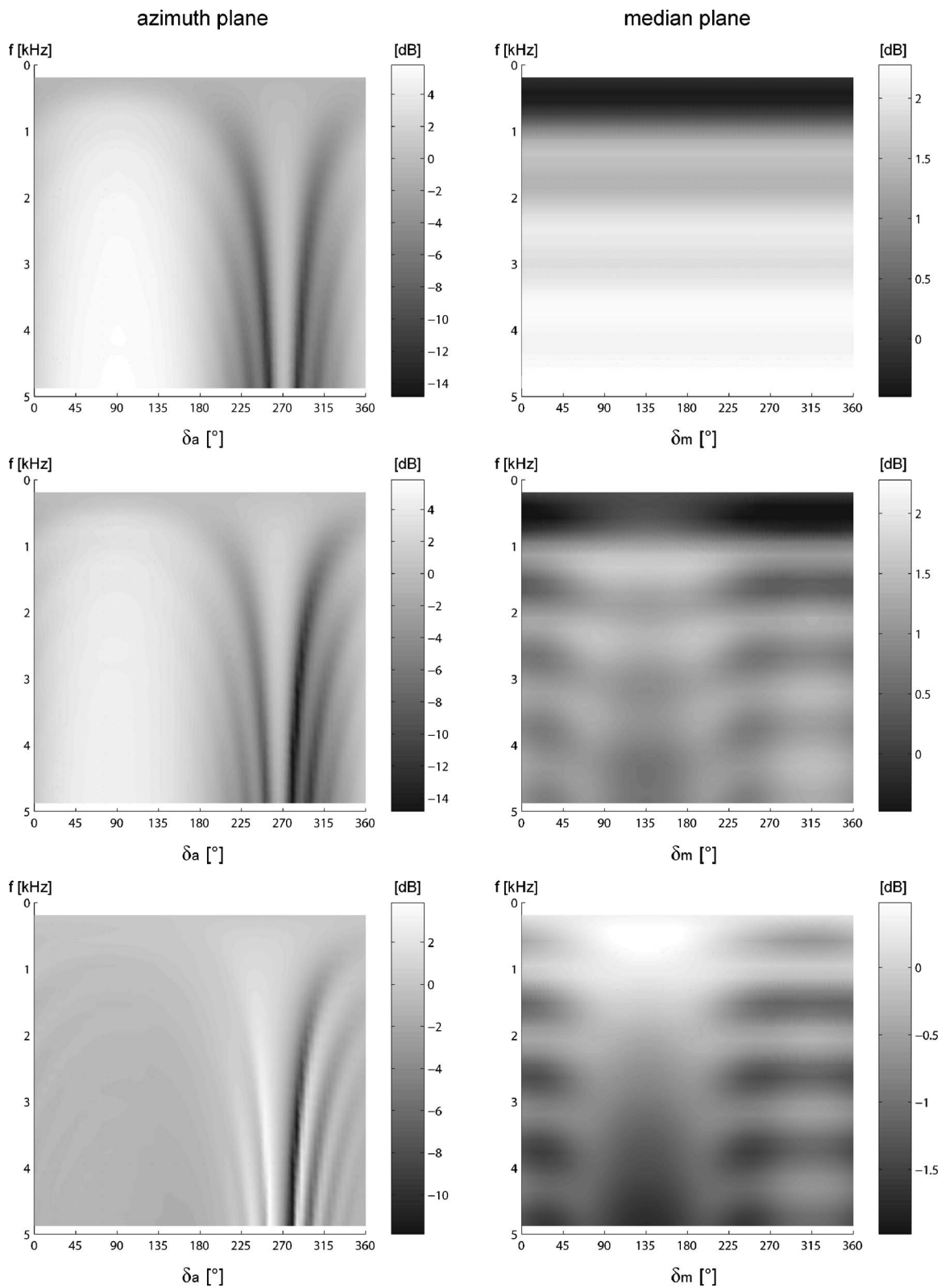


FIG. 3. Changes in the right ear head-related transfer function (HRTF) due to the addition of a hemispherical hair covering with a specific acoustic impedance of $|\zeta|=2$, $\angle\zeta=45^\circ$. The left panels display the azimuthal HRTF and the right panels the median HRTF. The upper panels correspond to a rigid sphere, the central panels include a hemispherical hair covering, and the lower panels show the decomposed HRTF change due to the hair covering.

There is a general decrease in the HRTF magnitude on the order of 1 to 2 dB. This becomes augmented as frequency increases. The changes are symmetric about 135° (or 315°), which corresponds to an axially incident source with refer-

ence to the coordinate system of the scattering model. Overall, the perturbations provided by the hair covering in the median plane are comparable in magnitude to those introduced by the sphere itself.

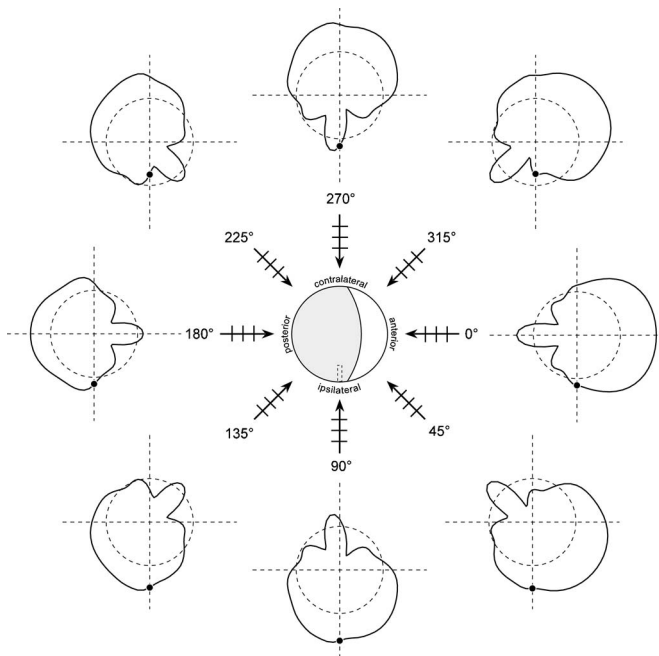


FIG. 4. Azimuthal circumferential surface pressure magnitudes at 2000 Hz for eight angles of source incidence for a spherical head with a hemispherical hair covering (where $|\zeta|=2$, $\angle\zeta=45^\circ$). The black dot on the polar plots corresponds to the pressure at the right ear and the dashed lines illustrate the polar axes and the unit circle for reference.

The general effect of hair on the azimuthal HRTF discussed here is in good agreement with experimental results presented by [Treeby *et al.* \(2007c\)](#) using a rigid spherical head and a hemispherical hair covering. The discussion is also in good agreement with experimental results presented by [Riederer \(2005\)](#), who examined the contribution of several hair coverings on the HRTF of a mannequin. The resolution and frequency range of Riederer's results make direct comparison of data below 5 kHz difficult, however, the general features are clearly evident. There is an increased disparity between the primary pressure nulls adjacent to the contralateral bright spot. For the right ear, this corresponds to a decrease for anterior contralateral angles and a slight increase for posterior, consistent with the present discussion. At higher frequencies, the asymmetric ipsilateral decrease for angles past 90° is also displayed. This posterior ipsilateral reduction becomes particularly augmented above 5 kHz. At very high frequencies (above 12 kHz), additional features are also noticeable, predominantly for posterior contralateral angles where a large decrease in the HRTF magnitude is shown. Again, the resolution of the results prohibit a more detailed explanation. Median plane changes analogous to those discussed here are not perceptible.

Only two additional studies investigating the effect of human hair are known to the authors. Preliminary results presented by [Katz \(2001\)](#) using the boundary element method illustrate that the inclusion of a nonrigid head surface introduces nontrivial variations to the HRTF. These changes are particularly noticeable for contralateral source angles on the interaural axis, consistent with the features discussed in the present study. [Wersényi and Illényi \(2005\)](#) also comment

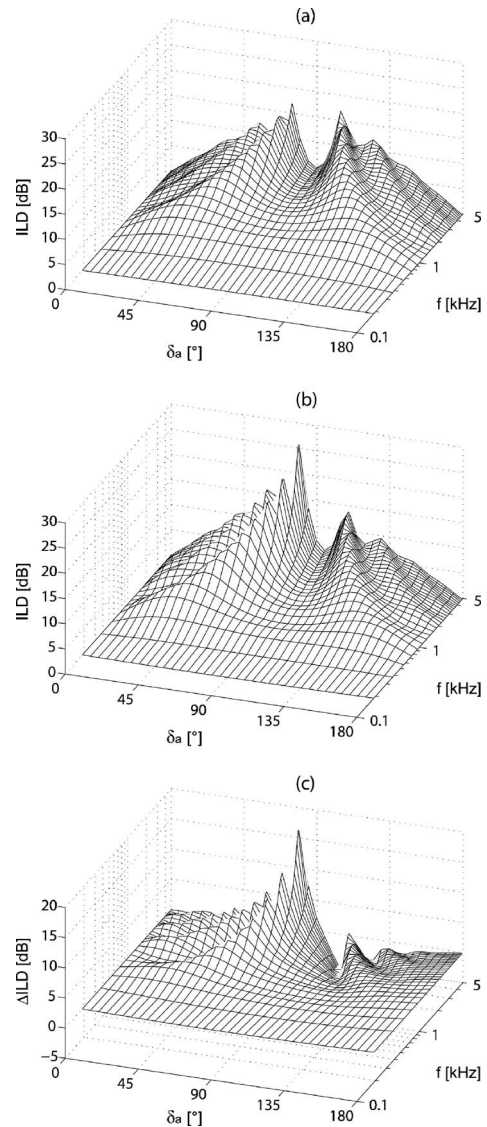


FIG. 5. Changes in the interaural level difference (ILD) due to the addition of a hemispherical hair covering (where $|\zeta|=2$, $\angle\zeta=45^\circ$). (a) ILD for a rigid sphere, (b) ILD including the hemispherical hair covering, and (c) decomposed ILD change due to the hair covering.

that the addition of a hair covering to a mannequin produces HRTF variations with frequency and source angle, including perturbation of the contralateral bright spot.

B. Changes to the interaural level difference

The changes to the ILD due to the addition of a hemispherical hair covering are reasonably intuitive given the previous discussion on the changes to the HRTF magnitude. The left ear response is a reflection of that from the right, and the ILD the difference between these. Figure 5 illustrates the corresponding ILD plots. Given the symmetric pinna alignment, for a rigid sphere [Fig. 5(a)] the ILD is symmetric about 90° . The general decrease in level difference for this angle is a result of the contralateral ear being coincident with the principal bright spot. Due to the symmetry of the scatterer about the median plane, ILDs for source angles past 180° are simply a reflection of those shown in Fig. 5. When the hair covering is added, the asymmetries present in the

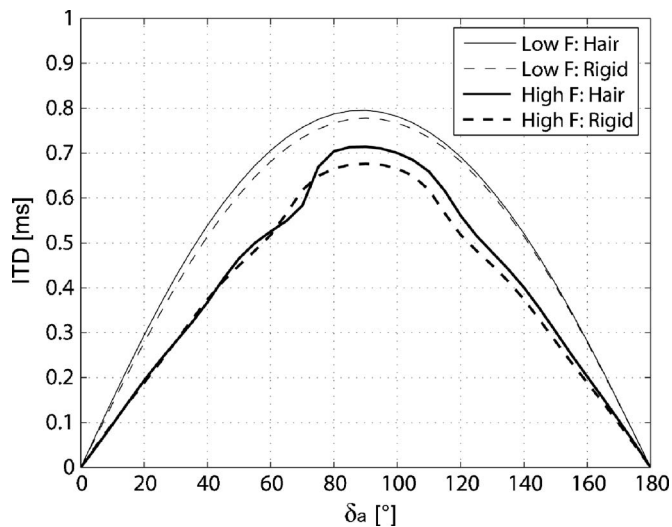


FIG. 6. Low (375 Hz) and high (3000 Hz) frequency interaural time differences (ITD) derived from a spherical head model either completely rigid ($\zeta=\infty$) or with a hemispherical hair covering ($|\zeta|=2, \angle\zeta=45^\circ$).

HRTF also cause the ILD to become asymmetrical. The corresponding ILD is shown in Fig. 5(b), with the decomposed changes due to the hair covering shown in Fig. 5(c). There is an increase in the ILD for source angles where the contralateral ear (left) coincides with a ridge of decreased pressure that has been further decreased. As the source moves past 90° , the contralateral ear coincides with a decreased pressure ridge that has been slightly increased, and the ILD is thus reduced. The asymmetry of the contralateral HRTF ridges with rotation angle also shifts the location of the 90° ILD minimum to source angles slightly posterior of the interaural axis.

C. Changes to the interaural time difference

When the surface is not rigid and the impedance is complex, the surface velocity is no longer in phase with the incoming pressure oscillations. As the ITD is dependent on the relative phase between two separate locations on the head (sphere), a shift in phase of the overall surface pressure at these locations will alter the ITD. Figure 6 shows the relative phase delays for both a rigid sphere and one including the hemispherical hair covering. The upper curves show the ITD at low frequency (375 Hz) and the lower curves at high frequency (3000 Hz). The low frequency ITD shows an increase due to the hair, but when using representative frequency dependent impedance characteristics, this difference becomes negligible (the surface is approximately rigid at low frequencies). At high frequencies, the inclusion of the hair covering generally increases the ITD. Near the interaural axis, these changes are on the order of $30\text{--}40\ \mu\text{s}$. They are asymmetrical about 90° and are greater for source angles past 75° .

For a uniformly covered sphere, a general increase in ITD is seen with a decrease in impedance magnitude, particularly for source angles near the interaural axis (Treeby *et al.*, 2007d). The difference in ITD between a completely rigid and highly absorbent sphere ($\zeta=1$) is around $100\ \mu\text{s}$. This value corresponds to an upper bound to the ITD pertur-

bation expected from a hemispherical hair covering. As discussed by Treeby *et al.* (2007d), altering the phase angle of the impedance in either direction produces a relative reduction in the ITD. This reduction is a maximum for acute impedance phase angles and source angles near the interaural axis. For realistic impedance characteristics, this can reduce the relative ITD increase by as much as 50% (when compared to a purely resistive impedance of the same magnitude).

D. Cue changes for measured values of human hair impedance

As discussed in detail by Treeby *et al.* (2007b), the measured equivalent acoustic impedance characteristics of human hair are primarily dependent on the overall sample thickness and density. Increasing either the bulk thickness or density of the sample decreases the impedance magnitude, whilst an increase in thickness additionally produces a relative increase in the impedance phase angle. For all pragmatic modifications, the equivalent impedance retains a stiffness-like reactance, with the impedance phase angle on the order of $10^\circ\text{--}50^\circ$. Figure 7 illustrates the change in HRTF due to a hemispherical hair covering utilizing three of the measured equivalent impedance properties discussed by Treeby *et al.* (2007b). These correspond to human hair within a 20 mm sample holder at $40\ \text{kg}/\text{m}^3$ [Fig. 7(a)], a 20 mm sample holder at $80\ \text{kg}/\text{m}^3$ [Fig. 7(b)], and a 40 mm sample holder at $40\ \text{kg}/\text{m}^3$ [Fig. 7(c)]. The form of the HRTF changes for all impedance values is consistent with the discussion given in Sec. III A. The corresponding ILD changes [not shown but similar in form to those in Fig. 5(c)] are a maximum for Fig. 7(c) (thicker hair sample) and are on the order of 4 dB. These results are limited by the frequency range of the measured impedance data, but the trends shown in Fig. 5 and the experimental results presented by Riederer (2005) indicate that these changes are further augmented at higher frequencies. Given an ILD JND of approximately 1 dB, it can be concluded that the changes to the ILD due to hair are in excess of the noticeable thresholds, particularly for source locations near the interaural axis.

The corresponding high frequency ITD changes are shown in Fig. 8. For all impedance values, these changes are consistent with the discussion given in Sec. III B. Again, the changes are a maximum for the thicker sample and are on the order $20\text{--}25\ \mu\text{s}$ for source angles past 75° . These values are below the broadband ITD JND values discussed by Mossop and Culling (1998) ($43\ \mu\text{s}$ for a $400\ \mu\text{s}$ reference ITD, $72\ \mu\text{s}$ for a $600\ \mu\text{s}$ reference). Whilst some studies have published ITD JND for low frequency pure tones near the interaural axis below $20\ \mu\text{s}$ (Hershkowitz and Durlach, 1969; Domnitz and Colburn, 1977), it remains unlikely that the inclusion of hair will produce a discernible shift in source location. In combination, the overall changes to the timing and magnitude properties of the HRTF due to hair will most likely introduce subtle audible features of a spectral nature. This supposition is somewhat confirmed by localization testing completed by Riederer (2005) using mannequin HRTF with various head treatments. Untrained listeners were able to perceive minor perceptual differences for large modifica-

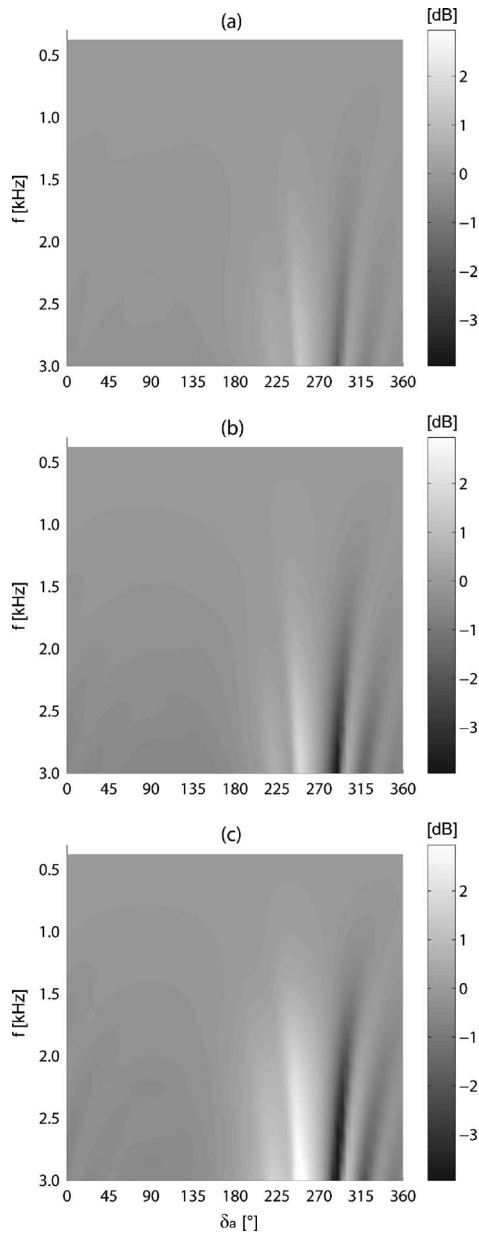


FIG. 7. Decomposed changes in the right ear HRTF due to the addition of a hemispherical covering of hair. The utilized impedance values correspond to human hair samples measured with varying sample thickness and bulk density, (a) 20 mm, 40 kg/m³, (b) 20 mm, 80 kg/m³, and (c) 40 mm, 40 kg/m³.

tions to the head covering. Overall, the inclusion of the hair did not improve localization performance (using nonindividualized HRTF), although the performance with and without hair coverings was weakly dissimilar indicating that the hair may also introduce some diminutive localization cues.

IV. EXPERIMENTAL VALIDATION OF THE IMPEDANCE SCATTERING PROBLEM

A. Experimental equipment and method

To validate the use of the hemispherical divided scattering solution to compute the auditory cue changes due to hair, a series of sphere scattering experiments were completed. These utilized a wooden sphere with an enlarged radius (12.4 cm) to obtain HRTF measurements both with and with-

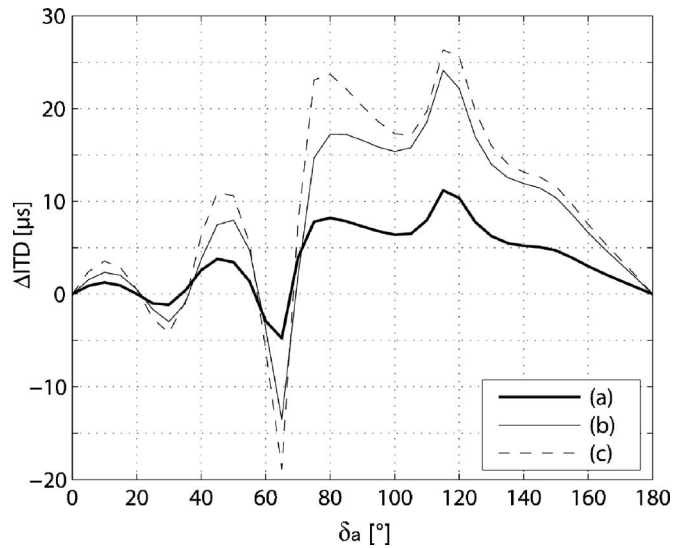


FIG. 8. Decomposed changes in the high frequency (3000 Hz) ITD due to the addition of a hemispherical hair covering to a spherical head model. The utilized impedance values correspond to human hair samples measured with varying sample thickness and bulk density, (a) 20 mm, 40 kg/m³, (b) 20 mm, 80 kg/m³, and (c) 40 mm, 40 kg/m³.

out a covering of synthetic hair material. The sphere was constructed to allow enough space (including cable relief) for two approximately diametrically opposed internal $\frac{1}{2}$ in. microphones and was supported by a thin steel rod which allowed rotation. Measurements were taken from one internal microphone (BSWA Tech MA211) which was positioned flush with the outside of the sphere surface for all experiments. All measurements were taken at 5° increments of sphere rotation starting with the internal microphone facing the frontal incident wave direction. The rotation angle was aligned using a laser level positioned at the base of the sound source (located approximately 3 m from the sphere) in conjunction with degree markings on the rotating sphere stand. The experiments were taken within an anechoic environment, and exposed areas of the stand were covered with a thick layer of highly absorbent material to minimize additional reflections.

Two series of measurements were taken starting with reference measurements of the rigid sphere. These were then repeated with the sphere hemispherically covered by a synthetic hair material. The equivalent impedance (impedance at a reference plane coincident with the inner rigid surface) and discussion of the acoustical properties of this material are given by [Treeby et al. \(2007b\)](#). The hair material came attached to a thin fabric backing and the covering was constructed from several pieces of this sewn together to make the correct shape. The covering was tailored circumspectly to maintain the overall distribution of the hair fibers, and so that it fitted neatly over the sphere surface without any significant deformation. A thin double-sided tape was used to adhere it securely. The hair line was elevated 45° from the median axis, with the internal microphone located along the hair boundary consistent with the alignment shown in Fig. 1. The hair covering was trimmed so that it did not cover the microphone as shown in Fig. 9.



FIG. 9. Experimental setup showing the wooden sphere with the synthetic hair covering. The hairline is elevated 45° and the microphone (ear) is set within the azimuth plane, offset from the frontal median axis by 90° .

For each test, impulse response measurements were obtained using maximum length sequences produced by the Brüel & Kjær DIRAC software and a Brüel & Kjær HP1001 unidirectional sound source. A sequence length of $2^{14}-1$ (the shortest available sequence length) with ten averages and a sampling frequency of 48 kHz was used. To remove the effects of the imperfectly anechoic measurement environment (the door surface was not properly treated), the impulse response peak onsets were located, and the tails then truncated to 128 samples and windowed using the second half of a cosine-tapered (Tukey) window with a 25% taper ratio. Each impulse response was then shortened to 256 samples (with

the timing information preserved) and converted to the frequency domain using a 256 point FFT. Extraction of experimental ILD and ITD information was done in the same manner as the analytical predictions.

For each of the experimental tests, comparative analytical results were derived using the equivalent impedance of the synthetic hair covering. As the experimental setup only facilitated azimuthal measurements, comparison of median plane predictions was not possible. Both experimental and analytical results were processed for frequencies from 375 to 3000 Hz. This corresponds to the available range of impedance data. Due to the enlarged radius of the test sphere, the equivalent frequency range for a head radius of 8.75 cm is 530–4250 Hz. Results shown in the following section are not frequency scaled.

B. Results for a hemispherical hair covering

Figure 10 shows the decomposed contribution of the synthetic hair covering to the HRTF and ILD. The left panels show the experimental response, and the right the corresponding analytical results. The main features of the experimental and analytical results are in good agreement and are consistent with previous discussion. The principal differences occur for posterior source angles, where the experimental HRTF changes illustrate a slightly decreased pressure magnitude and additional oscillatory behavior (evident for contralateral source angles). Consequently, the corresponding experimental and analytical results for the change in ILD also differ slightly for source angles past 90° . Predominantly, however, the analytical predictions account for the experi-

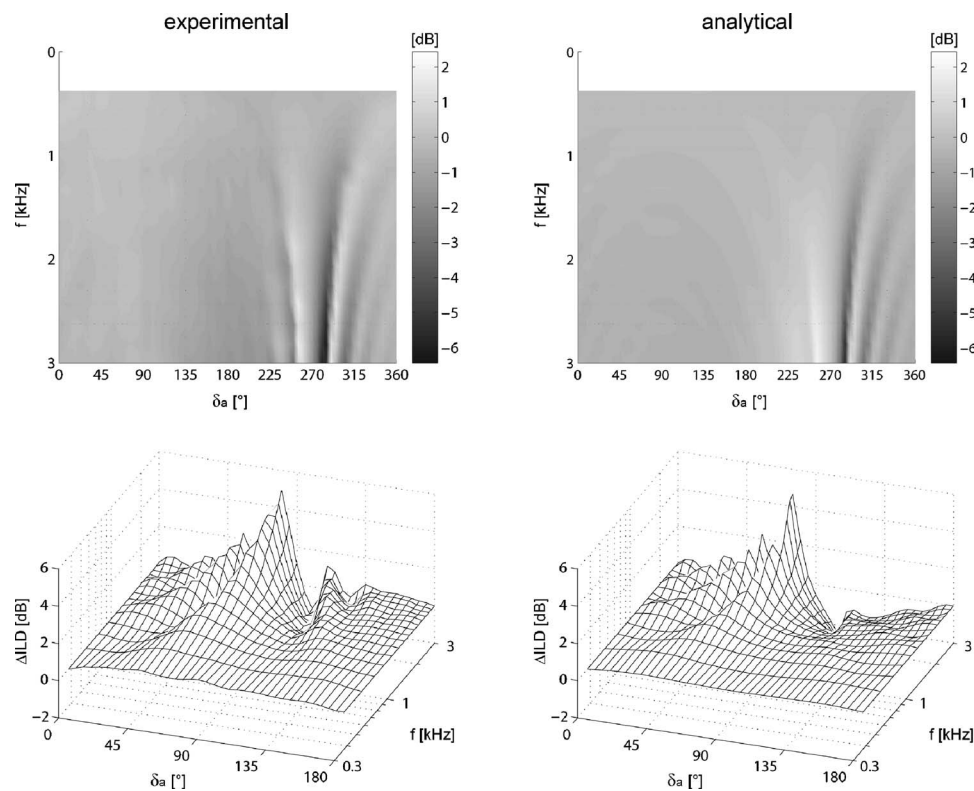


FIG. 10. Comparison of experimental (left panels) and analytical (right panels) results for the decomposed contribution of the synthetic hair material to the right ear HRTF and ILD.

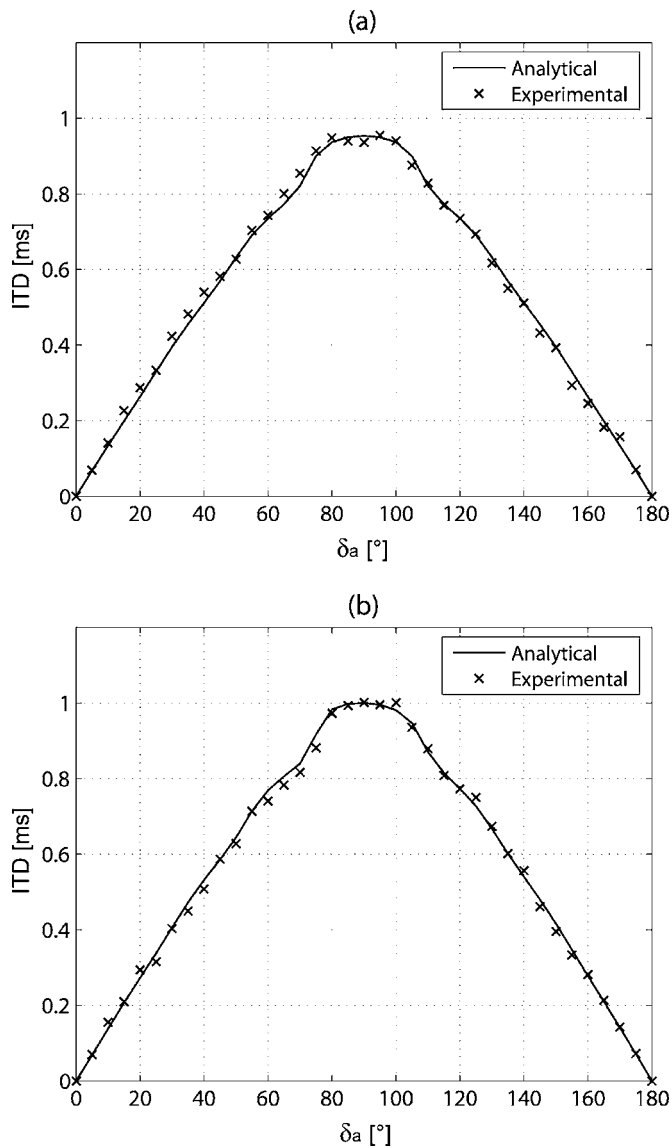


FIG. 11. Comparison of experimental and analytical values of the high frequency (3000 Hz) interaural time difference (ITD) for (a) a rigid sphere and (b) a rigid sphere with a hemispherical covering of synthetic hair.

mental changes. The variations may be attributed to slight inaccuracies in the utilized impedance characteristics of the hair covering, and experimental errors in maintaining exactly consistent rotation angles between tests with and without the hair covering. A comparison of analytical and experimental results for the high frequency ITD (3000 Hz) is shown in Fig. 11. For both the rigid sphere and the synthetic hair covering, the experimental and analytical results are in good agreement. For source angles near the interaural axis, the ITD is increased on the order of 30–50 μ s due to the addition of the hair covering. At low frequencies, the changes are negligible.

The synthetic hair sample utilized has a natural bulk density of approximately 30 kg/m³. This is at the mid to lower end of the pragmatic range expected for representative human subjects (Treeby *et al.*, 2007b). As discussed in Sec. III D, utilizing equivalent impedance values from denser or thicker hair samples yields even greater changes to the localization cues. The form of these changes, however, remains

the same, with the hair covering producing asymmetrical changes to the interaural cues between source directions in the frontal and rear hemifields. Analytical comparisons with less formal experimental measurements using an additional synthetic hair covering (which was comparatively much thicker and denser) showed that the predicted cue changes tended to underestimate the measured results. This is most likely due to the inability of the locally reactive equivalent impedance parameter to completely encapsulate the wave processes through the hair material.

V. THE CONTRIBUTION OF HAIR IN RELATION TO OTHER PERIPHERAL FEATURES

The contribution of hair to the auditory cues has thus far been discussed in relation to a single detached sphere with pinnae located on the interaural axis. Human subjects differ from this model in both head shape and pinna location. Additionally, the head is connected to other peripheral scattering bodies, for example the neck. It is important to examine whether the contribution of the hair covering remains consistent when these auxiliary features are included. The relative contributions of two additional anthropometric features are investigated in the subsequent sections. First, the effect of offsetting the pinnae to better match their physiological location is assessed. Second, the effect of connecting a semi-infinite neck to the spherical head is discussed in relation to additional experiments. The stimulus of these investigations is not to provide comprehensive discussion on the individual contributions of these ancillary features, but rather to study whether the contribution of the hair covering remains consistent and pertinent when they are included.

A. Effect of pinna offset

The location of the pinnae on an anthropometric head is commonly cited as being both downwards and backwards from the head center. For an anthropometric head shape, personalized measurements enable a relatively concise pinna position to be determined. However, when the head is modeled as a sphere this becomes considerably more complex. The spherical shape does not necessarily provide an accurate fit to the head length, width, and height measurements in isolation. Examining the relevant anthropometric data (Burkhard and Sachs, 1975; Farkas, 1981; Pheasant, 1986; Burandt *et al.*, 1991; Dreyfuss, 2002), the average ratio of auricular head height to craniofacial head height is 0.57. Assuming an 8.75 cm sphere radius, this proportionally corresponds to a 1.33 cm (8.7°) downward vertical offset of the pinnae. This offset represents the correct proportional vertical offset if the complete facial profile is taken into consideration. However, customized sphere radii are typically more dependent on the head length than the overall height of the facial profile (Algazi *et al.*, 2001b). Using the same anthropometric data, the average ratio of auricular head height to the head length is 0.68. If the sphere is aligned with the upper head surface, this ratio yields a downward vertical offset of 3.12 cm (20.9°), which is significantly more. Depending on the use for the spherical HRTF data, a value between these would most likely be appropriate. Examining

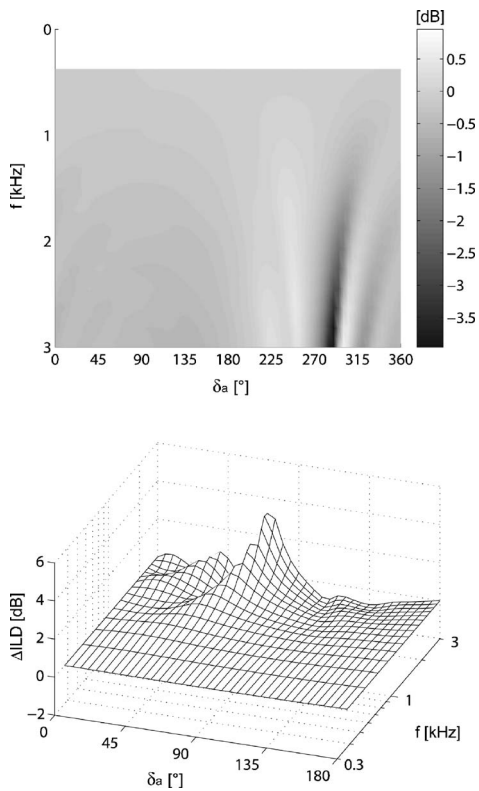


FIG. 12. The decomposed contribution of the synthetic hair material to the right ear HRTF and ILD relative to a rigid spherical head model with offset pinnae.

the horizontal offset, the average ratio of head-ear depth to head length is 0.51. Proportionally, this corresponds to a forward horizontal offset of 0.23 cm (1.5°). Assuming the same pinna position, if the sphere is aligned with the frontal head surface this becomes a backward horizontal offset of 0.46 cm (3.0°). For a spherical approximation of the head, the location of the pinnas is much closer to the interaural axis in horizontal location than in vertical. These values are consistent with [Algazi et al. \(2001a\)](#), who reported measured pinna offsets from an anthropometric mannequin for use with spherical head models to be 0.5 cm backwards and 3 cm downwards from the interaural axis (assuming a sphere radius of 8.5 cm).

To investigate the relative effect of the pinna position on the HRTF, the pinnas are assumed here to be located 10° downward and 1.5° backward of the interaural axis. This is comparable to aligning the spherical head slightly upward and forward of the geometric center of an anthropometric head. The impedance properties of the synthetic hair material are again used to characterize the contribution of hair covering. Figure 12 illustrates the decomposed effect of the hair covering in the azimuth plane relative to a rigid sphere with offset pinnas. The contribution of the hair covering remains robust regardless of the pinna offset, and the HRTF and ILD asymmetries are still clearly evident. The changes to the ITD also remain robust, with asymmetrical increases to the high frequency ITD similar in form to those shown in Figs. 6 and 8. In the median plane (not shown), the decomposed perturbations due to the hair covering are also consistent, however, the effect of the pinna offset dominates that of the hair.

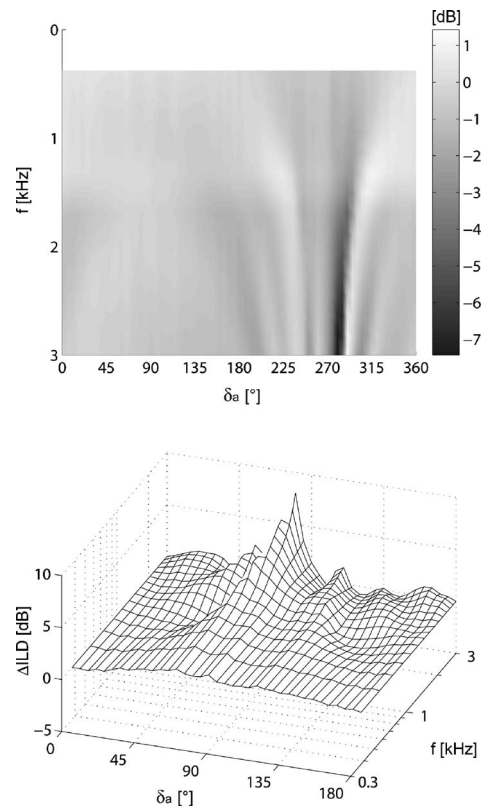


FIG. 13. Experimental changes to the right ear HRTF and ILD due to the addition of a hemispherical hair covering and a cylindrical neck to a rigid spherical head.

B. Effect of neck

If the head is examined in isolation, the major adjoining physiological features are the pinnas, face, hair, and neck. The relatively small nature of the pinnas and facial features means they contribute little to the scattering properties of the head at lower frequencies (below 3 to 4 kHz). However, it is useful to investigate the perturbation of the auditory cues by the neck, particularly in relation to the contribution of hair. [Treeby et al. \(2007c\)](#) experimentally examined the HRTF from a rigid sphere with a cylindrical neck and a hemispherical hair covering. Decomposition illustrated a reduction of the posterior bright spot due to addition of the neck on the order of 2–4 dB. This is consistent with expectations from anthropometric HRTF. The addition of the hair produced changes consistent with the discussion given here. The results from an additional but equivalent set of experiments are shown in Fig. 13. These use a 1.4 m long PVC pipe “neck” in place of the rotating sphere stand, 8 cm in radius with 1.2 cm thick walls. This corresponds to a 0.65 neck to sphere radius ratio, which is close to the anthropometric ratio of 0.675 if a 8.75 cm head radius is assumed. Whilst a 1.4 m neck does not match human physiology, it was chosen in favor of a shorter length to eliminate the effects of end scattering on the response (which also do not occur in human HRTF). Impulse response measurements were repeated for the sphere and neck arrangement, both with the hair covering and without.

The primary changes to both the HRTF and ILD shown in Fig. 13 are well explained by the individual contributions

of the two added features. There is a reduction in the contralateral bright spot due to the neck, and additional asymmetrical perturbations for both ipsilateral and contralateral angles consistent with the discussion on the effect of the hair covering. The HRTF perturbations produced by the hair are of comparable magnitude to the neck, and in the anterior contralateral region are significantly more. With regards to the ITD, the addition of the neck in isolation symmetrically reduces the overall value by up to 40 μ s. The subsequent addition of the hair produces relative ITD changes similar to those already discussed. Overall, the addition of both the neck and hair still increases the ITD, but only by a small amount (up to 20 μ s).

VI. SUMMARY AND DISCUSSION

This study utilizes a recently published analytical sphere scattering model to investigate the effect of hair on human auditory cues, a subject that has previously received little attention by the greater binaural community. The hair is shown to produce asymmetric perturbations to both the HRTF and interaural difference cues. In the azimuthal plane, the HRTF changes are characterized by two predominant features. First, the ridges of decreased pressure adjacent to the contralateral bright spot become asymmetrical, with the ridge noticeably more apparent on the anterior side. Second, for posterior ipsilateral source angles, there is a general decrease in the HRTF magnitude. This is a result of the increased absorption of the frontal surface seen by the source. Overall, the inclusion of a representative hair surface produces asymmetrical changes to the ILD and ITD on the order of 4 dB and 25 μ s, respectively (only frequencies up to 3 kHz are investigated due to the range of the available impedance data). These modifications remain robust regardless of the decomposition baseline (i.e., the inclusion of additional anthropometric features). In the azimuth plane, the addition of a cylindrical neck and the introduction of a pinna offset do not significantly influence the contribution from the hair. In the median plane, the HRTF changes introduced by the hair are similar in magnitude to those introduced by the sphere itself. However, the introduction of a pinna offset appears to modify the HRTF in this plane to a much greater extent.

The analytical results are experimentally validated using a series of azimuthal HRTF measurements from a sphere with and without a hemispherical hair covering. These results show a good agreement with analytical results for the same hair material. This mutually validates the use of the analytical formulation presented by [Treeby *et al.* \(2007a\)](#), and the equivalent impedance values given by [Treeby *et al.* \(2007b\)](#) for modeling the acoustic contribution of human hair. The additional trends in equivalent impedance for representative human hair samples discussed by [Treeby *et al.* \(2007b\)](#) subsequently allows simulation of the contribution of hair for a wide range of individuals. It should be noted that it is not difficult to think of head and hair characteristics that are not well approximated by a sphere with a hemispherically divided boundary condition (e.g., people who are partially bald, or have thick beards, etc). However, for physi-

cal understanding it is favorable to make astute simplifying assumptions that enable analytical investigation rather than attempt an exhaustive empirical study. As a sphere is only an approximation of the human head, a distribution of impedance that assimilates the general characteristics of the location of the hair is sufficient to investigate its contribution. In any case, for a large majority of people the boundary condition utilized here provides an adequate approximation. This boundary distribution may also be useful to simulate the contribution of other head coverings such as beanies (toques). Although such an addition may perturb the natural protrusion of the pinna, this is of little significance at lower frequencies.

The small magnitude of the cues produced by human hair (only the ILD changes are above JND thresholds) makes it unlikely that this topographical feature is significant in shaping the auditory percept. Given the relative plasticity of the auditory system to adapt to linear cue transformations with short term training (e.g., [Shinn-Cunningham *et al.*, 2005](#)) and complex transformations with long term training (e.g., [Hofman *et al.*, 1998](#)), the argument for including the effects of hair in auditory cue models is not strong [see [Wright and Zhang \(2006\)](#) for a recent review of auditory cue adaptation studies]. In contrast, however, for untrained listeners, even small perturbations to individualized HRTF can decrease localization accuracy within a virtual environment (e.g., [Wenzel *et al.*, 1993](#)). The contribution from hair, in combination with other peripheral and detailed features, may thus be important for accurately maintaining the spatial cues that a listener “normally” experiences. In relation to the use of spherical head models (particularly in isolation), the directional ambiguities created by the assumption of a rigid boundary limits their practical use. In this regard, the asymmetries introduced by the hair covering may assist in the discrimination between sources positioned in the front and rear hemifields. This may be particularly useful in the absence of other high frequency cues, or access to cue changes with head movement. In any case, the current study serves to clearly identify, quantify, and explicate the HRTF features that occur due to human hair.

ACKNOWLEDGMENTS

The authors would like to thank Frances Dooney for the construction of the synthetic hair covering. B.E.T. would also like to acknowledge the financial support of the Robert and Maude Gladden, and F. S. Shaw Memorial Postgraduate Scholarships.

- Akeroyd, M. A. (2006). “The psychoacoustics of binaural hearing,” *Int. J. Audiol.* **45**, S25–S33.
- Algazi, V. R., Avendano, C., and Duda, R. O. (2001a). “Elevation localization and head-related transfer function analysis at low frequency,” *J. Acoust. Soc. Am.* **109**, 1110–1122.
- Algazi, V. R., Avendano, C., and Duda, R. O. (2001b). “Estimation of a spherical-head model from anthropometry,” *J. Audio Eng. Soc.* **49**, 472–479.
- Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z. (2002a). “Approximating the head-related transfer function using simple geometric models of the head and torso,” *J. Acoust. Soc. Am.* **112**, 2053–2064.
- Algazi, V. R., Duda, R. O., and Thompson, D. M. (2002b). “The use of

- head-and-torso models for improved spatial sound synthesis," Proceedings of the 113th AES Convention, Audio Engineering Society, Los Angeles, CA, preprint 5712.
- Bernstein, L. R. (2004). "Sensitivity to interaural intensive disparities: Listeners' use of potential cues," *J. Acoust. Soc. Am.* **115**, 3156–3160.
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge).
- Brown, C. P., and Duda, R. O. (1998). "A structural model for binaural sound synthesis," *IEEE Trans. Speech Audio Process.* **6**, 476–488.
- Brungart, D. S., and Rabinowitz, W. M. (1999). "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.* **106**, 1465–1479.
- Burandt, U., Posselt, C., Ambrozus, S., Hosenfeld, M., and Knauff, V. (1991). "Anthropometric contribution to standardising manikins for artificial-head microphones and to measuring headphones and ear protectors," *Appl. Ergon.* **22**, 373–378.
- Burkhard, M. D., and Sachs, R. M. (1975). "Anthropometric manikin for acoustic research," *J. Acoust. Soc. Am.* **58**, 214–222.
- Chan, C.-T., and Chen, O. T.-C. (2000). "A 3D sound using the adaptive head model and measured pinna data," *IEEE International Conference on Multimedia and Expo*, New York, pp. 807–810.
- Chandler, D. W., and Grantham, D. W. (1992). "Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity," *J. Acoust. Soc. Am.* **91**, 1624–1636.
- Domnitz, R. H., and Colburn, H. S. (1977). "Lateral position and interaural discrimination," *J. Acoust. Soc. Am.* **61**, 1586–1598.
- Dreyfuss, H. (2002). *The Measure of Man and Woman: Human Factors in Design* (Wiley, New York).
- Duda, R. O., and Martens, W. L. (1998). "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.* **104**, 3048–3058.
- Farkas, L. G. (1981). *Anthropometry of the Head and Face in Medicine* (Elsevier, New York).
- Grantham, D. W., Hornsby, B. W. Y., and Erpenbeck, E. A. (2003). "Auditory spatial resolution in horizontal, vertical, and diagonal planes," *J. Acoust. Soc. Am.* **114**, 1009–1022.
- Hartmann, W. M., and Constan, Z. A. (2002). "Interaural level differences and the level-meter model," *J. Acoust. Soc. Am.* **112**, 1037–1045.
- Hershkowitz, R. M., and Durlach, N. I. (1969). "Interaural time and amplitude jnds for a 500-Hz tone," *J. Acoust. Soc. Am.* **46**, 1464–1467.
- Hofman, P. M., Van Riswick, J. G. A., and Van Opstal, A. J. (1998). "Re-learning sound localization with new ears," *Nat. Neurosci.* **1**, 417–421.
- Katz, B. F. G. (2000). "Acoustic absorption measurement of human hair and skin within the audible frequency range," *J. Acoust. Soc. Am.* **108**, 2238–2242.
- Katz, B. F. G. (2001). "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparison to real measurements," *J. Acoust. Soc. Am.* **110**, 2449–2455.
- Kuhn, G. F. (1977). "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.* **62**, 157–167.
- Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (1999). "Sensitivity of human subjects to head-related transfer-function phase spectra," *J. Acoust. Soc. Am.* **105**, 2821–2840.
- Mills, A. W. (1958). "On the minimum audible angle," *J. Acoust. Soc. Am.* **30**, 237–246.
- Mossop, J. E., and Culling, J. F. (1998). "Lateralization of large interaural delays," *J. Acoust. Soc. Am.* **104**, 1574–1579.
- Pheasant, S. (1986). *Bodyspace: Anthropometry, Ergonomics and Design* (Taylor and Francis, London).
- Riederer, K. A. J. (2005). "HRTF analysis: Objective and subjective evaluation of measured head-related transfer functions," Ph.D. dissertation, Helsinki University of Technology, Helsinki.
- Shaw, E. A. G. (1997). "Acoustical features of the external ear," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah), pp. 25–47.
- Shinn-Cunningham, B. G., Streeter, T., and Gyss, J.-F. (2005). "Perceptual plasticity in spatial auditory displays," *ACM T. Appl. Percept.* **2**, 418–425.
- Stern, R. M. J., Slocum, J. E., and Phillips, M. S. (1983). "Interaural time and amplitude discrimination in noise," *J. Acoust. Soc. Am.* **73**, 1714–1722.
- Treedy, B. E., Pan, J., and Paurobally, R. M. (2007a). "Acoustic scattering by a sphere with a hemispherically split boundary condition," *J. Acoust. Soc. Am.* **122**, 46–57.
- Treedy, B. E., Pan, J., and Paurobally, R. M. (2007b). "An experimental study of the acoustic impedance characteristics of human hair," *J. Acoust. Soc. Am.* **122**, 2107–2117.
- Treedy, B. E., Paurobally, R. M., and Pan, J. (2007c). "Decomposition of the HRTF from a sphere with neck and hair," Proceedings of the 13th International Conference on Auditory Display, Montreal, Canada, pp. 79–84.
- Treedy, B. E., Paurobally, R. M., and Pan, J. (2007d). "The effect of impedance on interaural azimuth cues derived from a spherical head model," *J. Acoust. Soc. Am.* **121**, 2217–2226.
- Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.* **94**, 111–123.
- Wersényi, G., and Illényi, A. (2005). "Differences in dummy-head HRTFs caused by the acoustical environment near the head," *Electronic Journal, Technical Acoustics* **1**, 1–15.
- Wright, B. A., and Zhang, Y. (2006). "A review of learning with normal and altered sound-localization cues in human adults," *Int. J. Audiol.* **45**, S92–S98.
- Zotkin, D. N., Duraiswami, R., Grassi, E., and Gumerov, N. A. (2006). "Fast head-related transfer function measurement via reciprocity," *J. Acoust. Soc. Am.* **120**, 2202–2215.
- Zotkin, D. N., Hwang, J., Duraiswami, R., and Davis, L. S. (2003). "HRTF personalization using anthropometric measurements," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE, New Paltz, NY, pp. 157–160.

Spectral integration and wideband analysis in gap detection and overshoot paradigms

Joseph W. Hall III,^{a)} Emily Buss, and John H. Grose

Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 24 October 2006; revised 13 September 2007; accepted 13 September 2007)

Several listening conditions show that energy remote from a target frequency can deleteriously affect sensitivity. One interpretation of such results entails a wideband analysis involving a wide predetection filter. The present study tested the hypothesis that both temporal gap detection and overshoot results are consistent with a wideband analysis, as contrasted with statistical combination of information across independent channels. For gap detection, stimuli were random or comodulated 50-Hz-wide noise bands centered on 1000, 1932, 3569, and 6437 Hz. For overshoot, the masker was an 8-kHz low-pass filtered noise, with 5-ms tone bursts presented at the same center frequencies used for gap detection. Signals were presented with either 0- or 250-ms delay after masker onset. In each paradigm, the target was introduced at only one frequency or at all four frequencies. Results from gap detection conditions did not favor a wideband analysis interpretation: Results in the random condition were consistent with an optimal combination of cues across frequency. An across-channel interference effect was also evident when only one of the four bands contained the gap. Although results from the overshoot conditions were consistent with a wideband analysis interpretation, they were more parsimoniously accounted for in terms of statistical combination of information.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799473]

PACS number(s): 43.66.Dc, 43.66.Ba, 43.66.Mk [RLF]

Pages: 3598–3608

I. INTRODUCTION

Most sounds in natural listening situations are spectrally complex, and the auditory system is adept at combining information from across the audio frequency range. This combination of information from across frequency has been referred to in several ways, depending in part on the particular auditory function under investigation and the paradigm employed to study it. The present study is concerned with two general mechanisms that have been proposed to underlie the combination of information across frequency, “wideband analysis” and “spectral integration.” As developed further in the following, the term wideband analysis is used to refer to a process compatible with a predetection filter having a bandwidth of at least several thousand hertz, and spectral integration refers to a process compatible with the statistical combination of information from two or more spectrally independent channels. This study considers these mechanisms within the specific contexts of gap detection and overshoot.

Gap detection results for spectrally complex markers often appear to be compatible with the assumption of a wideband analysis. Two characteristics are usually implied by this term. First, the frequency selectivity of the auditory periphery can be considered largely inconsequential to the gap detection performance for the majority of stimuli. The apparent absence of a peripheral filter effect has been demonstrated in a number of studies. For example, it has been reported that the function relating gap detection to the bandwidth of the noise that marks the gap is independent of the frequency

region being tested (Eddins *et al.*, 1992). Eddins *et al.* found that increasing the noise bandwidth from 50 to 400 Hz resulted in approximately the same improvement in gap detection whether the stimulus was centered on 4200 Hz (where the 400-Hz wide noise was within a single peripheral filter), or centered on 400 Hz (where the 400-Hz wide noise spanned multiple peripheral filters). The second characteristic implied by a wideband analysis as applied to gap detection is that the combination of information across frequency is obligatory. For example, the detection of a gap imposed on a noise band in one frequency region can be deleteriously affected by the presence of a nongapped noise band in a spectrally remote region (e.g., Grose and Hall, 1993; Moore *et al.*, 1993). Other temporal resolution data have also been modeled successfully with a system featuring a single broad predetection filter that exceeds in bandwidth any reasonable measure of peripheral auditory filtering (e.g., Viemeister, 1979; Green and Forrest, 1989). According to such a model, stimulus energy is pooled within a predetection filter and processed as a single channel, regardless of the frequency selectivity that can be demonstrated with other paradigms (e.g., notched-noise masking). It should be noted that although many temporal resolution findings are compatible with the above wideband interpretation, the results of experiments in which the leading and trailing gap markers are separated in frequency (e.g., Phillips *et al.*, 1997; Formby *et al.*, 1998a, b; Grose *et al.*, 2001) are not. In these experiments, spectral separations between the gap markers can result in very poor gap detection, a finding that would not be expected under the assumption of a wideband analysis.

^{a)}Electronic mail: jwh@med.unc.edu

Experiment 1 of the present study examines gap detection in a paradigm where bursts of four widely spaced narrow bands of noise are present either singly or as a complex, and either one of the bands or all of the bands receive a temporal gap. The experiment tests the hypothesis that the improvement in performance with increasing number of marker bands carrying the gap is in accord with a wideband analysis where performance improves more steeply than predicted from the statistical combination of information from independent frequency channels.

Whereas most of the above-considered temporal resolution findings are compatible with a single channel, wideband analysis, the results of studies investigating the detection of complex signals are generally more compatible with accounts framed in terms of statistical combination of cues across *multiple*, independent frequency channels. For example, the improvement in threshold with increasing number of signal components has been successfully accounted for in terms of the combination of d' 's associated with multiple, independent, frequency channels. Green and Swets (1966) demonstrated that, for optimal combination of independent cues of equal sensitivity, d' increases as a function of the square root of the number of signal components. Because d' for tone detection in noise is approximately proportional to intensity, threshold is expected to decrease as a function of $10 \log \sqrt{n}$, with a doubling of signal components resulting in a threshold improvement of approximately 1.5 dB. Several investigations using multiple-component signals have shown that detection thresholds agree well with those predicted from optimal or near optimal combination of information from independent channels (e.g., Spiegel, 1979; Buus *et al.*, 1986). There are two important assumptions associated with the spectral integration paradigm. First, it is assumed that the psychometric functions associated with the single-component signal and the multiple-component signal are parallel; i.e., they have the same slope. The second assumption is that the individual components making up the signal are sufficiently removed in frequency from each other so as to be spectrally independent; i.e., the frequency resolution of the auditory periphery is assumed to isolate the individual components adequately. In contrast, a simple pooling of components within a single broad filter (wideband analysis) would predict a 3-dB change in threshold with a doubling of signal energy.

The improvement in threshold associated with an increase in the number of spectral components can be greater than 1.5 dB/doubling when the spectral components are very short in duration. van den Brink and Houtgast (1990) have shown that for signal durations of 10 ms or less, thresholds change by approximately $16 \log \sqrt{n}$ with increases in component number, which translates to approximately 2.4 dB/doubling. This value is greater than the 1.5 dB/doubling found for long-duration signals, but still less than the 3 dB/doubling that would occur with wideband analysis. Hicks and Buus (2000) tested the hypothesis that differences in threshold improvement between brief and long-duration signals could be accounted for by differences in the psychometric function slope. If psychometric functions were shallower for very brief than for longer

duration signals, then a fixed \sqrt{n} improvement in d' would be associated with a greater change in threshold. This is because the intensity difference corresponding to a fixed \sqrt{n} improvement in performance (d') will be larger in the case of a shallower function. Contrary to this hypothesis, Hicks and Buus found steeper psychometric function slopes for brief than for long-duration stimuli. Based on these results they argued that the most likely explanation for the duration effect is related to the close temporal synchrony across frequency produced by short-duration signals. Although the steepness of the psychometric functions did not account for the pattern of results in the Hicks and Buus study, this factor is indeed pertinent to the predicted magnitude of spectral integration, and, as seen in the following, has an important role in the present investigation.

Experiment 2 of the present study examines the improvement in detection resulting from an increase from one to four brief spectral components in the context of an overshoot paradigm. Overshoot refers to the finding that, under some conditions, threshold for a brief signal presented near the onset of a gated masker (*short-delay* condition) is elevated relative to thresholds for the same signal presented after a longer delay from masker onset (*long-delay* condition) (e.g., Elliott, 1965). This effect has been shown to depend on masker energy distant from the signal frequency (McFadden, 1989), particularly at low signal frequency (Carlyon and White, 1992). There have been several approaches proposed to account for these results. One suggestion is that thresholds in the *short-delay* condition are elevated due to a reduced ability to hone in on a single frequency at stimulus onset, such that listeners are initially monitoring a broad range of frequencies for the presence of a signal. This would have the effect of reducing the average signal-to-noise ratio. A process of auditory filter narrowing that begins at the stimulus onset, for example, could also produce this effect (Wright, 1995). In all of these studies, stimulus energy outside the auditory filter centered on the signal has been shown to play a role in the observed effects, possibly making the results from this paradigm amenable to explanation in terms of a wideband analysis. As in experiment 1, experiment 2 employs a paradigm using four widely spaced signal frequencies. In experiment 2, any one of the signals or all four signals were presented under conditions of both short-delay and long-delay relative to masker onset. The experiment tests the hypothesis that the improvement in detection threshold with increasing number of signals is consistent with statistical summation from independent frequency channels for the long-delay conditions, but is consistent with a wideband analysis for short-delay conditions.

As noted by Hicks and Buus (2000), it is important to take into account the steepness of the psychometric function when considering the improvement in threshold associated with an increase in the number of spectral components. For example, whereas a large improvement in threshold with increasing number of spectral components might be consistent with a wideband analysis, it could also be accounted for by spectral integration of components that were associated with relatively shallow psychometric functions. Relevant to this

consideration, Oxenham and Moore (1995) reported that the psychometric function slope fitted to pure-tone detection data varied across masker conditions similar to those used to estimate overshoot. In that study, psychometric function slopes for a gated masker, analogous to the *short-delay* condition, were shallower than those for a long-duration masker, analogous to the *long-delay* condition. A caveat in interpreting these data is that the masker was pulsed on and off with the signal in the gated condition, whereas the masker continues after the signal in typical overshoot conditions. The finding nevertheless highlights the importance of considering psychometric function slope when interpreting overshoot data. We therefore included a means of estimating the slope of the psychometric function in the present experiments. A possibility that will be considered in the following is that olivocochlear efferent effects (e.g., Schmidt and Zwicker, 1991; von Klitzing and Kohlrausch, 1994) may contribute to differences in psychometric function slopes in short- and long-delay conditions, and may therefore contribute to differences in spectral integration.

In summary, the experiments described here examined whether gap detection and overshoot are more compatible with interpretations based upon wideband analysis or upon the statistical combination of information from multiple, spectrally independent channels. Threshold estimates for the multiple-frequency target condition were predicted from the single-frequency target conditions using the weighted combination of d' model (Green and Swets, 1966). If thresholds for the multifrequency target conditions could be accounted for by the model predictions, then interpretation in terms of statistical combination of information from independent channels would be favored over an interpretation in terms of a wideband analysis. The spectral characteristics of the stimuli in both paradigms were comparable, with the target (either a temporal gap or a brief tone) present either at one frequency or at four frequencies that were widely spaced relative to equivalent rectangular bandwidths for a normal ear (ERB_N) (Glasberg and Moore, 1990). Therefore, the gap detection and overshoot paradigms can be thought of as roughly analogous tests of sensitivity to brief reductions or brief increases in energy.

II. EXPERIMENT 1: GAP DETECTION

A. Observers

Observers were 6 adults, ages 19–49 years (mean 29 years). All had pure-tone detection thresholds of 15 dB HL or better at octave frequencies 250–8000 Hz.

B. Stimuli

The stimuli were composed of four 50-Hz wide bands of Gaussian noise centered at 1000, 1932, 3569, and 6437 Hz. The bands were presented either singly, or as a four-band complex. When presented together as a complex, the bands either shared the same fluctuation pattern (comodulated) or had fluctuations patterns that were independent across band (random).

The noise bands were digitally generated prior to the experiment (Matlab, Mathworks) and stored to disk. Using a

sampling rate of 24.414 kHz and a 2^{18} -point buffer, 10.7-s segments of each wave form were created. Each band was generated in the frequency domain using a single set of Gaussian-distributed draws for the real and imaginary components of each band. The resulting noise had Rayleigh-distributed amplitudes and uniformly distributed phases. For comodulated bands, the same draws were used for each center frequency. Stimulus gating and presentation were implemented using a DSP platform (TDT RPDvs communicating with an RX6). For a given condition, the stipulated wave form was read from disk, set to output continuously, and gated on in each observation interval for 400 ms, including 20-ms \cos^2 rise/fall ramps. In the target interval, the gap was initiated 100–150 ms after stimulus onset, at random, and the gap was imposed with 20-ms \cos^2 ramps to reduce the availability of cues related to spectral splatter. Gap duration was specified as the interval between the time at which the marker level began to decrease (turn off) and the time at which the marker level started to increase again (turn back on). For gap durations less than 20 ms, the gap minimum at which the bounding ramps reversed from decreasing level to increasing level occurred at a nonzero voltage. The shapes of the ramps were symmetrical around this minimum. Gap durations that would have extended beyond the end of the observation interval were coded as hitting ceiling. Any track that reached ceiling two or more times was excluded from the results and replaced. Each of the four 50-Hz-wide noise bands was presented at a level of approximately 67 dB SPL, resulting in a total level of approximately 73 dB SPL. Stimuli were presented through the right phone of a Sennheiser 580HD headset.

C. Procedure and conditions

Thresholds was obtained with a three-alternative, forced-choice (3AFC) procedure incorporating a three-down one-up adaptive rule, which continued until a total of eight reversals were obtained. The initial gap duration was not fixed but was set to a value that was likely to be detected by the observer. The initial step size was a factor of $\sqrt{2}$, and this factor was reduced to its square root (c. 1.2) after the first two reversals. Following three correct responses in a row, gap duration was reduced by one factor; following one incorrect response, it was increased by one factor. Threshold estimates reported in the following are based on the geometric mean of the gap durations at the last six reversals. Observers made responses via a handheld response box, and visual feedback was provided. Both the gap duration and the observer response for each trial were saved to disk. Observers participated in 14 conditions: Conditions 1–4 consisted of the four bands of noise presented singly; conditions 5–8 consisted of the four comodulated bands presented together, but with the gap imposed in only one of the frequency bands; conditions 9–12 consisted of the four random bands presented together, but with the gap imposed in only one of the frequency bands; conditions 13 and 14 consisted of the four bands presented together, either comodulated or random, with the gap presented synchronously in all bands. Observers practiced on all conditions until thresholds reached asymptote.

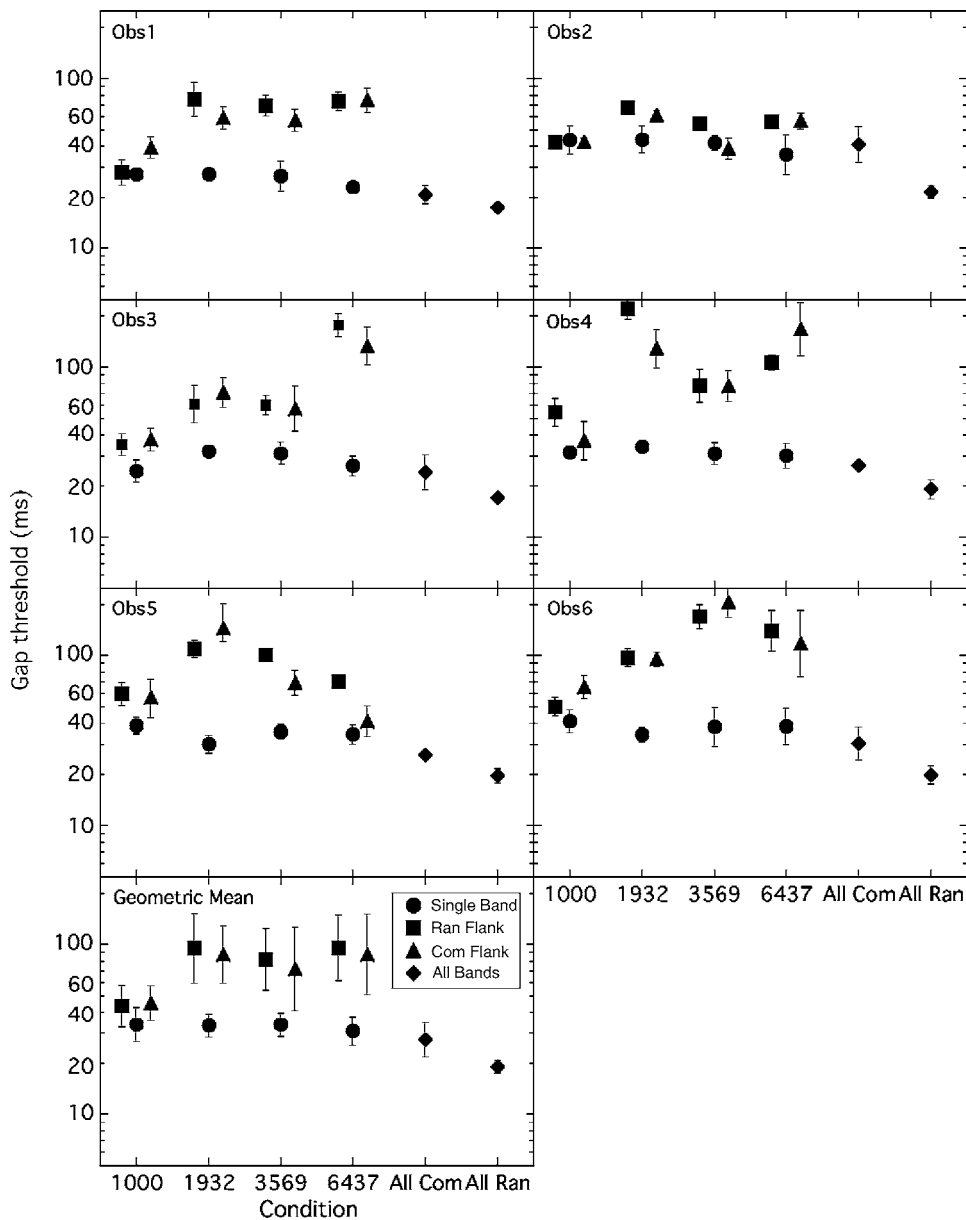


FIG. 1. Mean gap detection thresholds (ms) are plotted as a function of signal condition, with error bars indicating 1 s.d. computed in log units. The data from the conditions where the stimulus was a single band of noise are depicted by the circles. The data from the conditions where four noise bands were present but only one band received the gap are depicted by the triangles for the comodulated noise and by the squares for the random noise. The data from the condition where all four noise bands received a gap are depicted by the diamonds.

D. Psychometric function fits

Logit psychometric functions were fitted to the threshold track data for the single-band conditions using a maximum likelihood procedure (e.g., *Leek et al., 1992*). This function was defined as follows, where both μ (mean of the function) and k (a parameter determining slope) were free to vary,

$$\Phi(x) = \frac{1}{3} + \frac{2}{3} \left(\frac{1}{1 + e^{-k(x-\mu)}} \right). \quad (1)$$

Functions were fitted to the natural logarithm of the data, with sequential increments of 0.1 on μ and 0.1 on k . The modeling of *Leek et al. (1992)* indicated that the most reliable and accurate slope estimates are obtained from 3AFC and 4AFC (as opposed to 2AFC) procedures, using 100 or more trials. In the present 3AFC procedures, the median number of track trials per slope estimate was above this criterion (approximately 180 trials per estimate).

E. Results

Thresholds for gap detection conditions are shown in Fig. 1. The individual data are shown in the upper panels and the mean data are shown in the lower panel. In each panel, thresholds are plotted with reference to the frequency band carrying the gap or, if all bands carried the gap, with reference to whether the bands were comodulated or random. Although the data patterns appear quite variable across observers, several commonalities exist. For the single bands alone (conditions 1–4, circles), thresholds seem invariant with frequency, and thresholds appear lowest when the gap was carried synchronously in all four bands (conditions 13 and 14, diamonds). In the latter pair of conditions, thresholds were consistently lower when all bands were random than when they were comodulated. In general, thresholds were poorest when the gap was carried by one noise band in the presence of the other three noise bands, where the flanking bands were either comodulated (conditions 5–8, triangles) or

TABLE I. Slopes of the psychometric functions for conditions where a single gap marker was presented and the obtained and modeled gap detection thresholds (ms) when all four marker bands received a gap. Means are arithmetic for slope and geometric for gap.

Obs	Frequency (Hz)				Gap detection (ms)		
	1000	1932	3569	6437	Comod	Random	Modeled
1	8.0	8.0	4.1	6.0	20.8	17.4	19.3
2	1.7	4.2	2.2	1.4	40.9	21.6	22.5
3	4.0	3.4	4.7	4.6	24.0	17.1	19.9
4	3.8	8.0	5.8	4.8	26.4	19.2	17.2
5	3.8	4.3	6.2	3.8	26.0	19.6	23.6
6	2.5	4.2	2.4	3.0	30.4	19.9	24.4
Mean	4.0	5.3	4.2	3.9	27.4	19.1	21.0

random (conditions 9–12, squares). In these conditions, thresholds appeared to be best when the gap was carried by the lowest frequency marker (1000 Hz center frequency). Although the data pattern in the remainder of conditions 5–12 displays marked variability across observers, there appears to be a general consistency of pattern within an observer for comodulated versus random flanking bands.

In order to quantify the above-noted data trends, several repeated-measures analyses of variance (ANOVA) were undertaken. In all cases, the analyses were performed on the log transforms of the data to ensure homogeneity of variance. The first analysis was on the single bands alone (conditions 1–4), and indicated no effect of frequency ($F_{3,15}=1.34; p=0.30$). For each observer, therefore, thresholds for these conditions were collapsed and compared to thresholds with synchronous gaps in all bands (conditions 13 and 14). Here, the analysis indicated a significant effect of condition ($F_{2,10}=27.25; p<0.01$), and post hoc contrasts indicated that thresholds with multiple synchronous gaps were significantly better than for the single band alone, both for the comodulated bands ($F_{1,5}=20.50; p<0.01$) and the random bands ($F_{1,5}=62.20; p<0.01$). The latter two multigap conditions also differed significantly from each other ($F_{1,5}=11.68; p=0.02$), with a better gap detection threshold in the random condition. The next analysis examined the effect of modulation type (random or comodulated) and gap carrier frequency in those conditions where the gap was carried by one band in the presence of the other nongapped marker bands (conditions 5–12). The analysis indicated no effect of modulation type ($F_{1,5}=3.37; p=0.13$), a significant effect of gap carrier frequency ($F_{3,15}=6.13; p<0.01$), and no interaction between these factors ($F_{3,15}=0.34; p=0.80$). For each observer, the data were therefore collapsed across modulation type for each carrier frequency and compared to the respective single band alone threshold. This analysis indicated a significant effect of the presence of flanking bands ($F_{1,5}=39.90; p<0.01$), a significant effect of gap carrier frequency ($F_{3,15}=5.11; p=0.012$), and a significant interaction between these two factors ($F_{3,15}=6.49; p<0.01$). The interaction reflects the fact that, when only one of four marker bands received a gap, performance was best when the gapped marker was centered at 1000 Hz. Bonferroni-corrected paired comparisons between the thresholds with and without nongapped flanking

bands at each of the four frequencies all showed significantly poorer performance with the nongapped markers present (t_5 ranging from 4.06 to 5.46, $p\leq 0.01$).

For conditions in which a single marker band was present, psychometric functions fitted on a log(ms) scale produced parameter estimates of k that were highly variable across observer and condition (see Table I), but a repeated measures ANOVA indicated that k did not differ reliably as a function of frequency ($F_{3,15}=1.7; p=0.21$). The right-hand columns of Table I show the obtained and modeled gap detection thresholds for the case where all four markers received a gap. The modeled gap detection was determined for each observer based on optimal combination of d' . The fitted values of k and μ [see Eq. (1)] for each of the four conditions where a single marker received a gap were used to generate four psychometric functions for each observer. These functions were then converted into d' units and combined by taking the square root of the sum of squares associated with each of the four conditions. The combined d' was then converted back into percent correct and the 79% gap threshold for this composite function was determined. This value can be compared to the performance in the conditions where four markers were presented simultaneously and all four received a gap (see Table I). The gap detection threshold predicted in this way (21.0 ms) was significantly better than the threshold of 27.4 ms obtained in the comodulation condition ($t_5=3.2; p=0.02$). This prediction was not significantly different from the threshold of 19.2 ms obtained in the condition utilizing random bands ($t_5=1.9; p=0.10$).

F. Discussion

The present findings indicate close agreement between the gap detection modeled via the combination of d' associated with the single-marker conditions and the gap detection threshold obtained for the condition where four random bands were present and all received a gap. The random band results are therefore consistent with the combination of d' from independent channels and do not require a broad pre-detection filter. The gap detection threshold obtained in the condition where the four marker bands were comodulated was significantly higher than that modeled via the combination of d' . The relative independence of information across channels is pertinent to the interpretation of this result. In the

comodulated condition, although it is reasonable to assume that the internal noises associated with the frequency channels carrying marker information are independent across channels, the across-marker variability of the stimulus envelopes is nil. In a study assessing the combination of sequentially presented signal-plus-noise samples for cases where the external noise was the same across samples but internal noise was assumed to be independent, Swets *et al.* (1959) found an improvement with multiple samples, but less than the improvement that occurs when the external noise was also independent. In a similar vein, Hall *et al.* (1988) found that spectral integration for a multiple-tone signal was smaller for comodulated noise masking bands than for random noise bands. These past findings and the present finding of relatively poor performance in the comodulated case (compared to the random case and to the case where performance was modeled from the combination of d' from the single-band conditions) probably reflect the dependence of information across frequency in the comodulated case. Overall, the present data do not provide evidence that the gap detection threshold obtained when the gap is presented in all four markers is better than that expected from the combination of information across independent frequency channels.

The above-presented discussion suggests that it is not necessary to invoke a wideband predetection account for gap detection because the data appear to be compatible with a combination of information across independent frequency channels. Although this result does not rule out a wideband predetection filter account of gap detection, it should be pointed out that such an account does not appear to be compatible with the pattern of results in the conditions where only one of four bands received a gap. In these conditions, performance was best when the gap was applied to the lowest marker. This finding is important from the standpoint of evaluating the wideband analysis hypothesis. If all bands were subsumed within a broad predetection filter, the particular band receiving the gap would not be expected to make a difference. The present finding of a marker frequency effect when only one of four markers received a gap is therefore not consistent with an interpretation in terms of the wideband predetection filter.

A possible interpretation of the better performance when the lowest marker received the gap is related to upward spread of excitation. By such an account, the gap associated with a marker might be partially “filled in” by upward spread of excitation from a lower frequency marker. Because the 1000-Hz marker was the lowest marker, it would not be subject to such an effect. This upward spread of excitation hypothesis was evaluated using excitation pattern modeling (Moore *et al.*, 1997). In the first step of this analysis, excitation associated with each of the three higher-frequency marker bands (center frequencies of 1932, 3569, or 6437 Hz) was first determined. For each marker band, the excitation was calculated at the center frequency of that band. Then, for each of these center frequencies, the excitation arising from the marker band below it was determined. This analysis indicated that the excitation associated with a marker at its center frequency was at least 45 dB above the upward spread of excitation arising from a marker below that center fre-

quency. This analysis suggests that any “filling in” of a gap by a lower frequency marker would be negligible, and therefore the upward spread of excitation hypothesis was not supported.

Although aspects of the gap detection results obtained here do not favor an interpretation in terms of a wideband predetection filter, the findings nevertheless indicate that noise bands remote from a target band can reduce sensitivity to a temporal gap at a target frequency (see Fig. 1) when noise bands are present that do not contain a gap. One possible mechanism for this effect is that the presence of nongapped flanking bands mask some kind of off-frequency cue (perhaps spectral splatter) that is available in the case where the stimulus is a single-frequency gap marker. However, this is not likely to be a major factor as it has been shown that nongapped, continuous pure tones in spectrally remote regions have little or no deleterious effect on gap detection for a narrowband noise marker (Grose and Hall, 1993). It is also likely that energy splatter cues would be very weak in the present study, given the relatively long rise/fall of the stimulus gating.

It is possible that the deleterious effect of remote, nongapped markers may reflect a kind of across-channel interference, perhaps related to another psychoacoustical finding, modulation detection interference (e.g., Yost *et al.*, 1989). In modulation detection interference, the presence of amplitude modulation (AM) at one carrier frequency can interfere with the ability to detect AM at a remote carrier frequency. The deleterious effect of remote nongapped noise bands on gap detection for a target band may be interpreted in terms of a similar kind of across-channel interference. Note that by this account, the relatively good performance in the conditions where all four bands received a gap (compared to the conditions where only one of four bands received a gap) could be regarded as a release from across-channel interference. The best performance expected from a release from interference would be the gap detection threshold obtained in the absence of interferers (single marker cases). The thresholds obtained in the conditions where all bands received a gap were significantly lower than the thresholds obtained for a single marker for both comodulated and random masker types. Thus, although the better performance in the conditions where all bands received a gap relative to the performance in the conditions where a single band was present may arise in part from a release from across-channel interference, some other factor(s) (such as the combination of information across channels) must be invoked to account for the gap detection threshold in the condition where all four bands received a gap.

III. EXPERIMENT 2: OVERSHOOT

A. Observers

Observers were nine adults, with ages 16–28 years (mean 23 years). All had pure-tone detection thresholds of 15 dB HL or better at octave frequencies 250–8000 Hz.

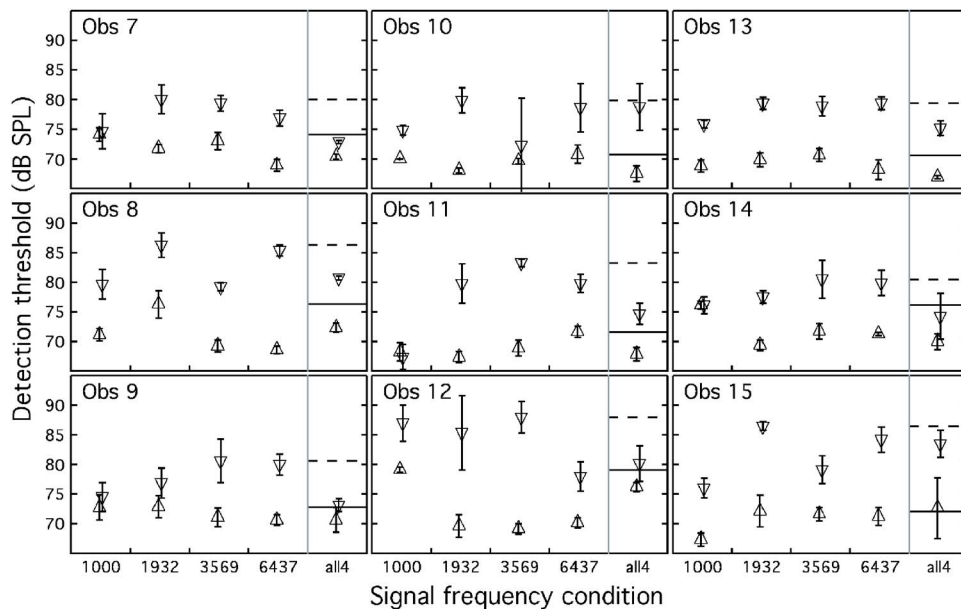


FIG. 2. (Color online) Mean detection thresholds (dB SPL) are plotted as a function of signal condition, with four single-frequency conditions and one multiple-frequency condition. Error bars indicate 1 s.d. around the mean. Each panel shows results for one observer. Symbols indicate the signal delay condition, *short-delay* (inverted triangle) and *long-delay* (triangle). The dotted line at the right of each panel is the maximum threshold obtained in the single-frequency *short-delay* conditions for that observer, which served as the reference level against which the multiple-frequency result is plotted. The solid line shows the analogous maximum *long-delay* threshold.

B. Stimuli

The masker for the overshoot conditions was a broadband Gaussian noise low passed at 8 kHz, having a duration of 409 ms, including 3-ms \cos^2 ramps. The masker presentation level was 40 dB/Hz SPL. The signal was composed of pure tones at the same four frequencies that were used as noise center frequencies in experiment 1 (1000, 1932, 3569, and 6437 Hz). The tones were presented either singly or as a four-tone complex. Tone duration was 8 ms, including 3-ms \cos^2 ramps, and the starting phase of each tone was chosen randomly prior to each presentation from a uniform distribution, 0 to 2π . The signal was delayed from masker onset by either 0 ms (*short delay*) or 250 ms (*long delay*). Stimuli were generated digitally (TDT, AP2), played out at a rate of 40 kHz, and presented monaurally via TDT modules to a Sony headphone (MDR-V6).

C. Procedure

Thresholds were obtained with a 3AFC procedure incorporating a three-down one-up adaptive rule, which continued until a total of six reversals were obtained. Threshold estimates reported in the following are based on the mean of the signal level at the last four reversals. Each threshold estimation track began with a level adjustment step of 4 dB, which was reduced to 2 dB after the second reversal. Thresholds for the single signal conditions were determined before the thresholds for the multiple-component signals were estimated. Initial threshold estimates for the conditions where the single-signal was presented were blocked by frequency, with three sequential runs collected. In some instances, observers showed clear practice effects across these runs. In these cases, the three sequential runs were not begun until the estimates appeared to reach asymptote. After completing all four target frequency blocks in random order, a fourth estimate was obtained for each frequency. If the fourth estimate was one or more standard deviations lower than the mean of the previous estimates, then the original estimates were discarded and those conditions were repeated. This pro-

cedure was used to reduce the possibility that observers were still learning detection cues by the end of testing. For the multiple-component signal, component levels were normalized to equal SL relative to the component having the highest single-frequency threshold. For example, if masked detection thresholds of 70, 70, 71, and 72 dB had been obtained for 1000, 1932, 3569, and 6437 Hz, respectively, then the levels for the four-component complex would be adjusted such that the 6437-Hz component was 1 dB higher than the 3569-Hz component and 2-dB higher than the components at 1932 and 1000 Hz. The threshold level of the complex was reported in terms of the highest-level component (in this example, the level of the 6437-Hz component). All thresholds in the *long-delay* conditions were estimated prior to those in the *short-delay* conditions.

D. Psychometric function fits

Logit psychometric functions described in Eq. (1) were fitted to the threshold track data using a maximum likelihood procedure. Functions were fitted to the data in units of decibels, with sequential increments of 0.5 on μ and 0.05 on k . As in the gap detection experiment, the median number of track trials per slope estimate was above the 100 trial criterion suggested from the modeling of Leek *et al.* (1992) (approximately 150 trials per estimate). Threshold estimates for the multiple-frequency condition were modeled from the psychometric functions fitted to the single-frequency conditions using the combination of d' (Green and Swets, 1966).

E. Results and discussion

Thresholds in the overshoot conditions are shown in Fig. 2, with each panel showing mean and standard deviation of threshold data for an individual subject. Because variability was relatively high across observers, mean data are not shown in Fig. 2. Overshoot, defined as the difference in thresholds in analogous *short-delay* (inverted triangle) and *long-delay* (triangle) conditions, varied across observers and across frequency for the single-signal conditions. The aver-

TABLE II. Slopes of the psychometric functions for conditions where one of four signal frequencies was present and the obtained and modeled spectral integration. Obtained spectral integration (dB) represents the improvement obtained in the condition where all four signals were presented in relation to the condition associated with the poorest threshold when only one signal was presented. Modeled spectral integration (dB) represents improvement in detection threshold estimated from the combination of d' from each of the single-signal conditions. Data are for the *long-delay* conditions.

Obs	Frequency (Hz)				Spectral integration (dB)	
	1000	1932	3569	6437	Obtained	Modeled
7	0.3	0.8	0.7	0.8	3.8	2.2
8	1.0	0.4	1.2	1.4	3.9	1.5
9	0.4	0.5	0.6	1.2	2.2	2.3
10	1.0	1.2	0.9	0.6	3.2	1.5
11	0.6	0.8	0.8	1.0	3.8	1.7
12	0.6	0.4	1.2	0.8	3.0	2.1
13	0.6	0.4	0.6	0.6	3.8	2.5
14	0.4	0.7	0.4	0.8	6.1	2.4
15	0.8	0.2	0.6	0.6	-0.5	2.9
Mean	0.6	0.6	0.8	0.9	3.2	2.1

age overshoot in the single-signal conditions across observers and across signal frequency was 8.6 dB (range: -0.9 to 16.8 dB). This magnitude and variability are similar to that which has been reported in other overshoot studies (e.g., Wright, 1995).

Because the levels of the component signals in the multiple-component signal conditions were adjusted relative to thresholds in the single-signal conditions for each observer, overall signal level in dB SPL is not a useful measure of threshold in that condition. Rather, threshold levels in the multiple-signal condition, plotted at the far right of each panel, are reported as the level of the most intense component in the shaped spectrum of the multiple-component signal complex. The points in this region of the graph show threshold in the multiple-component signal condition, while the horizontal lines indicate the level of the maximum amplitude component in the single-component signal conditions for the *short-delay* (horizontal dotted line) and the *long-delay* (solid line) conditions. The vertical distance between the line and the corresponding data point reflects the detection advantage conferred by inclusion of all four signals in the complex at equal sensation level. In the multiple-signal condition, overshoot was 6.5 dB on average (2.5–11.3 dB). Thresholds improved by an average of 5.7 dB relative to the maximum-amplitude component in the *short-delay* condition. In the *long-delay* condition, however, thresholds improved by an average of only 3.2 dB; for eight out of the nine observers, threshold improvement was less than the 4.8 dB value predicted based on van den Brink and Houtgast's (1990) results with short duration stimuli, more closely resembling the 3 dB predicted by a $10 \log(\sqrt{n})$ rule typical of longer duration stimuli. The difference in threshold advantage conferred by spectral combination between *short-* and *long-delay* conditions, however, was statistically significant (Wilcoxon signed ranks test, $p < 0.05$). Greater threshold improvement with increasing n in the *short-* than the *long-delay* conditions is consistent with the notion that processing bandwidth is initially broad and then sharpens as signal duration progresses (see Wright, 1995), but this interpretation is valid

only if the underlying psychometric functions are parallel across frequency and condition.

The effects of psychometric function slope were assessed via the Logit functions fitted to each observer's data in each of the single-component signal conditions. The parameter k , reflecting the slope of the best fitting psychometric function, differed significantly across the two delay conditions, with shallower functions associated with the *short-delay* than the *long-delay* conditions ($\bar{k}=0.38$ vs $\bar{k}=0.69$, respectively; $F_{1,8}=23.2$, $p < 0.001$). The slopes of the psychometric functions where one signal frequency was present are shown in Tables II and III (*long delay* and *short delay*, respectively), along with the obtained and modeled spectral integration. Modeled spectral integration was determined for each observer based on optimal combination of d' , using the average fitted psychometric function slopes for each of the four single-component signal conditions. The slope fits for each of the four conditions where a single signal was presented were used to generate four functions, with each function shifted such that the 79% correct point was located at 0 dB (to simulate equal detectability). These functions were then converted into d' and combined by taking the square root of the sum of squares associated with each of the four signals. The combined d' was then converted back into percent correct and the signal level associated with the 79% correct point for this composite function was determined. This value was subtracted from 0 to determine the modeled spectral integration.

The modeled spectral integration was on average 2.1 and 5.9 dB for *long-* and *short-delay* conditions (see Tables II and III). These values did not differ significantly from observed result for the *short-delay* condition ($t_8=0.2$; $p=0.87$) or the *long-delay* condition ($t_8=1.7$; $p=0.13$). Although the magnitude of the improvement in threshold from one to four signals observed for the *short-delay* condition is quite close to the 6-dB improvement expected from a wideband analysis, it would seem more parsimonious to explain these results in terms of differences in psychometric function slope and

TABLE III. Same information as in Table III, except that the data are for the *short-delay* conditions.

Obs	Frequency (Hz)				Spectral integration (dB)	
	1000	1932	3569	6437	Obtained	Modeled
7	0.2	0.2	0.2	0.6	7.2	4.6
8	0.2	0.4	0.8	0.7	5.6	3.0
9	0.4	0.2	0.2	0.2	7.5	5.9
10	0.4	0.2	0.1	0.1	1.1	9.7
11	0.3	0.2	0.2	0.3	8.6	5.6
12	0.1	0.1	0.1	0.1	7.8	13.9
13	0.7	0.8	0.6	0.4	4.2	2.4
14	0.2	0.8	0.2	0.4	6.2	4.0
15	0.4	0.4	0.2	0.4	3.0	4.1
Mean	0.3	0.3	0.3	0.4	5.7	5.9

near-optimal combination of information across frequency. The fact that the average psychometric function slope for the *short-delay* condition was substantially shallower than that estimated for the *long-delay* condition can account for the greater spectral integration obtained in that condition without relying on an additional process of pooling of auditory filter outputs into a wide predetection filter.

Several investigators have suggested that the olivocochlear efferent system may have a role in the overshoot effect (e.g., Schmidt and Zwicker, 1991; von Klitzing and Kohlrausch, 1994; Turner and Doherty, 1997; Strickland, 2001), and Zeng *et al.* (2000) have shown some reduction in overshoot in humans whose efferent connections have been severed surgically. It has been hypothesized that the cochlear input/output function becomes more linear (less compressive) following efferent activation (von Klitzing and Kohlrausch, 1994). Thus, the short-delay condition would be associated with a highly compressive region of the cochlear input/output function (e.g., Ruggero, 1992) where a large change in signal level is associated with a relatively small increase in cochlear output. For the long delay condition, however, efferent activation will have reduced cochlear amplification, resulting in a steeper input/output function. Compared to the short-delay condition, a given increase in signal level results in a greater output, a situation compatible with the overshoot phenomenon. The main relevance of this argument to the present data pertains to the slopes of the psychometric functions. The shallower cochlear input/output function hypothesized for the short-delay condition would be expected to be associated with a shallower psychometric function, which, in turn, would be associated with an amount of spectral integration greater than that for the long-delay condition.

IV. GENERAL DISCUSSION

Similar to previous results, gap detection was often quite poor in the condition where only one of the multiple narrow-band noise markers present received a gap. This was true even in the case where the multiple bands fluctuated coherently. There was little evidence that observers were able to use the potential gap detection cue of an across-frequency difference in pattern of fluctuation available in the comodulated conditions where only one of four bands received a

gap, although other work in our lab indicates that this can occur under special conditions. Gap detection thresholds in the multiple-marker gap conditions employing random bands were generally consistent with those expected based on the combination of d' . This result and the finding that sensitivity to the gap depended on the particular frequency receiving the gap in the case where flanking bands were present did not favor an obligatory wideband analysis. The deleterious effect of flanking noise bands on gap detection may be better accounted for in terms of a kind of across channel interference associated with the processing of temporal modulation.

The results from the overshoot paradigm also failed to produce strong support for the wideband listening hypothesis in the *short-delay* conditions. Although the threshold improvement associated with the increase from one to four signal components was larger in the *short-* than the *long-delay* conditions when expressed in decibels, this improvement was not in excess of that predicted based on the combination of d' . Thus, the patterns of data in the *short-* and *long-delay* conditions can both be accounted for via the same process of multiple-channel spectral integration. By this interpretation, threshold improvement was larger for the *short-delay* condition because of the shallower psychometric function slopes associated with that condition. This is consistent with the data of Oxenham and Moore (1995), which were discussed in terms of increased variance of the cue underlying signal detection at stimulus onset. This result is also consistent with an account based on efferent effects (Strickland, 2001), where the active mechanism responsible for amplification of low level sounds is hypothesized to adjust gain according to the level of the input signal, but with this adjustment occurring at some delay after an abrupt change in signal level (e.g., masker onset).

The present results indicated that the magnitude of overshoot, in decibels, was smaller for the four-component signal than for the one-component signal. This result is compatible with previous findings that have shown that overshoot is small or absent when a noise signal with a relatively wide bandwidth is used (e.g., Zwicker, 1965; Wright, 1995). For example, Zwicker (1965) found essentially no overshoot and Wright (1995) found overshoot of only approximately 1–4 dB when a wideband noise signal was employed. Although the present results are compatible with these past re-

sults, it should be noted that one factor complicating interpretation of the noise-signal data is the variability of the overshoot magnitude across signal frequency, a characteristic that is quite evident in the present results. Whereas the present experiment controlled for this factor by shaping the spectrum of the four-component signal, the previous studies employing noise signals did not address this factor. It is therefore not possible to make firm comparisons between the present study and the previous studies that used wideband noise signals.

One aspect of the overshoot results that deserves comment is that the spectral integration in the *long-delay* condition more closely resembled the $10 \log(\sqrt{n})$ improvement typically obtained for mid- to long-duration stimuli than the $16 \log(\sqrt{n})$ improvement that has been obtained previously for brief stimuli. Hicks and Buus (2000) showed that increased spectral integration for short duration signals is sensitive to the starting phase of the multiple-signal components, with the greatest effect for coherent phase at the amplitude peak of the Gaussian-windowed signal presentation. In experiment 2 signals were presented with random starting phase on each presentation and the steady portion of the signal presentation spanned 2 ms rather than being characterized by a single peak. Other parameters, such as a consistent signal duration across signal frequency in experiment 2 versus a consistent number of stimulus periods in Hicks and Buss, differed across studies and could have played a role in the reduced spectral integration observed in the *long-delay* condition. Also in contrast to the Hicks and Buus report, the approximately $16 \log(\sqrt{n})$ spectral integration obtained for brief signals was consistent with optimal combination of underlying psychometric functions, and so need not be interpreted in terms of processing unique to brief stimuli.

V. SUMMARY

The experiments described here examined the combination of signal information across frequency in gap detection and overshoot paradigms. In each case, sensitivity to brief targets at a localized frequency was compared to sensitivity to multiple targets distributed across frequency. In the case of gap detection for multiple, gated noise bands, spectral integration was generally consistent with the processing of gap information across frequency channels, a finding that did not support an obligatory pooling of information across frequency. Although the overshoot results were compatible with a wideband analysis, performance in the multiple-signal condition could more parsimoniously be accounted for in terms of spectral combination of independent cues.

ACKNOWLEDGMENTS

We thank Richard Freyman, Neal Viemeister, and three anonymous reviewers for comments on a previous version of this manuscript. This research was supported by NIH-NIDCD RO1 DC01507.

Buus, S., Schorer, E., Florentine, M., and Zwicker, E. (1986). "Decision rules in detection of simple and complex tones," *J. Acoust. Soc. Am.* **80**,

1646–1657.

- Carlyon, R., and White, L. (1992). "Effect of signal frequency and masker level on the frequency regions responsible for the overshoot effect," *J. Acoust. Soc. Am.* **91**, 1034–1344.
- Eddins, D. A., Hall, J. W., and Grose, J. H. (1992). "The detection of temporal gaps as a function of frequency region and absolute noise bandwidth," *J. Acoust. Soc. Am.* **91**, 1069–1077.
- Elliott, L. L. (1965). "Changes in the simultaneous masked threshold of brief tones," *J. Acoust. Soc. Am.* **38**, 738–746.
- Formby, C., Gerber, M. J., Sherlock, L. P., and Magder, L. S. (1998a). "Evidence for an across-frequency, between-channel process in asymptotic monaural temporal gap detection," *J. Acoust. Soc. Am.* **103**, 3554–3560.
- Formby, C., Sherlock, L. P., and Li, S. (1998b). "Temporal gap detection measured with multiple sinusoidal markers: Effects of marker number, frequency, and temporal position," *J. Acoust. Soc. Am.* **104**, 984–998.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Green, D. M., and Forrest, T. G. (1989). "Temporal gaps in noise and sinusoids," *J. Acoust. Soc. Am.* **86**, 961–970.
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Wiley, New York).
- Grose, J. H., and Hall, J. W. (1993). "Gap detection in a narrow band of noise in the presence of a flanking band of noise," *J. Acoust. Soc. Am.* **93**, 1645–1648.
- Grose, J. H., Hall, J. W., Buss, E., and Hatch, D. (2001). "Gap detection for similar and dissimilar gap markers," *J. Acoust. Soc. Am.* **109**, 1587–1595.
- Hall, J. W., Grose, J. H., and Haggard, M. P. (1988). "Comodulation masking release for multi-component signals," *J. Acoust. Soc. Am.* **83**, 677–686.
- Hicks, M. L., and Buus, S. (2000). "Efficient across-frequency integration: Evidence from psychometric functions," *J. Acoust. Soc. Am.* **107**, 3333–3342.
- Leek, M. R., Hanna, T. E., and Marshall, L. (1992). "Estimation of psychometric functions from adaptive tracking procedures," *Percept. Psychophys.* **51**, 247–256.
- McFadden, D. (1989). "Spectral differences in the ability of temporal gaps to reset the mechanisms underlying overshoot," *J. Acoust. Soc. Am.* **85**, 254–261.
- Moore, B. C. J., Glasberg, B. R., and Baer, T. (1997). "A model for the prediction of thresholds, loudness and partial loudness," *J. Audio Eng. Soc.* **45**, 224–240.
- Moore, B. C. J., Shailer, M. J., and Black, M. J. (1993). "Dichotic interference effects in gap detection," *J. Acoust. Soc. Am.* **93**, 2130–2133.
- Oxenham, A. J., and Moore, B. C. J. (1995). "Overshoot and the severe departure from Webers law," *J. Acoust. Soc. Am.* **97**, 2442–2453.
- Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (1997). "Detection of silent intervals between noises activating different perceptual channels: Some properties of 'central' auditory gap detection," *J. Acoust. Soc. Am.* **101**, 3694–3705.
- Ruggero, M. A. (1992). "Responses to sound of the basilar membrane of the mammalian cochlea," *Curr. Opin. Neurobiol.* **2**, 449–456.
- Schmidt, S., and Zwicker, E. (1991). "The effect of masker spectral asymmetry on overshoot in simultaneous masking," *J. Acoust. Soc. Am.* **89**, 1324–1330.
- Spiegel, M. F. (1979). "The range of spectral integration," *J. Acoust. Soc. Am.* **66**, 1356–1363.
- Strickland, E. A. (2001). "The relationship between frequency selectivity and overshoot," *J. Acoust. Soc. Am.* **109**, 2062–2073.
- Swets, J. A., Shipley, E. F., McKey, M. J., and Green, D. M. (1959). "Multiple observations of signals in noise," *J. Acoust. Soc. Am.* **31**, 514–521.
- Turner, C. W., and Doherty, K. A. (1997). "Temporal masking and the 'active' process in normal and hearing-impaired listeners," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ), pp. 386–396.
- van den Brink, W. A. C., and Houtgast, T. (1990). "Spectro-temporal integration in signal detection," *J. Acoust. Soc. Am.* **88**, 1703–1711.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.
- von Klitzing, R., and Kohlrausch, A. (1994). "Effect of masker level on overshoot in running- and frozen-noise maskers," *J. Acoust. Soc. Am.* **95**, 2192–2201.
- Wright, B. A. (1995). "Detectability of simultaneously masked signals as a function of signal bandwidth for different signal delays," *J. Acoust. Soc. Am.* **98**, 2493–2503.

Yost, W. A., Sheft, S., and Opie, J. (1989). "Modulation interference in detection and discrimination of amplitude modulation," *J. Acoust. Soc. Am.* **86**, 2138–2147.

Zeng, F. G., Martino, K. M., Linthicum, F. H., and Soli, S. D. (2000).

"Auditory perception in vestibular neurectomy subjects," *Hear. Res.* **142**, 102–112.

Zwicker, E. (1965). "Temporal effects in simultaneous masking by white-noise bursts," *J. Acoust. Soc. Am.* **37**, 653–663.

The Franssen effect illusion in budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*)

Micheal L. Dent,^{a)} Elizabeth M. McClaine, and Thomas E. Welch

Department of Psychology, University at Buffalo, the State University of New York, Buffalo, New York 14260

(Received 5 April 2007; revised 15 August 2007; accepted 18 September 2007)

The properties of the Franssen effect (FE) were measured in budgerigars and zebra finches. To elicit the FE, listeners are presented with a signal which has been split into a transient component, carrying an abrupt onset and ramped offset and separated in space from the sustained component which has a slowly rising onset and longer duration. When these two signals are played under certain conditions, the perception is that of a long-duration steady state tone being played at the location of the transient. The birds were trained using operant conditioning methods on a categorization task to peck a left key when presented with a stimulus from a left speaker and to peck a right key when presented with a stimulus from a right speaker. Once training was completed, FE stimuli were presented during a small proportion of trials. The FE was measured at speaker separations of 60° and 180° in both echoic and echoic-reduced conditions. Both species of birds exhibited the FE, although to varying degrees, across conditions. These results show that nonmammals also experience the FE illusion in confusing listening situations in a manner similar to mammals, suggestive of similar auditory processing mechanisms.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799493]

PACS number(s): 43.66.Gf, 43.66.Qp, 43.66.Pn, 43.80.Lb [JAS]

Pages: 3609–3614

I. INTRODUCTION

Studies of sound localization in animals have revealed large differences in abilities across species. Within birds alone, nocturnal predators such as marsh hawks (*Circus cyaneus*) and barn owls (*Tyto alba*) are able to localize to within 1°–2° of a sound source (Knudsen *et al.*, 1979; Rice, 1982), whereas smaller nonpredatory birds such as the bobwhite quail (*Colinus virginianus*) and zebra finch (*Poephila guttata*) show resolution closer to 100° in certain localization tasks (Gatehouse and Shelton, 1978; Park and Dooling, 1991). Budgerigars (*Melopsittacus undulatus*) are another species of bird shown to have fairly unremarkable simple sound localization abilities (resolution of 27° for broadband noise; Park and Dooling, 1991). Nelson and Suthers (2004) recently showed much better resolution in the eastern towhee (*Pipilo erythrophthalmus*) than has been previously seen in closely related birds, although birds in this study were allowed to move their heads freely during the experiment and they showed a lateral head turning bias while listening to sounds that may have aided in localizing the sound sources. Regardless of differences in methodology, these studies at least demonstrate that there is quite a bit of diversity between species of birds in their sound localization acuity in quiet listening conditions.

Several experiments measuring avian binaural hearing under more complex listening conditions have yielded some surprising differences between abilities in these tasks and the simple localization tasks. As mentioned previously, budgerigars are fairly poor at localizing pure tones in quiet environments. A study of spatial release from masking in budgeri-

gars, however, shows that thresholds for detecting a pure tone embedded in a broadband noise when both signals are emitting from the same location are about 11 dB higher than thresholds for detecting that same pure tone when the broadband noise is emitted from a different location in azimuth (Dent *et al.*, 1997). This improvement, or masking release, is an amount comparable to that found in humans tested under similar conditions (Saber *et al.*, 1991). Another series of experiments on the auditory illusion known as the precedence effect in budgerigars and zebra finches also show that these birds with poor abilities to locate a sound source show the same time courses for the phases of the precedence effect (summing localization, localization dominance, and echo thresholds) as those found in humans (Dent and Dooling, 2003; 2004). Here, the delays between a leading sound and an identical lagging sound (mimicking an echo) have an influence on where birds perceive the sound. If the delays are short, they perceive a phantom sound towards the midline. At longer delays of about 1 ms, the birds perceive only one auditory image located at the leading stimulus location. At even longer delays, both sounds are perceived at their respective locations. These experiments demonstrate that these birds experience the noisy binaural world with overlapping or competing sounds in a manner similar to humans and other animals, despite not having the localization acuity of these other animals.

The Franssen effect (FE) is an auditory illusion thought to be related to the precedence effect. To experience the illusion, listeners are played two sounds simultaneously from two locations surrounding the midline in azimuth. One of those sounds is a transient tone with an abrupt onset and slowly decaying offset. The other sound is a pure tone of the same frequency with a slowly ramping onset (equal in dura-

^{a)}Electronic mail: mdent@buffalo.edu

tion to the offset of the transient) and much longer total duration. When the illusion is operating, listeners will report hearing only one sound, with the location of that sound being at the location of the transient stimulus (Franssen, 1962). This single perceived sound lasts as long as the longer sustained sound lasts; it is merely perceived at a position where no sound is actually being presented. Studies in humans show that the illusion occurs only for narrowband sounds in reverberant environments (Hartmann and Rakerd, 1989; Yost *et al.*, 1997). If sounds are broadband or listening conditions are anechoic, listeners hear both sounds at their respective locations.

A study comparing cats and humans found that the illusion works in cats equally well in reverberant versus anechoic environments, but that the other characteristics of the illusion were similar (Dent *et al.*, 2004). As mentioned earlier, the FE illusion is thought to occur in conditions where localization is difficult. In humans, these are for pure tones between 1.5 and 3 kHz, where neither interaural time nor level difference cues can be maximally utilized for sound localization (Hartmann and Rakerd, 1989; Stevens and Newman, 1936; Yost *et al.*, 1997). Cats showed more evidence for a perceived illusion at higher frequency ranges than found in humans, but these frequencies correspond to frequencies where cats have difficulty utilizing interaural time and level differences: 4–6 kHz (Dent *et al.*, 2004). Thus, conditions that are less-than-optimal for localization seem to promote the FE.

FE and sound localization experiments published to date all support the idea that humans and other animals that can accurately locate a stimulus using many cues such as broad bandwidths, rapid onsets, and anechoic environments should not experience the FE illusion (Hartmann and Rakerd, 1989; Yost *et al.*, 1997). However, when onsets of stimuli are obscured or reverberations are introduced, listeners do not seem to notice that the transient stimulus has ended and that a new stimulus is now playing from another location, and the FE is robust. The experiments here were designed to measure the FE in two species of animals that are very poor at localizing even in the most optimal conditions: budgerigars and zebra finches. We hypothesized that, because of the poor sound localization acuity of these birds, they would experience a greater incidence of the FE than any other animal tested to date and that the frequencies where the FE is most robust would match frequencies where these animals cannot accurately localize pure tones in isolation. For both species of birds, these are frequencies ranging from 500 to 1000 Hz (Park and Dooling, 1991).

II. METHODS

A. Subjects

Four adult budgerigars (3 males and 1 female) and five adult zebra finches (2 males and 3 females) were used as subjects in these experiments. All of the birds were individually housed in a vivarium at the University at Buffalo and were kept on a day/night cycle corresponding to the season. The birds were either purchased from a local pet store or bred in the vivarium. They were kept at approximately 90%

of their free-feeding weight during the course of the experiment. All procedures were approved by the University at Buffalo, SUNY's Institutional Animal Care and Use Committee and complied with NIH guidelines for animal use.

B. Testing apparatus

The psychoacoustic experiments took place in one of four identical psychoacoustic testing setups. The setups consisted of a wire test cage ($61 \times 33 \times 36$ cm³) mounted in a sound-attenuated chamber (Industrial Acoustics Company, Small Animal Chamber) lined with sound-absorbent foam (10.2 cm Sonex, Ilbruck Co.). For the echoic conditions, all of the foam was covered with heavy-duty Plexiglas panels. The test cage consisted of a perch, an automatic food hopper (Med Associates Standard Pigeon Grain Hopper), and two vertical response keys extending downwards from the inside of the hopper in front of the bird. The response keys were two sensitive microswitches with 1 cm green (left key) or red (right key) buttons glued to the ends. The birds pecking the colored keys tripped the microswitch. A small 7 W light at the top of the test cage illuminated the chamber and served as the experimental house light. An additional 30 W bulb remained on in the chamber for the entire session. The behavior of the animals during test sessions was monitored at all times by an overhead web-camera (Logitech QuickCam Pro, Model 4000). Two speakers (Morel Acoustics, Model MDT-29) were hung at the level of the bird's head, 30.5 cm away from the bird during testing (one to the right and one to the left of midline). The experiments were controlled by a Dell microcomputer operating Tucker-Davis Technologies (TDT, Gainesville, FL) modules and SYKOFIZX software.

C. Stimuli and calibration

On each day of testing, the birds were presented with a wide range of different stimuli from two possible locations ($\pm 30^\circ$ azimuth in half the sessions and $\pm 90^\circ$ azimuth in the other sessions). Schematics of the stimuli are depicted in Fig. 1. The stimuli were either broadband noise bursts or pure tones ranging from 1 to 10 kHz. The single source (SS) stimuli had a total duration of 500 ms, which was comprised of an immediate onset, a 400 ms sustained portion, and a 100 ms linear ramp offset. The FE stimuli consisted of two signals presented simultaneously, one to the left location and one to the right location. The transient signal had an immediate onset, a 50 ms linear ramp offset, and a total duration of 50 ms. The sustained signal had a 50 ms linear ramp onset, a 350 ms sustained portion, and a 100 ms linear ramp offset, for a total duration of 500 ms. The stimuli were presented so that the maximum amplitudes of their envelopes were the same as the maximum amplitude of the corresponding sustained stimulus that produced a sensation level of 30 dB. The stimuli were roved by ± 5 dB across trials and were generated at a sample rate of 50 kHz.

Stimulus calibration was performed with a Larson-Davis sound level meter (Model 825) and 20 ft extension cable. For all measurements, a $\frac{1}{2}$ in. microphone was placed in the position normally occupied by the bird's head during testing. We used the method described by Yost *et al.* (1997) and Dent

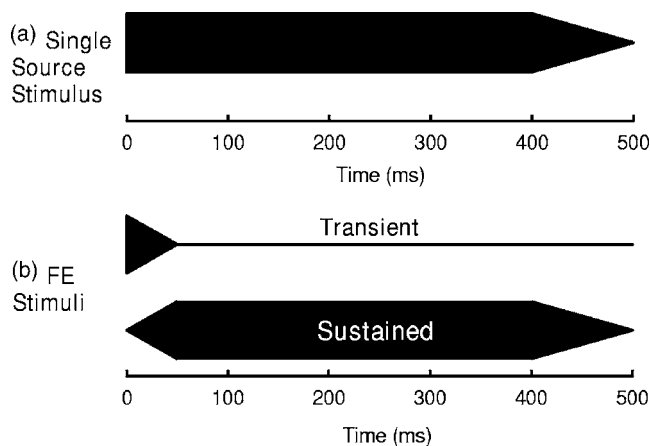


FIG. 1. Schematics of (a) SS and (b) FE stimuli. The SS stimulus was a single stimulus presented to either the right or left. This stimulus had an abrupt onset, a 100 ms offset ramp, and lasted 500 ms total. The FE stimuli were presented to two locations simultaneously (transient to the left and sustained to the right or vice versa). The transient stimulus had an abrupt onset, a 50 ms offset ramp, and lasted a total of 50 ms. The sustained stimulus had a 50 ms onset ramp, a 100 ms offset ramp, and lasted 500 ms total.

et al. (2004) to measure the reverberation characteristics of the room. The microphone was placed approximately where the bird's head was during testing. For each condition, we presented fifty 100 μ s clicks, and then averaged and band-pass filtered (500–12 000 Hz) the responses. We then compared the rms level of the averaged response for the first 5 ms of the recording to the rms level of the recording for the next 25 ms. We then took the $20 \log_{10}$ of the ratio of these two levels to compare the reflections to the initial click. For the echo-reduced chamber, the last 25 ms was down 19.9 dB from the initial 5 ms, and for the echoic condition the ratio was lowered by only 17.2 dB. Hence, there was a 2.7 dB difference in chamber conditions, which was lower than the 6.5 dB difference obtained in the *Dent et al.* (2004) study. So, although adding the plastic panels to the chambers changed the characteristics of the listening condition, these differences were somewhat small.

D. Procedures

The birds were trained using operant conditioning procedures to peck the microswitches for food reinforcement. First, they pecked at the left key to start a trial. After a variable interval of 2–7 s, the SS sound (as depicted in Fig. 1) was presented from the left or right location with equal probability. The birds were trained to peck the left key again when they localized the sound to the left location and peck the right key when they localized the sound to the right location. If they correctly identified the location of the sound within 1.5 s following its presentation, they were rewarded with 1.5 s access to hulled millet from the illuminated food hopper for 70% of the correct trials. They were rewarded with the hopper light only for 1.5 s in the other 30% of the correct trials. The birds were required to wait to respond until the entire stimulus was completed. If they responded either during the sound's presentation or later than 1.5 s after the sound's completion, no reward was given. If they re-

sponded incorrectly, either by indicating the wrong location of the stimulus or by responding too early, the house light was extinguished for 5 s.

The animals were randomly assigned to training with the SS stimuli on one of the four conditions: 180° echo reduced, 180° echoic, 60° echo reduced, and 60° echoic. Chance performance in these experiments is 50%. Once a training criterion performance of at least 85% correct across all frequencies was reached, testing began, involving the introduction of FE trials, which were interspersed with the SS stimuli on approximately 10% of all trials. As with the SS stimuli, the FE stimulus pairs were presented from both the left and right sides with equal probability (transient to the left and sustained to the right, and vice versa) and results were collapsed across those two positions. The reward contingency for the FE stimuli differed from our usual criteria; as these stimuli are illusory and have no “correct” response, the birds were rewarded for responding as long as their responses fell within the appropriate time window (from the conclusion of the stimulus presentation to 1.5 s later).

All four budgerigars were tested at the 180° and 60° speaker separations, but two of the zebra finches were unable to be trained on the 60° speaker separation so results from only three zebra finches are included for that experiment. Otherwise, all birds completed all four conditions in a random order and a different random order was used for each subject.

E. Data analysis

Each data point for each bird was based on 20 trials. SS errors were calculated for each subject at each frequency as the percent of trials where the birds responded incorrectly and these percent error values were then averaged across all subjects. Percent FE was calculated as the percent of trials where the birds responded that they perceived the stimulus to be located at the transient speaker's location. Two-way repeated measures ANOVAs (echo condition \times frequency) were conducted for each species and each speaker separation to determine if there were differences in the incidence of the FE.

III. RESULTS

Results for the 180° speaker separation configuration are shown in Fig. 2 for the budgerigars and in Fig. 3 for the zebra finches. Overall, the budgerigars were better than the zebra finches at localizing the SS stimuli [Figs. 2(a) and 3(a)]. The budgerigars showed the lowest errors for the broadband noise bursts as expected, and had mean errors ranging from 5% to 13% for the pure tones. The zebra finches were not better at localizing the noise bursts compared to the pure tones, and their mean errors ranged from 5% to 25%.

In the 180° condition, both species of birds exhibited a strong tendency to respond to the paired FE stimuli at the location of the 50 ms transient signal [Figs. 2(b) and 3(b)], despite the fact that they were required to wait until the end of the 500 ms sustained sound before making their responses. In budgerigars, this incidence of the FE was lowest

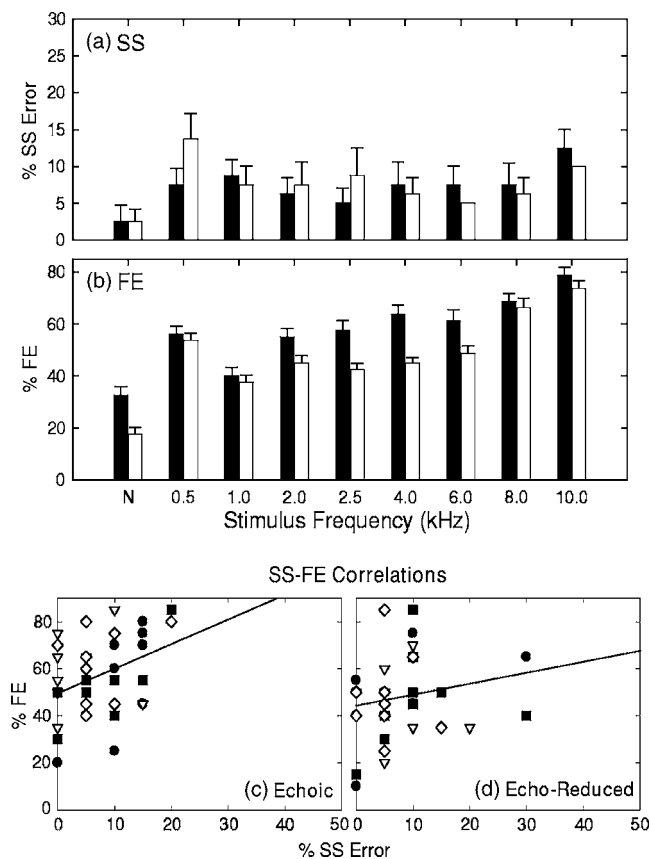


FIG. 2. Results for budgerigars with a 180° speaker separation distance. (a) The mean percent of trials where subjects incorrectly classified the SS stimuli as a function of stimulus frequency under echoic (black bars) and echo-reduced (white bars) conditions. Error bars represent between-subject standard errors. (b) The mean percent of trials where subjects indicated that they perceived the paired FE stimuli at the location of the transient as a function of stimulus frequency under echoic (black bars) and echo-reduced (white bars) conditions. Error bars represent between-subject standard errors. (c) Correlations between the percent of SS error and the percent of trials the subjects indicated that they perceived the paired FE stimuli at the location of the transient under echoic conditions. Each symbol type represents a different individual subject and each data point represents a different frequency. Regression lines for all data are also shown. (d) The same graph as in (c), except the results were collected under echo-reduced conditions.

for the broadband noise bursts and higher for the pure tones. The budgerigars showed significant effects of both frequency [$F(1,3)=10.70, p<0.05$] and echo condition [$F(8,24)=14.72, p<0.001$]. Holm-Sidak post-hoc t -tests comparing performance across frequencies found many significant pairs, including the noise bursts versus all pure tones except 1 kHz, and 10 kHz versus all other pure tones except 8 kHz ($p<0.05$). In zebra finches, there were no significant differences between the echo conditions [$F(1,4)=3.53, p>0.05$] but the differences across frequencies were significant [$F(8,32)=5.32, p<0.001$]. Post-hoc tests for the zebra finches revealed significant differences between the noise bursts and 6, 8, and 10 kHz, as well as between several other pairs ($p<0.05$).

To test the hypothesis that the FE is experienced more readily for stimuli that are inherently difficult to localize, correlations between the SS and FE conditions were calculated. Correlations between errors in localizing the SS stimuli and the incidence of the FE are shown in Figs. 2(c)

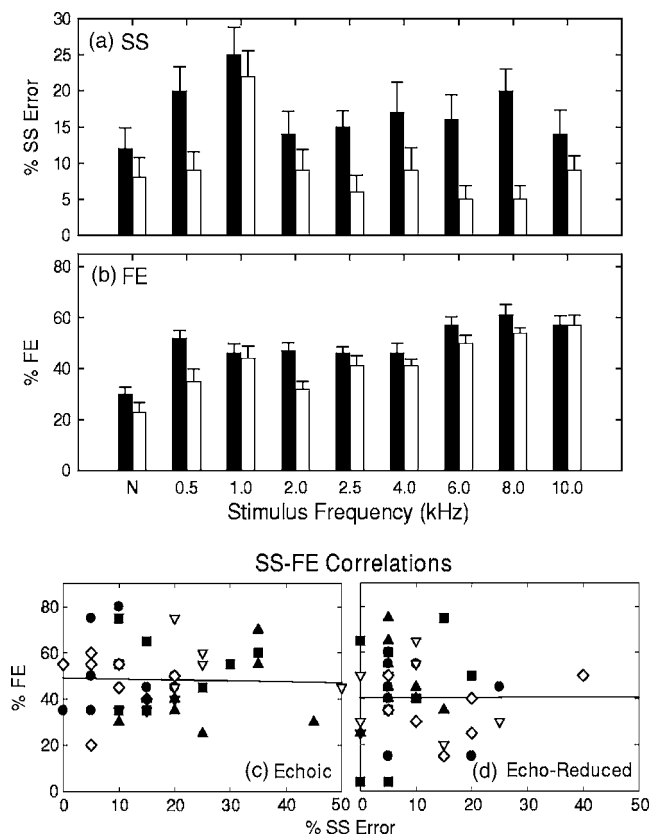


FIG. 3. Results for zebra finches with a 180° speaker separation distance. Plots are as in Fig. 2.

and 2(d) for the budgerigars and in Figs. 3(c) and 3(d) for the zebra finches. The budgerigars showed positive correlations under both echoic and echo-reduced conditions. Mean r -values for the budgerigars in the echoic condition were $+0.38$ ($p<0.05$) and for the echo-reduced condition were $+0.20$ ($p>0.05$). The zebra finches showed minimal correlations under both echoic and echo-reduced conditions. Mean r -values for the zebra finches in the echoic condition were -0.03 ($p>0.05$) and for the echo-reduced condition were $+0.002$ ($p>0.05$).

Results for the 60° speaker separation configuration are shown in Fig. 4 for the budgerigars and in Fig. 5 for the zebra finches. Unlike in the 180° speaker separation conditions, here the budgerigars and zebra finches were equally poor at localizing the SS stimuli [Figs. 4(a) and 5(a)]. Errors for localizing both the noise bursts and pure tones were much higher at this speaker separation distance (ranging from 22% to 53% overall).

In the 60° speaker separation condition, both species of birds again exhibited a strong tendency to respond to the paired FE stimuli at the location of the transient signal [Figs. 4(b) and 5(b)]. In budgerigars, the differences across frequency were not significant [$F(1,3)=0.54, p>0.05$], nor were there differences between echo conditions [$F(8,24)=2.11, p>0.05$]. In zebra finches, there were no significant differences between the echo conditions [$F(1,2)=9.14, p>0.05$] but the differences across frequencies were significant [$F(8,16)=2.63, p<0.05$].

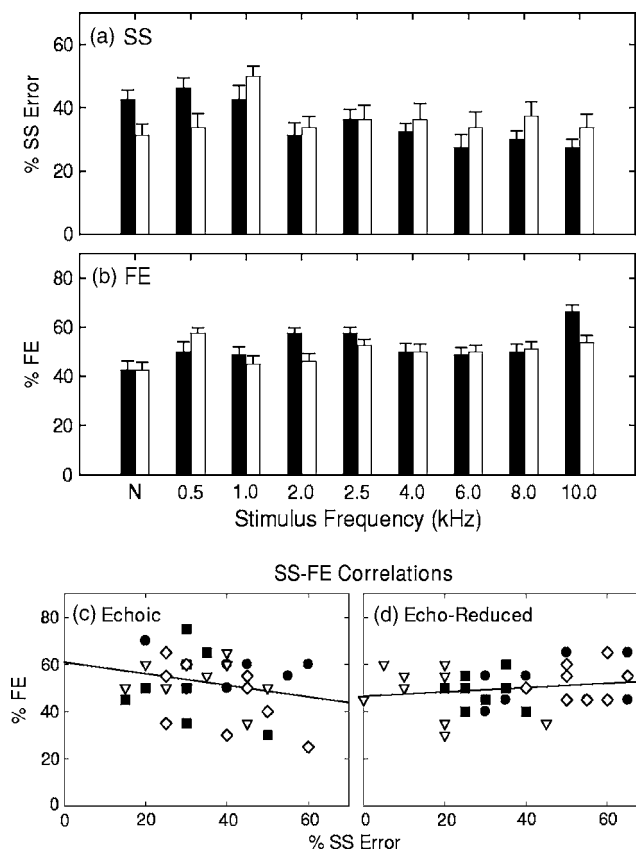


FIG. 4. Results for budgerigars with a 60° speaker separation distance. Plots are as in Fig. 2.

Correlations between errors in localizing the SS stimuli and the incidence of the FE in the 60° speaker separation condition are shown in Figs. 4(c) and 4(d) for the budgerigars and in Figs. 5(c) and 5(d) for the zebra finches. Both species showed positive correlations under the echo-reduced condition but not the echoic condition. Mean r -values for the budgerigars in the echoic condition were -0.25 ($p > 0.05$) and for the echo-reduced condition were $+0.18$ ($p > 0.05$). Mean r -values for the zebra finches in the echoic condition were -0.16 ($p > 0.05$) and for the echo-reduced condition were $+0.14$ ($p > 0.05$).

IV. DISCUSSION

The birds were trained on an operant conditioning task to categorize SS stimuli into categories of right or left. They were required to wait until the completion of the stimulus before making their response. After training was accomplished, each subject was presented with FE trials on a small portion of all trials, interspersed with the SS trials. Here, the responses were always rewarded as the animals might be experiencing an auditory illusion and we did not wish to bias their responses by rewarding only one type of response. As long as the animal waited until the end of both stimuli to respond, and responded within 1.5 s following the stimulus, they were rewarded. These methods were specifically chosen so that the sustained stimulus, which continued for 450 ms past the end of the transient sound, would be the most salient cue for the birds. If the birds were not experiencing the Franssen illusion, they would have responded to that more

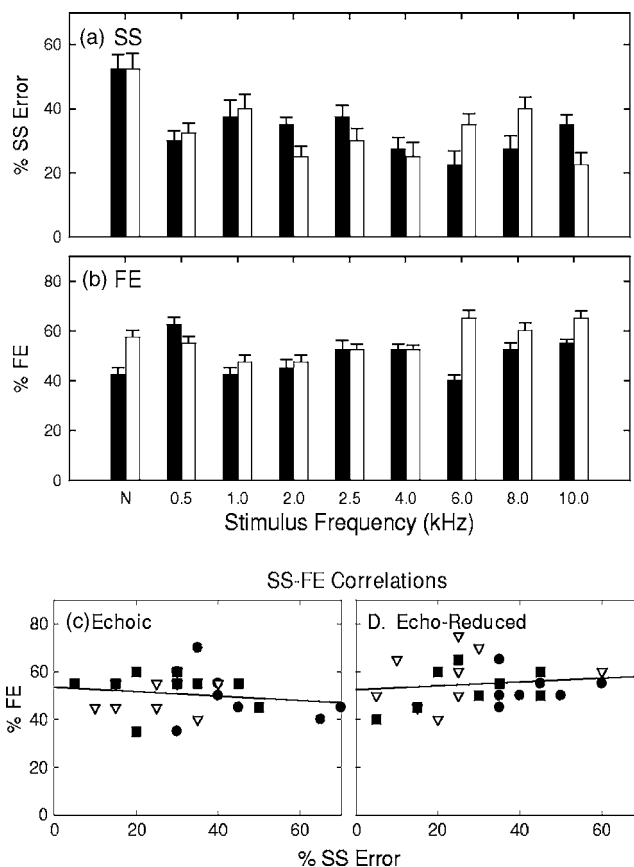


FIG. 5. Results for zebra finches with a 60° speaker separation distance. Plots are as in Fig. 2.

salient, less ambiguous, longer-lasting stimulus during FE trials. For many trials, however, this was not the case. The birds reliably chose the transient's location.

Surprisingly, the only “classic” Franssen illusion we found was experienced by the budgerigars in the 180° speaker separation condition. They showed a positive correlation between an inability to localize pure tones and the incidence of the FE across frequencies. Errors for localizing the SS broadband noise bursts were low, as was the incidence of the FE for noise. The budgerigars also showed a higher incidence of the FE under echoic than echo-reduced conditions at this speaker separation condition. These results mimic results from humans (Hartmann and Rakerd, 1989; Yost *et al.*, 1997).

For the budgerigars tested at 60° and for the zebra finches tested at both speaker separations, there was still evidence that the birds were experiencing the Franssen illusion, although the results are somewhat more difficult to interpret. The birds often indicated the perceived location of the FE stimuli pairs at the location of the transient long after the end of that stimulus, and this difference generally differed across frequency. This suggests that the birds are hearing a phantom sound at the location of the transient stimulus. However, the correlations between localizing SS stimuli and the incidence of the FE were not high, or even positive in some instances. Therefore, these results deviate from what is typically seen in FE experiments, although results from other animals tested on the Franssen illusion also diverge from human results.

Domestic cats showed little FE for broadband noise bursts compared to tones, similar to humans, although no differences were found between echoic and echo-reduced environments in this species (Dent *et al.*, 2004). Rhesus monkeys (*Macaca mulatta*) tested under echo-reduced conditions also showed high numbers of gaze shifts to a transient stimulus using high frequency pure tones, but less when broadband noise bursts were used (Dent *et al.*, 2005; Populin, 2006).

One of the requirements for the illusion discovered by Franssen is that the speakers cannot be too far apart. Franssen (1962) suggested this value to be no more than 40° in humans, whereas Hartmann and Rakerd (1989) measured it at 60°. In our experiment, we tried to account for the localization abilities of these birds by choosing much larger speaker separation distances. Park and Dooling (1991) found that budgerigars had absolute localization thresholds of between 45° and 90° for the pure tones tested here, and their thresholds were 27° for broadband noises. Zebra finches had even worse tone detection thresholds, ranging from 71° to immeasurable. Their absolute localization thresholds for broadband noises averaged 101°. In contrast, humans have pure tone localization thresholds of between 1° and 3° (Mills, 1958), while cats have pure tone thresholds of between 3° and 25° (Casseday and Neff, 1973; Martin and Webster, 1987). Thresholds for localizing broadband noises in cats are about 3°–5° (Casseday and Neff, 1973; Martin and Webster, 1987), and about 5° for humans (Stevens and Newman, 1936). The speaker separation distances where the FE is effective in humans (below 40°–60°) and cats (60°) represent separations *much* higher than the localization thresholds for simple pure tones (Dent *et al.*, 2005; Franssen, 1962).

As the budgerigars and zebra finches are so poor at sound localization, even larger speaker separation distances were chosen here. This was a relative, not absolute localization task, however, so we assumed that at least at 180°, our birds would be able to localize the SS stimuli fairly easily. If we compare our relative SS localization data to the absolute localization thresholds collected by Park and Dooling (1991), the budgerigars would be better-than-threshold (if we also use a 75% correct, or 25% incorrect, criterion) for all pure tones at 180° and worse-than-threshold at 60°. The zebra finches would have already been localizing below threshold for the 1 kHz pure tones and very close to threshold at several other tone frequencies at 180° and were attempting to localize well above threshold for all pure tones at 60°. Thus, it is not surprising that the results for the 60° condition in both species and for the 180° in zebra finches are not what has been typically seen in humans and cats. However, the results here that show that the birds will look towards the location of the transient at a high rate across conditions suggest that their general inability to localize stimuli in a complex listening environment contributes to their confusion when categorizing the paired FE stimuli.

Our findings add to earlier data in the same species of birds on sound localization abilities in complex, multidimensional listening environments. Although these birds were found to be poor at localizing sounds in quiet environments

(Park and Dooling, 1991), budgerigars exhibited a spatial release from masking for pure tones embedded in broadband noise in amounts comparable to those found in humans (Dent *et al.*, 1997). In a series of experiments on the precedence effect, a phenomenon thought to be related to the FE, budgerigars and zebra finches exhibited summing localization, localization dominance, and echo thresholds at time-courses extremely similar to those found in humans (Dent and Dooling, 2004). The results here showing that birds exhibit the FE, and that they exhibit it at very high rates across a number of frequencies, add to the wealth of comparative data on hearing in complex listening environments suggesting that the processing of auditory information is extremely similar across species.

ACKNOWLEDGMENTS

The authors would like to thank the numerous graduate and undergraduate students at UB for their invaluable experimental assistance and Dr. D.J. Tollin for assistance with the reverberation calculations. This work was supported by grants from the DRF and NOHR.

- Casseday, J. H., and Neff, W. D. (1973). "Localization of pure tones," *J. Acoust. Soc. Am.* **54**, 365–372.
- Dent, M. L., and Dooling, R. J. (2003). "Investigations of the precedence effect in budgerigars: Effects of stimulus type, intensity, duration, and location," *J. Acoust. Soc. Am.* **113**, 2146–2158.
- Dent, M. L., and Dooling, R. J. (2004). "The precedence effect in three species of birds," *J. Comp. Psychol.* **118**, 325–331.
- Dent, M. L., Larsen, O. N., and Dooling, R. J. (1997). "Free-field binaural unmasking in budgerigars (*Melopsittacus undulatus*)," *Behav. Neurosci.* **111**, 590–598.
- Dent, M. L., Tollin, D. J., and Yin, T. C. T. (2004). "Cats exhibit the Franssen effect illusion," *J. Acoust. Soc. Am.* **116**, 3070–3074.
- Dent, M. L., Weinstein, J. M., and Populin, L. C. (2005). "Localization of stimuli that produce the Franssen effect in mammals," Association for Research in Otolaryngology, New Orleans, LA.
- Franssen, N. V. (1962). *Stereophony* (Phillips Technical Library, Eindhoven, The Netherlands).
- Gatehouse, R. W., and Shelton, B. R. (1978). "Sound localization in the bobwhite quail (*Colinus virginianus*)," *Behav. Biol.* **22**, 533–540.
- Hartmann, W. M., and Rakerd, B. (1989). "Localization of sound in rooms IV. The Franssen effect," *J. Acoust. Soc. Am.* **86**, 1366–1373.
- Knudsen, E. I., Blasdel, G. G., and Konishi, M. (1979). "Sound localization in the barn owl (*Tyto alba*) measured with the search coil technique," *J. Comp. Physiol.* **133**, 1–11.
- Martin, R. L., and Webster, W. R. (1987). "The auditory spatial acuity of the domestic cat in the interaural horizontal and median vertical planes," *Hear. Res.* **30**, 239–252.
- Mills, A. W. (1958). "On the minimum audible angle," *J. Acoust. Soc. Am.* **30**, 237–246.
- Nelson, B. S., and Suthers, R. A. (2004). "Sound localization in a small passerine bird: Discrimination of azimuth as a function of head orientation and sound frequency," *J. Exp. Biol.* **207**, 4121–4133.
- Park, T. J., and Dooling, R. J. (1991). "Sound localization in small birds: Absolute localization in azimuth," *J. Comp. Psychol.* **105**, 125–133.
- Populin, L. C. (2006). "Monkey sound localization: Head-restrained versus head-unrestrained orienting," *J. Neurosci.* **26**, 9820–9832.
- Rice, W. R. (1982). "Acoustical location of prey by the marsh hawk: Adaptation to concealed prey," *Auk* **99**, 403–413.
- Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (1991). "Free-field release from masking," *J. Acoust. Soc. Am.* **90**, 1355–1370.
- Stevens, S. S., and Newman, E. B. (1936). "The localization of actual sources of sound," *Am. J. Psychol.* **48**, 297–306.
- Yost, W. A., Mapes-Riordan, D., and Guzman, S. J. (1997). "The relationship between localization and the Franssen Effect," *J. Acoust. Soc. Am.* **101**, 2994–2997.

Detection and discrimination of simple and complex sounds by hearing-impaired Belgian Waterslager canaries

Amanda M. Lauer^{a)} and Robert J. Dooling

Department of Psychology, University of Maryland, College Park, Maryland 20742

Marjorie R. Leek

National Center for Rehabilitative Auditory Research, Portland VA Medical Center, 3710 SW

US Veterans Hospital Road, Portland, Oregon 97207

Kirsten Poling^{b)}

Department of Psychology, University of Maryland, College Park, Maryland 20742

(Received 31 January 2007; revised 17 September 2007; accepted 18 September 2007)

Belgian Waterslager canaries (BWC) are bred to produce a distinctive low-pitched song with energy restricted to a lower range of frequencies than in other types of canaries. Previous studies have shown a high frequency hearing loss primarily above 2000 Hz that is related to hair cell abnormalities in BWC, but little is known about auditory perception in these birds. Here, frequency, duration, and intensity discrimination, temporal integration, gap detection, and discrimination of temporally reversed harmonic complexes in BWC were measured and compared to normal-hearing non-BWC. BWC had excellent frequency discrimination ability at 1000 Hz, but showed poor frequency discrimination compared to non-BWC at frequencies in the region of hearing loss. Duration and intensity discrimination were not adversely affected in BWC. Temporal integration was reduced in BWC, except at 2000 Hz. Gap detection and discrimination of temporally reversed stimuli were somewhat better in BWC than in non-BWC. Those tests that relied primarily on temporal processing were less affected by the cochlear damage in BWC than tests that probably relied more on audibility and spectral analysis. Thus, despite significant high frequency hearing loss and extensive damage along the basilar papilla, BWC retain relatively good hearing abilities under many conditions. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2799482]

PACS number(s): 43.66.Gf, 43.66.Sr, 43.80.Lb, 43.66.Fe [JAS]

Pages: 3615–3627

I. INTRODUCTION

The canary (*Serinus canaria*) is a species of cardueline finch that produces intricate songs during a yearly breeding season. Domesticated canary song is noted for its long sequences of “tours” consisting of consecutive repetitions of relatively tonal notes or syllables (e.g., [Güttinger, 1985](#)). Several strains of canaries are bred for particular song characteristics, while other strains are bred for body shape or plumage. The Belgian Waterslager canary (BWC) is one of the types bred for song. BWC song contains distinct syllables referred to by breeders as “water notes.” The frequency range of BWC song lies mainly between 1000 to 4000 Hz ([Nottebohm and Nottebohm, 1978](#); [Güttinger, 1985](#); [Wright et al., 2004](#)). The song repertoire consists of approximately 20–35 different syllable types ([Marler and Waser, 1977](#); [Nottebohm and Nottebohm, 1978](#); [Güttinger, 1985](#)). In contrast, songs produced by non-BWC strains typically have broader frequency ranges (1000–6500 Hz) and more syllable types ([Güttinger, 1985](#); [Lohr et al., 2004](#)).

BWC have been the subject of many behavioral and neurobiological studies of song learning and production. Male BWC normally learn their songs from their fathers and male siblings, though they can learn from other sources of acoustic input ([Marler and Waser, 1977](#); [Waser and Marler, 1977](#)). Recent work has shown that song learning by juvenile BWC is largely influenced by tutor song, and that young birds are capable of imitating synthesized song with a structure that does not resemble normal adult song, highlighting the substantial influence of acoustic input during song learning ([Gardner et al., 2005](#)).

BWC have been found to have a hereditary hearing loss, primarily above 2000 Hz, linked to hair cell abnormalities ([Gleich et al., 1994, 1995](#); [Okanoya and Dooling, 1985, 1987](#); [Okanoya et al., 1990](#); [Wright et al., 2004](#)), which is presumably due to breeding for special song characteristics. Many hair cells are missing or damaged, and many of the remaining hair cells have abnormal stereocilia bundles ([Gleich et al., 1994, 1995](#); [Weisleder and Park, 1994](#); [Weisleder et al., 1996](#)). Despite a reduction in the number of hair cells and auditory nerve fibers in BWC ([Gleich et al., 2001](#)), cell number is not significantly reduced, compared to non-BWC, in auditory brainstem structures responsible for encoding temporal information (nucleus magnocellularis and nucleus laminaris; [Kubke et al., 2002](#)). However, the overall volumes of nucleus magnocellularis and nucleus laminaris are reduced in BWC, an effect that has been attributed to smaller

^{a)}Currently at Johns Hopkins University, 521 Traylor Research Building, 720 Rutland Avenue, Baltimore, MD 21205. Electronic mail: alauer2@jhmi.edu

^{b)}Currently at Department of Biological Sciences, University of Windsor, 401 Sunset Avenue, Windsor, Ontario, Canada N9B 3P4.

cell size. Intriguingly, the hearing deficit and hair cell abnormalities develop after hatch, but before the bird reaches adulthood (Brittan-Powell *et al.*, 2002; Ryals and Dooling, 2002). Thus, the peripheral auditory system of the BWC is compromised during the period of initial song learning, as well as during yearly periods of song modification in adults.

Auditory information is crucial for the development and maintenance of species-specific vocalizations in canaries. Canaries raised in isolation or masking noise develop relatively simple songs compared to birds raised in the presence of tutors (Marler and Waser, 1977). Surgical deafening during adulthood can result in the degradation of song in canaries (Nottebohm and Nottebohm, 1976). The ability of a bird to resolve acoustic changes in tutor song must directly influence the content of its vocalizations. The animal needs to distinguish one call or note from another in order to develop and maintain a normal vocal repertoire, and it must be able to focus its attention on important acoustic information to effectively communicate with other members of its strain or species. More specifically, the bird must be able to resolve changes in frequency, time, and intensity. It must also be able to combine complex interactions of the acoustic dimensions to perceive and respond to naturally occurring sounds.

The behavioral consequences of the pathology in BWC, other than elevated high frequency thresholds, are not well understood because few studies have investigated auditory perception in BWC. Separate studies have shown that frequency selectivity and phase effects on masking are reduced in BWC (Lauer and Dooling, 2002; Lauer *et al.*, 2002; 2006). These studies suggest that there are deficits in the active processing mechanisms of the basilar papilla that result in abnormal encoding in the auditory periphery of BWC. The effects of the hair cell pathology on other aspects of auditory perception are unknown.

The unique vocalizations coupled with abnormal inner ear pathology in BWC may be related to strain-specific perceptual predispositions. These perceptual predispositions are likely to influence song learning and preference and within-strain communication. The following series of experiments was chosen to encompass some of the types of acoustic cues that are present in canary vocalizations. We investigated discrimination of changes in the frequency, duration, and intensity of pure tones, temporal integration of pure tones, gap detection, and discrimination of temporally reversed harmonic complexes in BWC and non-BWC. These studies are part of a larger effort to assess auditory perception in BWC. These birds are the only known animal in which we can investigate the relationships among heredity, auditory system structural abnormalities, vocal learning and vocal production, hair cell regeneration, and hearing abilities. Together, these experiments provide perhaps the most comprehensive investigation of hearing abilities in an animal with early-onset hereditary hearing loss.

II. GENERAL METHODS

A. Subjects

Adult BWC and adult non-BWC were used in each experiment. The same birds did not participate in all experi-

ments due to deaths. A total of 8 BWC and 7 non-BWC were used. Birds were housed in an avian vivarium at the University of Maryland and kept on a 12/12 h photoperiod. All birds were maintained at approximately 85–90% of their free-feeding weight, and had free access to water and grit. The Animal Care and Use Committee of the University of Maryland (College Park, MD) approved the care and use of animals in this study (A3270).

B. Apparatus

Birds were tested in a wire cage (26 × 18 × 14 cm) mounted in a sound-attenuated chamber (Industrial Acoustics Company, Bronx, NY, IAC-3) lined with acoustic foam. The test cage consisted of a perch, an automatic feeder (food hopper), and two response keys made of red and green 8 mm light-emitting diodes (LEDs) attached to two microswitches. The left key (red LED) was designated as the observation key, and the right key (green LED) was designated as the report key. A speaker (KEF Model 80C, England) was mounted from the roof of the sound-attenuated chamber at a 45° angle aimed toward the front of the bird, approximately 35 cm from the bird's head. Birds were monitored at all times by an overhead video camera system during testing.

The experiments were controlled by an IBM Pentium III microcomputer operating Tucker-Davis Technologies (TDT, Gainesville, FL) System 2 modules. Stimuli were generated with a 40,000 Hz sampling rate prior to the beginning of the experiment, stored digitally, and output via a timing generator (TDT, Model TG6) to a four-channel D/A converter (TDT, Model DA3-4). Each signal was then output from a separate channel of the D/A converter to a digital attenuator (TDT, Model PA4) and amplifier (TDT, Model HB6) to the speaker. Stimulus calibration was performed periodically using a Larson–Davis sound level meter (Model 825, Provo, UT) attached to a $\frac{1}{2}$ in. microphone positioned in the place normally occupied by the birds' head during testing.

C. Training and testing procedures

Birds were trained to peck the observation key for a random interval of 2–6 s during a repeating background sound or in quiet. The background sound was alternated with a target sound twice after this random interval. The bird was required to peck the report key within 2 s of this target/background alternation to receive a food reward. A report key peck during this time was recorded as a hit. If the bird failed to peck the report key within 2 s of the target/background alternation, it was recorded as a miss. Incorrect report key pecks were punished with a time-out period during which the chamber lights and LEDs were extinguished. Time-out periods lasted from 1 to 10 s depending on an individual bird's performance. On 30% of all trials, sham trials were presented during which there was no target/background alternation. Pecks to the report key during sham trials were recorded as false alarms and punished with time-out periods. This procedure has been described in more detail elsewhere (Dooling and Okanoya 1995).

Experimental sessions consisted of approximately 50–100 trials, and birds were tested twice a day, 5 days a week.

Within a block of 10 trials, the bird was presented with 7 target sounds and 3 sham trials in a random order. All test sessions were automated using a custom-designed Visual Basic computer program. Data were stored digitally and analyzed using commercially available statistics software and a custom-designed analysis program.

The general procedures for measuring detection and discrimination thresholds in Experiments 1–5 were similar. The order of conditions tested was randomly chosen for each individual bird. Birds were required to peck the observation key during silence (detection) or a repeating background (discrimination). After a random interval, the background was alternated with a target sound. Target sounds were presented using the Method of Constant Stimuli. Thresholds were defined as the frequency of the target detected 50% of the time (P_c), corrected for the false alarm (FA) rate [$P_c^* = (P_c - FA) / (1 - FA)$] (Gescheider, 1985; Dooling and Okanoya, 1995). In Experiment 6, all target sounds were identical during a testing session. No thresholds were measured in Experiment 6. Instead, percent correct discriminations were measured.

Each bird ran a minimum of 300 trials on each experimental condition, and the last 200 trials once behavior stabilized were used for analysis. Behavior was considered stable if the threshold did not change by more than 1/3 of the increment step size within the last two 100-trial blocks.

Prior to testing in the experiments described in the following, absolute thresholds for a range of pure tone frequencies were measured for each bird to confirm normal hearing in non-BWC and hearing impairment in BWC. Average absolute thresholds and thresholds from individual birds are shown in Fig. 1. These thresholds are consistent with previously reported behavioral thresholds in BWC and non-BWC (Okanoya and Dooling, 1985, 1987; Okanoya *et al.*, 1990).

III. EXPERIMENT 1—FREQUENCY DISCRIMINATION

Canary vocalizations consist of mainly tonal elements (e.g., Güttinger, 1985). Despite the importance of frequency as a salient feature of avian vocalizations, most tests reveal that birds are not especially sensitive to frequency changes compared to other vertebrates. In general, birds are able to detect less than a 1% change in frequency between 1000 and 4000 Hz, whereas humans can detect less than a 0.5% change (reviewed in Dooling *et al.*, 2000). In this experiment, frequency difference limens (FDLs) at 1000, 2000, and 4000 Hz were measured in BWC and non-BWC at a range of sound levels. Budgerigars with mild residual hearing losses 4–6 weeks following kanamycin exposure do not show significant increases in FDLs for 1000 and 2860 Hz tones presented at 65 dB sound pressure level (SPL) (Dooling *et al.*, 2006). Accordingly, BWC may only show increased FDLs at frequencies where the hearing loss is most severe (above 2000 Hz).

A. Methods: Stimuli and procedures

Background stimuli were 1000, 2000, and 4000 Hz pure tones. Target stimuli were pure tones with frequencies ranging from 10 to 700 Hz above the background frequency,

with a step size of 10, 20, 50, or 100 Hz depending on the background frequency and the bird's estimated threshold. All stimuli were 400 ms in duration with rise/fall times of 20 ms. Birds were tested at a range of sound levels at each frequency. Stimuli were presented at 50, 60, 70, and 80 dB SPL for the 1000 Hz background condition; 40, 50, 60, 70, and 80 dB SPL for the 2000 Hz background condition; and 30, 40, 50, 60, 70, and 80 dB SPL for the 4000 Hz background condition. It was not possible to test BWC at all of the levels that non-BWC were tested on due to their high absolute thresholds. To reduce potential loudness cues, the sounds were randomly roved by ± 6 dB on each stimulus presentation.

Thresholds for detecting increments in tone frequency were measured in 4 non-BWC and 4 BWC. The average false alarm rate was 2.8% for non-BWC and 3.6% for BWC. Data from sessions with false alarm rates larger than 18% or with a hit rate less than 80% for the two targets with the largest frequency change were excluded from analysis. Four percent of the data from non-BWC were discarded, and 5% of the data from BWC were discarded.

B. Results and discussion

Mean FDLs expressed as percent of the background frequency are shown in Fig. 2 for BWC and non-BWC for (A) 1000, (B) 2000, and (C) 4000 Hz. Data points for individual BWC are shown where no average data could be computed because it was not possible to measure a FDL at that point for all subjects due to stimulus inaudibility. As reported in other species of small birds (Dooling and Saunders, 1975; Kuhn and Saunders, 1980; Sinnott *et al.*, 1980), non-BWC were able to detect a change in frequency as small as 1–2% at high sound levels. Non-BWC showed a decrease of about three percentage points in FDLs with increasing level at 1000 Hz, and less than one percentage point at 2000 and 4000 Hz. The largest FDLs were at 1000 Hz for non-BWC. It should be noted that of the three frequencies tested, non-BWC showed the highest absolute thresholds at 1000 Hz (see Fig. 1).

BWC showed larger FDLs than non-BWC at 2000 and 4000 Hz. The largest FDLs occurred at 4000 Hz, where absolute thresholds were most elevated. At 1000 Hz, BWC showed a trend toward slightly better FDLs at the two lowest levels tested. A one-tailed t-test revealed that BWC had significantly lower FDLs for 1000 Hz presented at 60 dB SPL [$t(6) = -2.687$, $p < 0.05$]. BWC showed very little change in FDLs with increasing level at 1000 and 2000 Hz. The two BWC that were tested at multiple levels at 4000 Hz showed a decrease in FDL with increasing sound level. Because BWC were not tested at all of the levels at which non-BWC were tested, it was not possible to perform an analysis of variance (ANOVA) on the entire set of data. Thus, a mixed factor ANOVA (strain \times frequency) was conducted on the FDLs for the 80 dB SPL condition only. There was no significant effect of frequency; however, there was a significant effect of strain [$F(1, 6) = 30.484$, $p = 0.001$]. The interaction

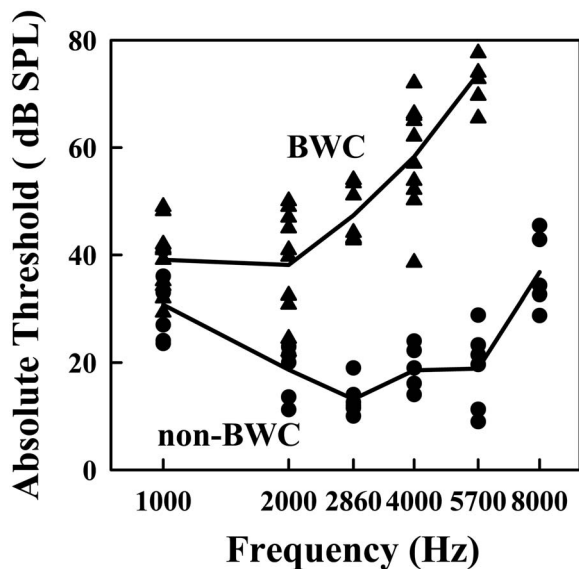


FIG. 1. Pure tone thresholds for average (solid lines) and individual non-BWC (circles) and BWC (triangles).

between factors was not significant. Strain differences in FDLs (BWC FDL—non-BWC FDL) at 80 dB SPL averaged across frequency are shown in Fig. 2(d).

The increased FDLs in the region of hearing loss in BWC compared to non-BWC are consistent with reports of reduced frequency discrimination abilities in animals with combined outer and inner hair cell damage. In cats and chinchillas, complete destruction of outer hair cells (OHCs) in the region of the basilar membrane corresponding to the test frequency does not result in increased FDLs; however, damage that results in the destruction of over 50% of inner hair cells (IHCs) and complete destruction of OHCs does result in

increased FDLs (Nienhuys and Clark, 1978; Prosen *et al.*, 1989). BWC show damage primarily to efferently innervated hair cells, but also show abnormal afferent hair cells. Thus, the decreased sensitivity to changes in frequency at 2000 and 4000 Hz is not surprising. However, the fact that frequency discrimination is better in BWC compared to non-BWC at 1000 Hz at 60 dB SPL despite the presence of significant hair cell abnormalities across the entire basilar papilla suggests that (1) there is not enough damage to hair cells with characteristic frequencies near 1000 Hz to impair frequency discrimination or (2) frequency discrimination at high and low frequencies is accomplished through different mechanisms. This hypothesis has been suggested by Sek and Moore (1995) to explain frequency discrimination results from human listeners. Temporal mechanisms of frequency discrimination may be dominant in lower frequencies, whereas, as phase locking is reduced at higher frequencies, spectral mechanisms have more importance. BWC may have particularly good temporal processing, but poorer than normal spectral processing due to their cochlear damage, providing at least a partial explanation for good frequency discrimination at 1000 Hz, but relatively poor performance, compared to normal canaries, at 2000 and 4000 Hz. Little effect of frequency might be observed in normal canaries, where spectral coding mechanisms may aid temporal coding mechanisms in the frequency region where phase locking declines.

The relationship between absolute threshold and frequency discrimination ability in BWC and non-BWC also indicates that spectral mechanisms are in play for higher frequencies and temporal mechanisms support frequency discrimination at lower frequencies. FDLs at 80 dB SPL are significantly correlated with absolute threshold in BWC and

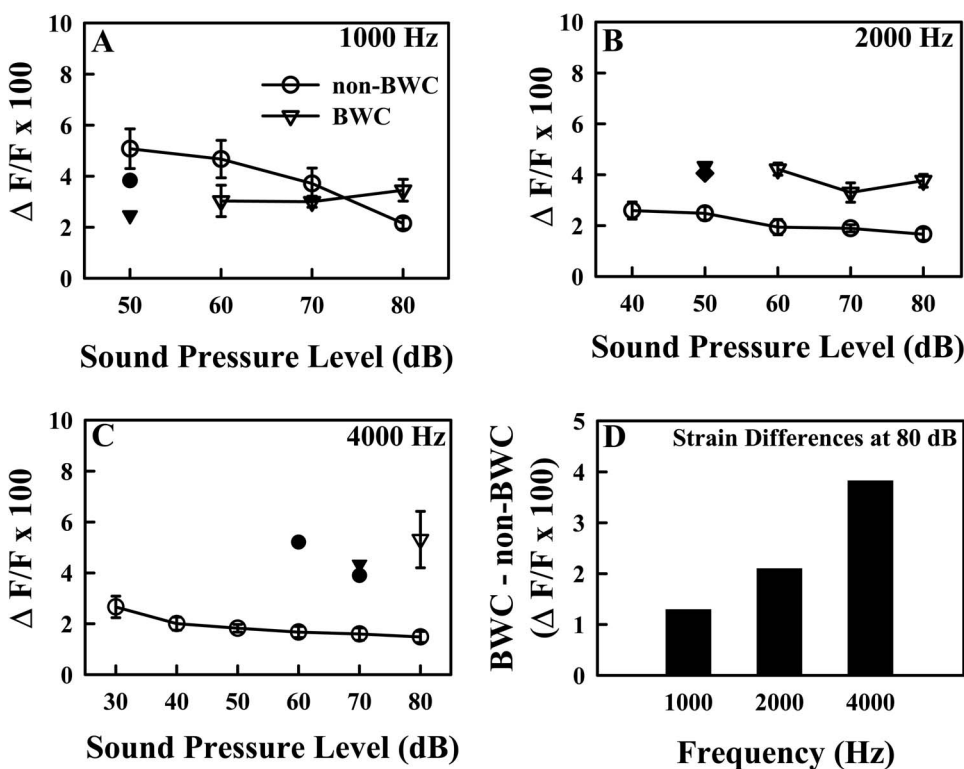


FIG. 2. Frequency difference limens for non-BWC (open circles; $n=4$) and BWC (open triangles; $n=4$) for (a) 1000, (b) 2000, and (c) 4000 Hz tones, and (d) strain differences at 80 dB SPL. Open symbols represent average data and closed symbols show individual data where average data were not available because not all subjects could be evaluated due to stimulus inaudibility. Error bars indicate standard error.

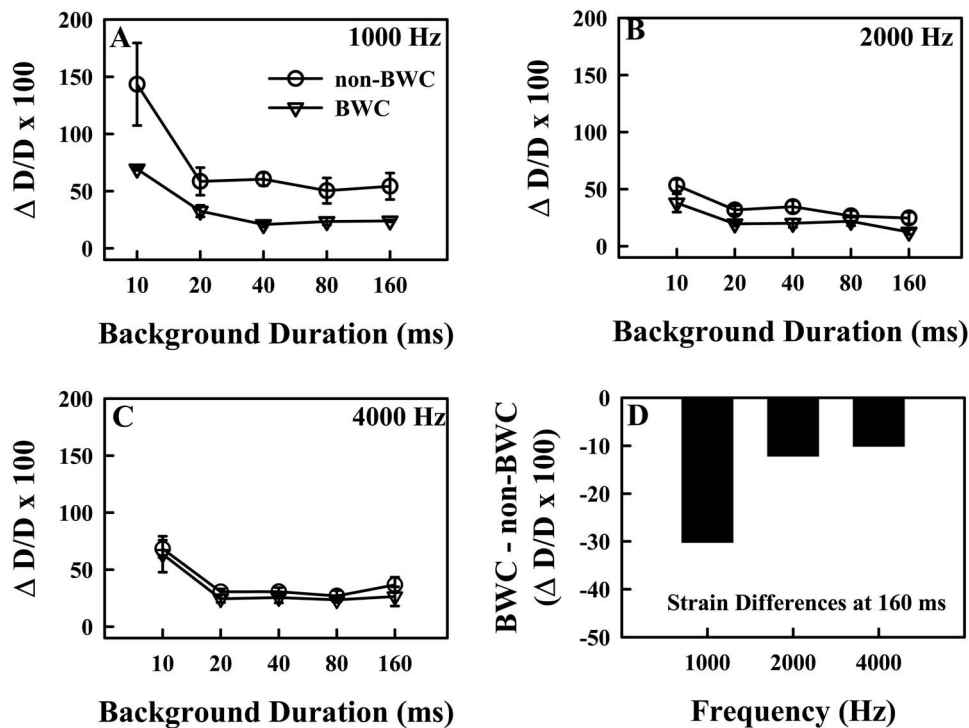


FIG. 3. Average duration difference limens for non-BWC (open circles; $n=4$) and BWC (open triangles; $n=4$) for (a) 1000, (b) 2000, and (c) 4000 Hz tones, and (D) strain differences at 160 ms. Error bars indicate standard error.

non-BWC when all reference frequencies are taken into account ($r^2=0.627$, $p<0.0001$). However, a stronger correlation occurs between FDLs at 80 dB SPL and absolute threshold if the 1000 Hz data are excluded ($r^2=0.814$, $p<0.0001$). Further, an analysis of the 1000 Hz data alone reveals no significant correlation between FDLs and absolute thresholds ($r^2=-0.021$, $p=0.732$). These correlations, coupled with the fact that damage along the BWC basilar papilla is worse in the apex and midsection than in the basal region (Gleich *et al.*, 1994; Weisleder and Park, 1994), suggest that temporal coding mechanisms are responsible for frequency discrimination at low frequencies in canaries.

Differences in frequency discrimination ability between BWC and non-BWC may be related to differences in vocalizations. The vocalizations of BWC contain most of their energy below 4000 Hz (Nottebohm and Nottebohm, 1978; Güttinger, 1985; Okanoya *et al.*, 1990; Wright *et al.*, 2004). In contrast, the vocalizations of non-BWC often have a significant amount of energy present at frequencies up to 6000 or 7000 Hz (Güttinger, 1985; Lohr *et al.*, 2004). Reduced frequency discrimination ability at higher frequencies and good discrimination ability at lower frequencies might actually aid BWC in attending to strain-specific vocalizations.

IV. EXPERIMENT 2—DURATION DISCRIMINATION

Despite the temporal stereotypy of most birdsong, few studies have investigated the ability to detect changes in sound duration in birds. Only two species have been tested. Budgerigars and starlings can detect about a 10–20% increase in the duration of tones (Dooling and Haskell, 1978; Maier and Klump, 1990). Analysis of canary song structure reveals components that occur on several time scales (Güttinger, 1979; 1981, 1985). Singing bouts can last for many minutes, tours (repetitions of a single syllable) typically last

several seconds, syllables range from approximately 50 to 300 ms, and individual notes range from about 10 to 300 ms. Sometimes, different canary syllables have similar frequency structure but differ in duration. Thus, it is of interest to determine duration discrimination ability in BWC and non-BWC. In this experiment, duration difference limens (DDLs) were measured for 1000, 2000, and 4000 Hz tones in BWC and non-BWC for a range of reference durations similar to the range of durations that are characteristic of canary song syllables.

A. Methods: Stimuli and procedures

Stimuli were 1000, 2000, and 4000 Hz pure tones with 5 ms \cos^2 rise/fall times. Background durations of 10, 20, 40, 80, and 160 ms were used. Stimuli were presented at 80 dB SPL to ensure that the tones were at least 10 dB above absolute thresholds for BWC at all frequencies. Thresholds for detecting increases in duration were measured in 4 BWC and 4 non-BWC. Target tones were presented in increments of 10 or 20% of the background durations, depending on the bird's performance. The average false alarm rate was 2.7% for non-BWC and 3.1% for BWC. Based on the same criteria used in Experiment 1, 2% of the data from non-BWC were discarded, and 4% of the data from BWC were discarded.

B. Results and discussion

DDLs for (a) 1000, (b) 2000, and (c) 4000 Hz tones for BWC and non-BWC are shown in Fig. 3. Strain differences (BWC DDL—non-BWC DDL) for the 160 ms tone are shown in Fig. 3(d). Non-BWC were able to detect an increase of approximately 25–30% in duration for stimuli that were longer than 10 ms at 2000 and 4000 Hz, and 50–60% at 1000 Hz. These thresholds are somewhat larger than DDLs reported in other bird species (Dooling and Haskell,

1978; Maier and Klump, 1990). BWC were able to detect increases of 20–30% in duration above 10 ms for all frequencies. DDLs increased at the shortest duration tested (10 ms) for both non-BWC (50–140%) and BWC (40–70%) and varied more among individual birds than at other durations. Budgerigars and starlings also show increased DDLs for very short durations (Dooling and Haskell, 1978; Maier and Klump, 1990). A strain \times frequency \times duration mixed factor ANOVA revealed significant effects of strain [$F(1, 6) = 26.897, p = 0.002$] and duration [$F(4, 24) = 28.128, p < 0.0001$], and significant interactions between duration and strain [$F(4, 24) = 5.903, p = 0.002$] and between duration and frequency [$F(8, 48) = 24.752, p < 0.0001$]. The significant interaction between duration and strain indicates that there may be at least some differences between BWC and non-BWC for certain reference durations. The significant interaction between duration and frequency also indicates that duration discrimination ability is not completely independent of frequency in these birds. There was not a significant main effect of frequency, and no other interactions were significant.

These results show that the inner ear abnormalities in BWC have no detrimental effect on duration discrimination. Considering that the stimulus presentation level was 80 dB SPL, BWC were listening at a reduced sensation level (SL) compared to non-BWC. However, this reduced audibility of the stimuli did not have a negative effect on performance. Surprisingly, BWC were actually slightly better than non-BWC at discriminating changes in the duration of a 1000 Hz tone, though the effect is not statistically significant (i.e., there was no significant strain by frequency interaction). Differences in DDLs as a function of frequency have not been reported in humans or other nonhuman animals.

V. EXPERIMENT 3—INTENSITY DISCRIMINATION

Most avian species can detect intensity changes as small as 1–4 dB (reviewed in Dooling *et al.*, 2000). This ability is assumed to be based on the perceived loudness of sounds. Clearly, the dynamic range of the BWC auditory system is restricted at higher frequencies as a consequence of the hearing loss. However, it is unknown how other aspects of the perception of sound intensity are affected in BWC.

In an earlier experiment, intensity difference limens (IDLs) for continuous broadband noise were measured in BWC (Okanoya and Dooling, 1985). IDLs ranged from approximately 2.9 to 1 dB for noise levels between 60 and 90 dB SPL. Between 70 and 90 dB SPL, IDLs changed very little with increasing level. This result is consistent with predictions from Weber's law and with reported IDLs for noise in budgerigars (Dooling and Searcy, 1981). Performance in BWC was worse at lower sound levels presumably because of the inaudibility of high frequency noise components. To test whether discrimination of intensity changes in pure tones are also affected in BWC, IDLs were measured as a function of frequency and level in BWC and non-BWC. Birds were tested at equal SLs in order to make comparisons at points that were an equivalent amount above absolute threshold across the dynamic range of each strain.

A. Methods: Stimuli and procedures

Stimuli were 1000, 2000, and 4000 Hz pure tones with durations of 400 ms with 20 ms \cos^2 rise/fall times. Target tones were presented with an increment size of 1 or 2 dB, depending on the bird's performance. Thresholds for the smallest detectable increase in intensity were measured in 4 BWC and 4 non-BWC at levels of 10, 20, 30, and 40 dB SL. It was not possible to test BWC at all SLs for all frequencies due to the high levels of sound necessary to determine thresholds. The average false alarm rate was 2.5% for non-BWC and 3.3% for BWC. Based on the same criteria used in Experiment 1, 2% of the data from non-BWC were discarded, and 4% of the data from BWC were discarded.

B. Results and discussion

IDLs for BWC and non-BWC at equal SLs are plotted in Fig. 4 for (a) 1000, (b) 2000, and (c) 4000 Hz. Individual data points are plotted where no average data were computed because not all birds could perform the task at all SLs. IDLs for non-BWC ranged from approximately 3 to 6 dB. Strain differences (BWC IDL—non-BWC IDL) for the 10 dB SL condition are shown in Fig. 4(d). Overall, BWC had somewhat lower IDLs than non-BWC, but IDLs decreased with increasing level for both strains. The range of IDLs reported here for non-BWC is consistent with those reported for pure tones in other bird species (Dooling and Saunders, 1975; Dooling and Searcy, 1979; Hienz *et al.*, 1980; Klump and Baur, 1990; Wright *et al.*, 2003). As intensity difference limens in BWC and non-BWC expressed in dB did not remain constant across testing levels, intensity discrimination of pure tones in both BWC and non-BWC deviates from Weber's law. This result is also consistent with intensity discrimination studies in other bird species and other vertebrates (Fay, 1988), in which IDLs to noise follow Weber's law and IDLs to tones do not.

Because BWC were not tested at all of the levels at each frequency that non-BWC were tested on, it was not possible to perform an ANOVA on the entire set of data. Thus, a mixed factor ANOVA (strain \times frequency) was conducted on the IDLs for the 10 dB SL condition only. This condition was chosen because all animals were tested at 10 dB SL at all frequencies. There was a significant effect of strain [$F(1, 6) = 36.865, p = 0.001$], and a marginally significant effect of frequency [$F(2, 12) = 3.909, p = 0.049$]. The interaction between strain and frequency was not significant.

Although IDLs were consistently smaller in BWC than in non-BWC at 10 dB SL, this finding may reflect the higher SPL levels used in testing BWC and not an enhanced discrimination ability in the hearing-impaired birds. At the points where BWC and non-BWC can be compared at equivalent SPLs, BWC performed about as well as non-BWC. This is consistent with earlier reports of intensity discrimination in hearing-impaired humans, who often perform better than listeners with normal hearing at equal SLs, but not at equal SPLs (Turner *et al.*, 1989; Glasberg and Moore, 1989). Despite the significant amount of missing and damaged hair cells along and across the BWC basilar papilla and the reduced number of auditory nerve fibers, BWC must ob-

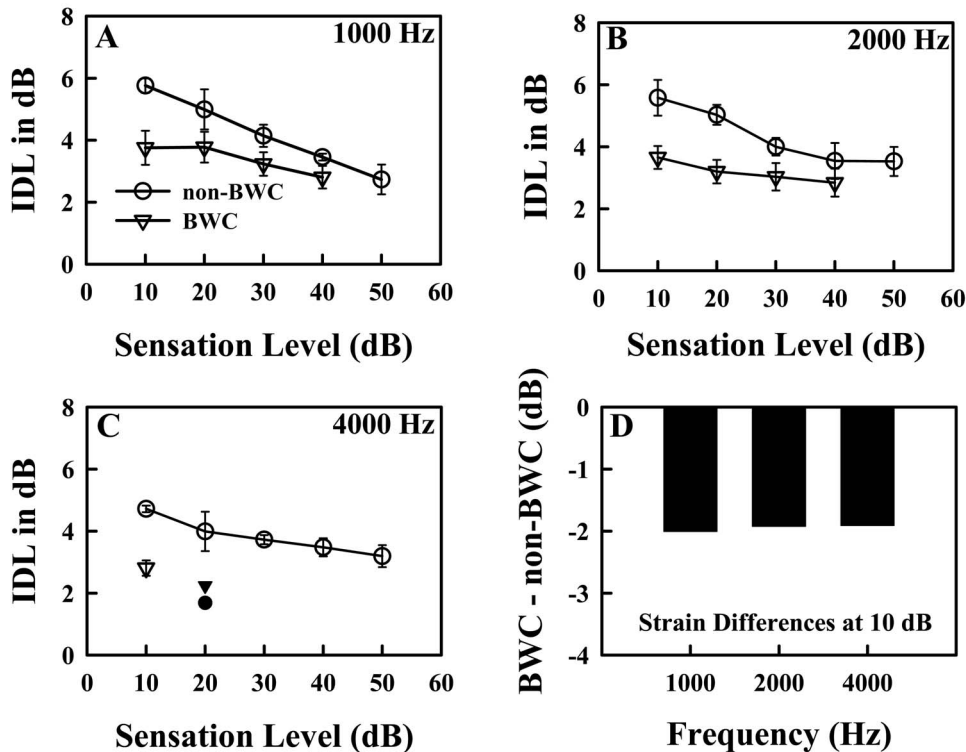


FIG. 4. Intensity difference limens for non-BWC (open circles; $n=4$) and BWC (open triangles; $n=4$) for (a) 1000, (b) 2000, and (c) 4000 Hz tones, (d) and strain differences at 10 dB SL. Open symbols represent average data. Closed symbols show individual data where average data were not available. Error bars indicate standard error.

viously have enough afferent input to encode sound intensity sufficiently to produce normal behavioral results. This is consistent with Viemeister (1988) who suggested that changes in firing rate in a small number of neurons can account for intensity discrimination abilities in mammals.

VI. EXPERIMENT 4—TEMPORAL INTEGRATION

Temporal integration refers to the auditory system's ability to integrate acoustic energy over time, and is reflected in the relationship between the duration of a sound and the threshold for detection of that sound (Hughes, 1946). Detection thresholds decrease exponentially as the duration of a sound increases from a few milliseconds to several hundred milliseconds. No further significant decreases in threshold are observed beyond the asymptotic value. The rate of threshold improvement with increasing duration is typically 2–3 dB/doubling of duration (Saunders and Salvi, 1993). Earlier studies of temporal integration in birds indicate about a 10–20 dB threshold improvement with increasing stimulus duration (Dooling, 1979; Dooling and Searcy, 1985; Klump and Maier, 1990; Saunders and Salvi, 1993).

Temporal integration in BWC and non-BWC was measured for several tone frequencies. Saunders *et al.* (1995) demonstrated that temporal integration is reduced in chickens with temporary hearing loss resulting from hair cell damage. Similarly, BWC were expected to show smaller changes in threshold with increasing duration compared to non-BWC.

A. Methods: Stimuli and procedures

Pure tones of 1000, 2000, and 4000 Hz with durations of 5, 10, 20, 40, 80, 160, 240, 320, and 480 ms and \cos^2 rise/fall times of 2 ms were used as target stimuli. Thresholds for detecting tones of different durations presented in quiet were

measured in 4 non-BWC and 4 BWC. Tones were presented in increments of 5 dB within a block of 10 trials. The average false alarm rate was 3.0% for non-BWC and 2.9% for BWC. Using the same criteria as in the previous experiments, 13% of the data from non-BWC were discarded, and 5% of the data from BWC were discarded. The amount of discarded data is somewhat high for the non-BWC because the birds initially had very unstable behavior when detecting the tones with the smallest durations.

B. Results and discussion

To facilitate comparisons of the amount of threshold change in non-BWC and BWC, relative thresholds (threshold at duration x —threshold at longest duration) were calculated for individual birds at each frequency. Average relative thresholds as a function of tone duration for (A) 1000, (B) 2000, and (C) 4000 Hz are shown in Fig. 5 for non-BWC and BWC. Strain differences in the amount of threshold improvement between 5 and 480 ms are shown in Fig. 5(d). Threshold-by-duration functions were shallower for BWC than for non-BWC at all durations tested. Non-BWC showed a decrease in threshold of about 10 to 15 dB with increasing stimulus duration for all frequencies. This rate of threshold change over the range of durations tested here of 1.5–2.5 dB/doubling of duration is generally consistent with reports in other species of birds (Dooling, 1979; Dooling and Searcy, 1985; Klump and Maier, 1990; Saunders and Salvi, 1993).

Relative thresholds improved by about 7 dB with increasing duration at 2000 Hz in BWC, decreasing at a rate of about 1.2 dB/doubling, but showed less than 5 dB of improvement at 1000 and 4000 Hz (less than 1 dB/doubling threshold improvement). The amount of temporal integration

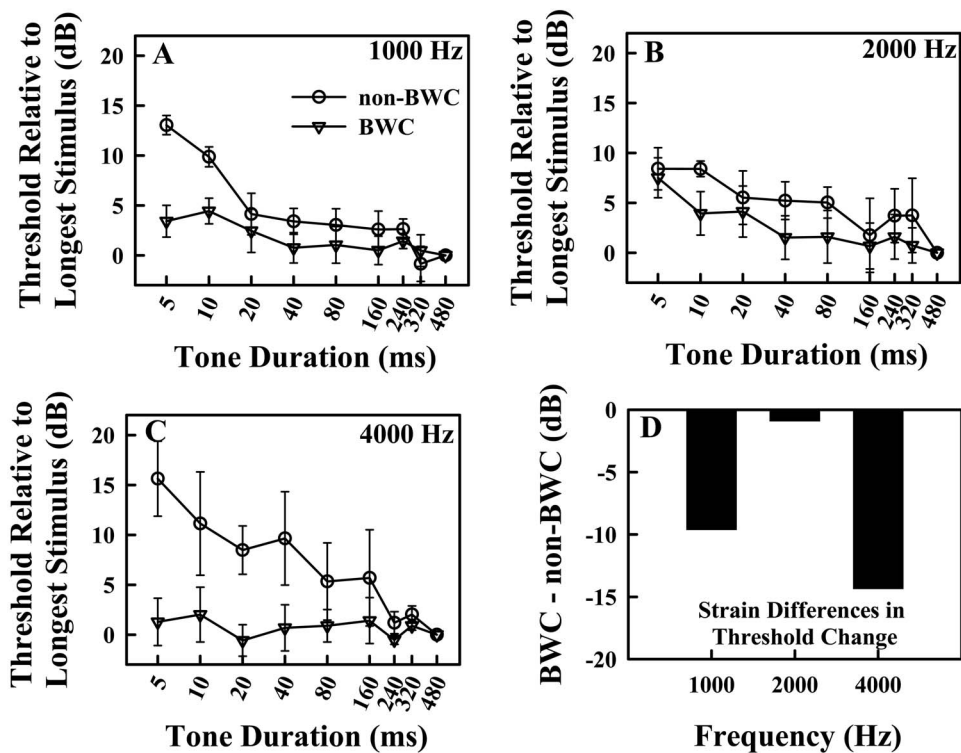


FIG. 5. Average absolute thresholds for pure tones as a function of duration for non-BWC (open circles; $n=4$) and BWC (open triangles; $n=4$) for (a) 1000, (b) 2000, and (c) 4000 Hz, and (d) strain differences in threshold improvement. Error bars indicate standard error.

in BWC is consistent with the amount of temporal integration reported in chickens with temporary hearing loss (Saunders *et al.*, 1995). A strain \times frequency \times duration mixed factor ANOVA revealed significant effects of duration [$F(8, 48)=28.409, p<0.0001$] and strain [$F(1, 6)=14.099, p=0.009$], and significant interactions of duration and strain [$F(8, 48)=7.264, p<0.0001$] and frequency, duration, and strain [$F(16, 96)=2.017, p=0.019$]. These interactions indicate that duration-dependent changes in threshold showed differing dependence on frequency in BWC and non-BWC. No other interactions were significant. There was no significant effect of frequency, indicating that temporal integration does not change systematically with frequency. Two stimulus durations, 5 and 240 ms, were chosen for post hoc comparisons using the Bonferroni method. Relative thresholds at 5 ms were higher in non-BWC than in BWC at 1000 Hz [$t(6)=5.166, p=0.002$] and 4000 Hz [$t(6)=3.905, p=0.008$], but not at 2000 Hz. Relative thresholds at 240 ms were not significantly different between non-BWC and BWC at any frequency. Thus, threshold improved more with increasing duration (indicating increased temporal integration) for non-BWC than BWC at 1000 and 4000 Hz, but not at 2000 Hz. The amount of threshold improvement at 2000 Hz was similar in BWC and non-BWC. The audiogram shows the best sensitivity at about 2000 Hz in BWC. It is possible that the portion of the stimulus which is effective in driving the hair cells is most unaffected by damage in this region, thereby supporting good temporal integration (Neubauer and Heil, 2004).

Smaller changes in threshold with increasing stimulus duration have typically been attributed to reduced temporal integration resulting from reduced peripheral compression in hearing-impaired human listeners (Gengel and Watson, 1971; Pedersen and Eberling, 1973; Elliott, 1975; Chung, 1981;

Hall and Fernandes, 1983; Carlyon *et al.*, 1990) and cats with experimentally induced hearing loss (Solecki and Gerken, 1990). However, a new analysis of the data from cats indicates that the reduction in threshold change associated with hearing loss is actually due to changes in the effective portion of the stimulus rather than changes in the temporal integration mechanism (Neubauer and Heil, 2004). That is, not only is there a reduction in sensitivity with hearing loss, but also an elevation in the baseline above which sound pressure is effective in exciting the auditory system. A similar explanation may hold for the smaller threshold change with increasing stimulus duration observed in BWC.

VII. EXPERIMENT 5—GAP DETECTION

Although measures of temporal integration describe how an organism combines auditory information over a period of time, these measures do not describe the ability to resolve temporal fluctuations in sounds. A simple and convenient measure of temporal resolution of the auditory system can be obtained by measuring thresholds for detecting temporal gaps, or brief silent periods, in noise. Starlings, zebra finches, budgerigars, and barn owls can detect about a 2–3 ms gap in noise with levels exceeding 20 dB SL (Okanoya and Dooling, 1990; Klump and Maier, 1989; Klump *et al.*, 1998). Gap detection thresholds (GDTs) for birds increase at lower noise levels (Okanoya and Dooling, 1990; Klump and Maier, 1989; Klump *et al.*, 1998). Okanoya and Dooling (1990) found that a reciprocal relationship between resolution of gaps in noise and spectral resolution exists in the zebra finch. Thresholds for gaps in octave-band noise were easily predicted from critical ratios in zebra finches, where larger GDTs corresponded to smaller critical ratio values. These results are consistent with the time/

frequency resolution trade-off described in theories of the filtering properties of the mammalian ear (Duifhuis, 1973; de Boer, 1985). A similar relationship was not found in budgerigars, probably due to their unusual critical ratio function (Okanoya and Dooling, 1990). These species differences in the relationship between frequency selectivity and temporal resolution may reflect more general differences between songbirds and parrots.

In the present experiment, thresholds for detecting gaps in broadband noise-bursts were measured in BWC and non-BWC. If an inverse relationship between frequency selectivity and temporal resolution exists in canaries as it does in zebra finches, then BWC should have smaller GDTs than non-BWC provided all components of the stimuli are clearly audible. However, if temporal coding of stimuli is compromised in BWC, then they should show larger GDTs than non-BWC. GDTs were expected to decrease with increasing sound level in both strains.

A. Methods: Stimuli and procedures

Background stimuli were 300 ms bursts of Gaussian noise with 5 ms \cos^2 rise/fall times, sampled at 40 kHz and lowpass filtered at 15 kHz. The target sounds were noise-bursts with silent gaps of different durations centered within the noise-burst. Rise/fall times of the noise on either side of the gap were essentially instantaneous (i.e., the minimum time specified by the stimulus generation software). The total duration of the target sounds, including gaps and noise-bursts, was kept at a constant 300 ms. Thresholds were measured in 4 BWC and 4 non-BWC.

Absolute thresholds for noise-bursts were measured to establish audibility levels for noise-bursts. GDTs were measured by randomly alternating target sounds (noise-bursts containing gaps) with repeating background noise-bursts without gaps. GDTs were measured for noise-bursts with levels of 60, 65, 70, and 75 dB SPL in BWC and 40, 50, 60, 65, 70, and 75 dB SPL in non-BWC. Noise-bursts with gaps of different durations were presented with a step size of 1 or 2 ms, depending on the birds' behavior. The average false alarm rate was 2.6% for non-BWC and 4.0% for BWC. Non-BWC initially showed unstable behavior at the lower sound levels tested. Based on criteria used in the previous experiments, 17% of the data from non-BWC were discarded, and 7% of the data from BWC were discarded.

B. Results and discussion

Average absolute thresholds for noise-bursts were significantly lower for non-BWC (mean=22.08 dB SPL, SD=3.91) than for BWC (mean=47.23 dB SPL, SD=5.89) [$t(6)=-7.089, p<0.0001$]. This difference in thresholds for noise-bursts between non-BWC and BWC is comparable to differences in pure tone thresholds between the two strains for frequencies above 2000 Hz. Presumably, thresholds for noise-bursts are higher in BWC than in non-BWC due to the inaudibility of the high frequency components of the noise.

Average gap detection thresholds for BWC and non-BWC are shown in Fig. 6. Thresholds increased from 3.62 ms at 75 dB SPL to 6.49 ms at 40 dB SPL for non-

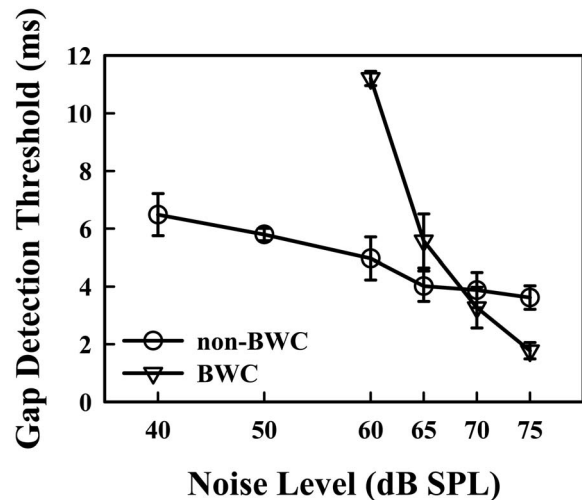


FIG. 6. Average gap detection thresholds as a function of sound pressure level for BWC (open circles; $n=4$) and non-BWC (open triangles; $n=4$). Error bars indicate standard error.

BWC. These GDTs are within the range of those reported in other species of birds (Klump and Maier, 1989; Okanoya and Dooling, 1990; Klump *et al.*, 1998). Thresholds for BWC increased from 1.78 ms at 75 dB SPL to 11.21 ms at 60 dB SPL. BWC were not tested at lower levels due to audibility constraints. In the range of 60 to 75 dB SPL, BWC showed a much larger change as a function of level than non-BWC.

A strain \times level mixed factor ANOVA revealed significant effects of strain [$F(1, 16)=224.464, p<0.0001$] and level [$F(3, 18)=62.909, p<0.0001$], and a significant interaction between strain and level [$F(3, 18)=48.023, p<0.0001$]. Post hoc analyses using the Bonferroni method showed that GDTs for BWC were not significantly different from GDTs for non-BWC at 65 and 70 dB SPL. GDTs for BWC were larger than GDTs for non-BWC at 60 dB SPL [$t(4)=8.875, p<0.05$], and were lower than for non-BWC at 75 dB SPL [$t(4)=-3.851, p<0.05$].

The sharp decrease in resolution of gaps between 75 and 60 dB SPL in BWC may be related to the low audibility of high frequency components of the noise-bursts at 60 and 65 dB. Recall that thresholds for noise-bursts were approximately 25 dB lower in non-BWC than in BWC. Thus, 60 dB SPL is approximately 35 dB SL in non-BWC, but only 10 dB SL in BWC. As the higher frequency components become less audible, BWC performance decreases.

The superior temporal resolution at higher sound levels in BWC may be related to the wider bandwidth of auditory filters in the area of hearing loss. An earlier experiment showed impaired frequency resolution at high frequencies in BWC (Lauer and Dooling, 2002; Lauer *et al.* 2002, 2006). The wider auditory filters associated with poor frequency resolution may result in less smoothing of the input wave form over time, thus preserving more of the fine details of the signal that aid gap detection. Additionally, wider filters have less ringing than narrow filters. Ringing may obscure the abrupt offset of the noise when a gap occurs, thereby increasing GDTs for non-BWC.

VIII. EXPERIMENT 6—DISCRIMINATION OF CHANGES IN TIME-REVERSED HARMONIC COMPLEXES

The vocalizations of birds can be quite temporally complex, with both slow envelope fluctuations and fast within-period fluctuations. Canary syllables and calls often include upward or downward frequency sweeps. The majority of studies of temporal resolution in birds have focused on resolution of slow overall changes in the envelope of sounds. Recently, [Dooling *et al.* \(2002\)](#) and [Lohr *et al.* \(2006\)](#) presented evidence that birds are superior to humans when envelope and frequency cues are removed and discrimination of sounds must rely on cues in the temporal fine structure (within-period temporal fluctuations).

The differences in resolution of within-period temporal changes between birds and humans are thought to be related in part to the width of the auditory filters. Broader auditory filters should lead to better within-channel temporal resolution in a linear system because there is better preservation of phase relationships as more components fall within one channel ([Duifhuis, 1973](#); [de Boer, 1985](#)). Behavioral estimates of auditory filter bandwidth suggest that birds generally have broader filters than humans (reviewed in [Dooling *et al.*, 2000](#)). However, tuning curves of some auditory nerve fibers in birds are more narrowly tuned than in mammals (reviewed in [Gleich and Manley, 2000](#)). Thus, the relationship between temporal acuity and peripheral auditory filtering mechanisms remains unclear.

Experiment 5 demonstrated that BWC are actually better than non-BWC at detecting changes in the envelope of sounds under certain conditions. The present experiment addresses the question of whether or not resolution of fast within-period temporal changes is also enhanced in BWC. The ability to discriminate changes in temporal fine structure was measured in BWC and non-BWC using procedures identical to those of [Dooling *et al.* \(2002\)](#).

A. Methods: Stimuli and procedures

The stimuli have been described in detail elsewhere ([Dooling *et al.*, 2002](#)). Harmonic complexes were composed of equal amplitude components with component starting phases selected according to the [Schroeder \(1970\)](#) algorithm [$\theta_n = \pm \pi n(n+1)/N$, where θ =component starting phase, n =component number within the complex, and N =total number of components]. Reversing the sign of the phase algorithm results in two wave forms that are the reverse of each other in time. Complexes with fundamental frequencies of 150, 200, 300, 400, 600, 800, and 1000 Hz were created. The periods of these stimuli ranged from 1 to 6.67 ms. The wave forms were 260 ms in duration, with 20 ms \cos^2 onset/offset ramps. Stimuli were presented at 80 dB SPL. Sections of negative and positive-phase stimuli with a fundamental frequency of 200 Hz are shown in Fig. 7(a).

Three BWC were tested using procedures identical to those of [Dooling *et al.* \(2002\)](#). The data from BWC were compared to data from 3 non-BWC reported by [Dooling *et al.* \(2002\)](#). Birds were trained to discriminate between negative-phase and positive-phase wave forms at each fundamental frequency. As birds required extra training when

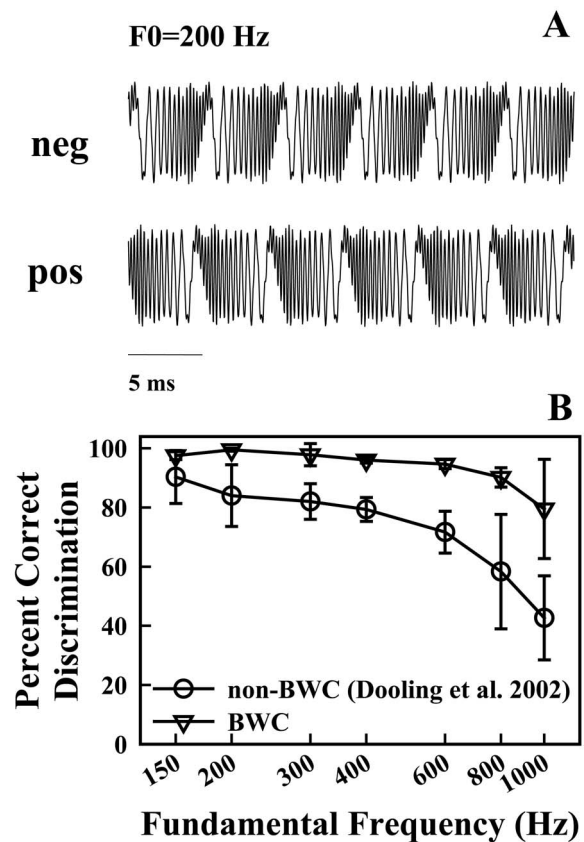


FIG. 7. (A) Examples of harmonic complexes used as stimuli in Experiment 6. Complexes were constructed by summing equal amplitude components with starting phases selected according to the Schroeder algorithm. Complexes with higher fundamental frequencies have shorter period durations. (B) Average percent correct discriminations of harmonic complexes for non-BWC (open circles; $n=3$; [Dooling *et al.*, 2002](#)) and BWC (open triangles; $n=3$). Error bars indicate standard error.

switching between negative-phase and positive-phase background sounds, all fundamental frequencies were tested for a given phase selection (negative or positive) before switching to the opposite phase selection. Whether a bird began with the negative-phase or the positive-phase background sounds was chosen randomly.

Behavior was considered stable if the percent correct for a given target did not change more than 10% within the last two 100-trial blocks. If the bird's percent correct remained at 100% for the first two blocks of 100 trials, the bird was not run on more trials for that condition. Percent correct discrimination values for each experimental condition were taken as the mean percent correct over the last 200 trials run. The average false alarm rate was 4.5% for BWC. Data from sessions with false alarm rates larger than 18% were excluded from analysis. Seven percent of the data were discarded.

B. Results and discussion

Figure 7(b) shows percent correct discriminations for harmonic complexes with different fundamental frequencies for BWC along with data from non-BWC (from [Dooling *et al.*, 2002](#)). Average BWC discrimination ability was high (80% correct or better) for complexes with fundamental fre-

quencies up to 1000 Hz, whereas non-BWC discrimination performance decreased to less than 60% correct for complexes with fundamental frequencies of 800 and 1000 Hz. BWC were able to perform above 80% correct even for stimuli with periods as short as 1.0 ms. This is well beyond the human limit, which falls at about 300–400 Hz (Dooling *et al.*, 2002). Overall, percent correct discriminations for BWC were almost as high as those reported for zebra finches and were slightly higher than in budgerigars (Dooling *et al.*, 2002).

BWC showed better performance than non-BWC at most fundamental frequencies; however a mixed factor (strain \times fundamental frequency) ANOVA showed that the main effect of strain was not significant. The lack of a significant main effect of strain is probably due to similar performance between non-BWC and BWC at some fundamental frequencies. There was a significant effect of fundamental frequency [$F(6,24)=16.056$, $p<0.0001$] and a significant interaction between fundamental frequency and strain [$F(6,24)=2.667$, $p=0.040$]. Estimates of effect sizes were calculated in the absence of a significant effect of strain. Large effects occurred for fundamental frequency [$\eta^2=0.801$] and strain [$\eta^2=0.636$]. A moderate effect size was found for the interaction between strain and fundamental frequency [$\eta^2=0.40$]. Post hoc comparisons using the Bonferroni method showed that BWC had higher percent correct discriminations only for complexes with fundamental frequencies of 300 Hz [$t(4)=3.354$, $p<0.05$] and 600 Hz [$t(4)=3.137$, $p<0.05$]. The excellent resolution of fast temporal changes seen in BWC may be attributed in part to wider filter bandwidths at higher frequencies. Wider filters allow more components of the harmonic complexes to fall within one channel, thereby preserving more of the phase relationships between components than in narrower filters.

IX. GENERAL DISCUSSION

This study examined in detail the hearing abilities in an animal with early-onset hereditary high-frequency hearing loss bred for its unique low-pitched vocalizations—the Belgian Waterslager canary. Remarkably, only some aspects of hearing investigated in this study were impaired, despite the extensive damage that occurs along the BWC basilar papilla. Frequency discrimination was quite good at 1000 Hz in BWC; however, discrimination of changes in higher frequencies was worse in BWC than in non-BWC. Duration discrimination was also similar in BWC and non-BWC overall, but was slightly better in BWC at 1000 Hz. Intensity discrimination was also not adversely affected in BWC. Temporal integration was reduced at 1000 and 4000 Hz but not at 2000 Hz in BWC, gap detection was especially good in BWC provided the high frequency components of the stimuli were clearly audible, and discrimination of fast within-period temporal changes was also somewhat better than normal in BWC. Taken all together, these results seem to point to deficits that are mostly related to impaired spectral resolution that typically results from sensorineural hearing loss, and an accompanying enhancement in temporal processing.

The auditory sensitivities that are clearly abnormal in this population are absolute thresholds, frequency resolution and frequency discrimination at high frequencies (where temporal processing may suffer from reductions in phase locking in the auditory nerve), and temporal integration. The latter may be impaired because of loss of effective bandwidth due to reduced audibility, also a spectral effect. Interestingly, the enhancement of temporal resolution, which may have had a role in most if not all of the normal or supernormal results reported here, occurred in spite of the fact that BWC show some abnormalities in the auditory brainstem nuclei involved in temporal processing. Nucleus magnocellularis and nucleus laminaris show normal cell number and organization in adult birds; however, both have reduced volumes attributed to smaller cell size (Kubke *et al.*, 2002). Presumably, the smaller cell size is a result of the progressive reduction of auditory input from an early age. It is possible that either the reduced volume is simply not enough to have a profound impact on temporal resolution, or that some subcellular mechanism such as an increased thickness of the postsynaptic densities compensates for the reduction in cell size.

The results from these experiments show that there are significant perceptual consequences of the early-onset hearing loss in BWC other than elevated pure tone thresholds. Still, BWC hear remarkably well given the extensive degree of basilar papillar damage. The unique perceptual profile in BWC can, in some ways, be viewed as a compensation for loss of high-frequency hearing. Although these birds have poor detection and resolution of high frequency sounds, they have excellent temporal resolution and intensity discrimination. In addition to enhanced temporal resolution, which may be a serendipitous byproduct of hair cell damage, BWC may have developed more central compensatory mechanisms that promote processing of strain-specific sounds in the absence of reliable high frequency cues. This success is even more remarkable if one considers that the BWC are actually listening at a lower sensation level than non-BWC in their everyday environment.

The link between perceptual abilities and vocal characteristics in BWC highlights the uniqueness of this animal system. These birds show a unique pattern of perceptual abilities that may enhance their ability to learn and produce strain-specific vocalizations. This model is exceptional in that it is the only animal in which we can investigate the link between genetics, auditory system structural abnormalities, vocal learning, vocal production, hair cell regeneration, and hearing abilities. The close correspondence between auditory perception and vocal characteristics in BWC suggests that the auditory pathology is a product of artificial selection. As breeders mated birds with desirable low-pitched song elements, they may have inadvertently selected for auditory system abnormalities. In essence, breeders may have artificially produced a specialization in BWC over several hundred years.

The BWC model has a unique potential to further our understanding of the evolution of vocal learning and production and the role of genetics in hearing and auditory pathology. We now have an extensive behavioral assay to comple-

ment the many studies describing the auditory pathology in BWC. To our knowledge, this is the most comprehensive psychophysical exploration in an animal with hereditary hearing loss. Future studies should investigate the specific genes responsible for the hair cell abnormalities, the relationships among basilar papilla microstructure, hearing abilities, and the physiological response of the auditory system, and the role of genetic predispositions in song learning in BWC.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. DC-01372 to RJD, Grant No. DC-005450 to AML, Grant No. DC-00626 to MRL, and Grant No. DC-04664. Portions of this work were presented at the 2003 and 2004 Midwinter Meeting of the Association for Research in Otolaryngology, Daytona Beach, FL and the 147th Meeting of the Acoustical Society of America.

- Brittan-Powell, E. F., Dooling, R., Wright, T., Munding, P., and Ryals, B. (2002). "Development of auditory sensitivity in Belgian Waterslager (BWC) canaries," *Assoc. Res. Otolaryngol. St. Petersburg, FL*.
- Carlyon, R. P., Buus, S., and Florentine, M. (1990). "Temporal integration of tone pulses by normal and by cochlearly impaired listeners," *J. Acoust. Soc. Am.* **87**, 260–268.
- Chung, D. Y. (1981). "Masking, temporal integration, and sensorineural hearing loss," *J. Speech Hear. Res.* **24**, 514–520.
- de Boer, E. (1985). "Auditory time constants: A paradox?" *Time Resolution in the Auditory System*, edited by A. Michelson (Springer, Berlin), pp. 141–158.
- Dooling, R. J. (1979). "Temporal summation of pure tones in birds," *J. Acoust. Soc. Am.* **65**, 1058–1060.
- Dooling, R. J., and Haskell, R. J. (1978). "Auditory duration discrimination in the parakeet (*Melospittacus undulatus*)," *J. Acoust. Soc. Am.* **63**, 1640–1642.
- Dooling, R. J., Leek, M. R., Gleich, O., and Dent, M. L. (2002). "Auditory temporal resolution in birds: Discrimination of harmonic complexes," *J. Acoust. Soc. Am.* **112**, 748–759.
- Dooling, R. J., Lohr, B., and Dent, M. L. (2000). "Hearing in birds and reptiles," *Comparative Hearing: Birds and Reptiles*, edited by R. J. Dooling, R. R. Fay, and A. N. Popper (Springer, Berlin), pp. 197–248.
- Dooling, R. J., and Okanoya, K. (1995). "The method of constant stimuli in testing auditory sensitivity in small birds," *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhauser, Basel), pp. 161–169.
- Dooling, R. J., Ryals, B. M., Dent, M. L., and Reid, T. L. (2006). "Perception of complex sounds in Budgerigars (*Melospittacus undulatus*) with temporary hearing loss," *J. Acoust. Soc. Am.* **119**, 2524–2532.
- Dooling, R. J., and Saunders, J. C. (1975). "Auditory intensity discrimination in the parakeet (*Melospittacus undulatus*)," *J. Acoust. Soc. Am.* **58**, 1308–1310.
- Dooling, R. J., and Searcy, M. H. (1979). "Relation among CRs, critical bands, and intensity difference limens in the parakeet (*Melospittacus undulatus*)," *Bull. Psychon. Soc.* **13**, 300–302.
- Dooling, R. J., and Searcy, M. H. (1981). "Amplitude-modulation thresholds for the parakeet (*Melospittacus undulatus*)," *J. Acoust. Soc. Am.* **143**, 383–388.
- Dooling, R. J., and Searcy, M. H. (1985). "Temporal integration of acoustic signals by the budgerigar (*Melospittacus undulatus*)," *J. Acoust. Soc. Am.* **77**, 1917–1920.
- Duifhuis, H. (1973). "Consequences of peripheral frequency selectivity for nonsimultaneous masking," *J. Acoust. Soc. Am.* **54**, 1471–1488.
- Elliott, L. L. (1975). "Temporal and masking phenomena in persons with sensorineural hearing loss," *Audiology* **14**, 336–353.
- Fay, R. R. (1988). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay Associates, Winnetka, IL).
- Gardner, T. J., Naef, F., and Nottebohm, F. (2005). "Freedom and rules: The acquisition and reprogramming of a bird's learned song," *Science* **308**, 1046–1049.
- Gengel, R. W., and Watson, C. S. (1971). "Temporal integration: I. Clinical implications of a laboratory study. II. Additional data from hearing-impaired subjects," *J. Speech Hear. Disord.* **36**, 213–224.
- Gescheider, G. A. (1985). *Psychophysics: Method, Theory, and Application* (Lawrence Erlbaum & Associates, New York).
- Glasberg, B. R., and Moore, B. C. J. (1989). "Psychoacoustic abilities in subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech," *J. Acoust. Soc. Am.* **32**, 1–25.
- Gleich, O., Dooling, R. J., and Manley, G. A. (1994). "Inner-ear abnormalities and their functional consequences in Belgian Waterslager canaries (*Serinus canarius*)," *Hear. Res.* **79**, 123–136.
- Gleich, O., Dooling, R. J., and Ryals, B. M. (2001). "A quantitative analysis of the nerve fibers in the VIIIth nerve of Belgian Waterslager canaries with a hereditary sensorineural hearing loss," *Hear. Res.* **151**, 141–158.
- Gleich, O., Klump, G. M., and Dooling, R. J. (1995). "Peripheral basis for the auditory deficit in Belgian Waterslager canaries (*Serinus canarius*)," *Hear. Res.* **82**, 100–108.
- Gleich, O., and Manley, G. A. (2000). "The hearing organ of birds and crocodylian," *Comparative Hearing: Birds and Reptiles*, edited by R. J. Dooling, R. R. Fay, and A. N. Popper (Springer, Berlin), pp. 197–248.
- Güttinger, H. R. (1979). "Integration of learnt and genetically programmed behavior—Study of hierarchical organization in songs of canaries, greenfinches and their hybrids," *Z. Tierpsychol.* **49**, 285–303.
- Güttinger, H. R. (1981). "Self-differentiation of song organization rules by deaf canaries," *Z. Tierpsychol.* **56**, 323–340.
- Güttinger, H. R. (1985). "Consequences of domestication on the song structures in the canary," *Behaviour* **94**, 254–278.
- Hall, J. W., and Fernandes, M. A. (1983). "Temporal integration, frequency resolution, and off-frequency listening in normal-hearing and cochlear-impaired listeners," *J. Acoust. Soc. Am.* **74**, 1172–1177.
- Hienz, R. D., Sinnott, J. M., and Sachs, M. B. (1980). "Auditory intensity discrimination in blackbirds and pigeons," *J. Community Psychol.* **94**, 993–1002.
- Hughes, J. W. (1946). "The threshold of audition for short periods of stimulation," *Proc. R. Soc. London, Ser. B* **133**, 486–490.
- Klump, G. M., and Baur, A. (1990). "Intensity discrimination in the European starling (*Sturnus vulgaris*)," *Naturwiss.* **77**, 545–548.
- Klump, G. M., and Maier, E. H. (1989). "Gap detection in the starling (*Sturnus vulgaris*)," *J. Comp. Physiol., A* **164**, 531–538.
- Klump, G. M., and Maier, E. H. (1990). "Temporal summation in the European starling (*Sturnus vulgaris*)," *J. Comp. Psychol.* **104**, 94–100.
- Klump, G. M., Schwenzfeier, C., and Dent, M. L. (1998). "Gap detection in the barn owl (*Tyto alba*)," *Assoc. Res. Otolaryngol. St. Petersburg, FL*.
- Kubke, M. F., Dent, M. L., Hodos, W., Carr, C. E., and Dooling, R. J. (2002). "Nucleus magnocellularis and nucleus laminaris in Belgian Waterslager and normal strain canaries," *Hear. Res.* **164**, 19–28.
- Kuhn, A., and Saunders, J. C. (1980). "Psychophysical tuning curves in the parakeet: A comparison between simultaneous and forward masking procedures," *J. Acoust. Soc. Am.* **68**, 1892–1894.
- Lauer, A. M., and Dooling, R. J. (2002). "Frequency selectivity in canaries with a hereditary hearing loss," *J. Acoust. Soc. Am.* **111**, 2392.
- Lauer, A. M., Dooling, R. J., and Leek, M. R. (2006). "Impaired frequency resolution in canaries with hereditary hearing loss," *J. Acoust. Soc. Am.* **119**, 1251.
- Lauer, A. M., Dooling, R. J., Leek, M. R., and Lentz, J. J. (2002). "Masking by harmonic complexes in the hearing impaired Belgian Waterslager canary," *Assoc. Res. Otolaryngol. St. Petersburg, FL*.
- Lohr, B., Dooling, R. J., and Bartone, S. (2006). "The discrimination of temporal fine structure in call-like harmonic sounds by birds," *J. Comp. Psychol.* **120**, 239–251.
- Lohr, B. L., Lauer, A. M., Newman, M. R., and Dooling, R. J. (2004). "Hearing in the red-billed firefinch (*Lagonosticta senegalensis*) and the Spanish timbardo canary (*Serinus canaria*): The influence of natural and artificial selection on auditory abilities and vocal structure," *Bioacoustics* **14**, 83–98.
- Maier, E. H., and Klump, G. M. (1990). "Auditory duration discrimination in the European starling (*Sturnus vulgaris*)," *J. Acoust. Soc. Am.* **88**, 616–620.
- Marler, P., and Waser, M. S. (1977). "Role of auditory feedback in canary song development," *J. Comp. Physiol. Psychol.* **91**, 8–16.
- Neubauer, H., and Heil, P. (2004). "Towards a unifying basis of auditory thresholds: The effects of hearing loss on temporal integration reconsidered," *J. Assoc. Res. Otolaryngol.* **5**, 436–458.
- Nienhuys, T. G., and Clark, G. M. (1978). "Frequency discrimination following the selective destruction of cochlear inner and outer hair cells,"

- Science **199**, 1356–1357.
- Nottebohm, F., and Nottebohm, M. (1976). "Left hypoglossal dominance in the control of canary and white-crowned sparrow song," J. Comp. Physiol., A **108**, 171–192.
- Nottebohm, F., and Nottebohm, M. (1978). "Relationship between song repertoire and age in the canary, *Serinus canarius*," Z. Tierpsychol **46**, 298–305.
- Okanoya, K., and Dooling, R. J. (1985). "Colony differences in auditory thresholds in the canary (*Serinus canarius*)," J. Acoust. Soc. Am. **78**, 1170–1176.
- Okanoya, K., and Dooling, R. J. (1987). "Strain differences in auditory thresholds in the canary (*Serinus canarius*)," J. Comp. Psychol. **101**, 213–215.
- Okanoya, K., and Dooling, R. J. (1990). "Detection of gaps in noise by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Poephila guttata*)," Hear. Res. **50**, 185–192.
- Okanoya, K., Dooling, R. J., and Downing, J. D. (1990). "Hearing and vocalizations in hybrid Waterslager-Roller canaries (*Serinus canarius*)," Hear. Res. **46**, 271–275.
- Pedersen, C. B., and Eberling, C. (1973). "Temporal integration of acoustic energy in patients with presbycusis," Acta Oto-Laryngol. **75**, 32–37.
- Prosen, C. A., Halpern, D. L., and Dallos, P. (1989). "Frequency difference limens in normal and sensorineural hearing impaired chinchillas," J. Acoust. Soc. Am. **85**, 1302–1313.
- Ryals, B. M., and Dooling, R. J. (2002). "Development of hair cell stereovilli bundle abnormalities in Belgian Waterslager canary," Assoc. Res. Otolaryngol. St. Petersburg, FL.
- Saunders, S. S., and Salvi, R. J. (1993). "Psychoacoustics of normal adult chickens: thresholds and temporal integration," J. Acoust. Soc. Am. **94**, 83–90.
- Saunders, S. S., Salvi, R. J., and Miller, K. M. (1995). "Recovery of thresholds and temporal integration in adult chickens after high-level 525-Hz pure-tone exposure," J. Acoust. Soc. Am. **97**, 1150–1164.
- Schroeder, M. R. (1970). "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," IEEE Trans. Inf. Theory **16**, 85–89.
- Sek, A., and Moore, B. C. J. (1995). "Frequency discrimination as a function of frequency, measured in several ways," J. Acoust. Soc. Am. **97**, 2479–2486.
- Sinnott, J. M., Sachs, M. B., and Hienz, R. D. (1980). "Aspects of frequency discrimination in passerine birds and pigeons," J. Comp. Physiol. Psychol. **94**, 401–415.
- Solecki, J., and Gerken, G. M. (1990). "Auditory temporal integration in the normal-hearing and hearing-impaired cat," J. Acoust. Soc. Am. **88**, 779–785.
- Turner, C. W., Zwislocki, J. J., and Filion, P. R. (1989). "Intensity discrimination determined with two paradigms in normal and hearing-impaired subjects," J. Acoust. Soc. Am. **86**, 109–115.
- Viemeister, N. F. (1988). "Psychophysical aspects of auditory intensity coding," *Auditory Function*, edited by G. M. Edelman, W. E. Gall, and W. A. Cowan (Wiley, New York), pp. 213–241.
- Waser, M. S., and Marler, P. (1977). "Song learning in canaries," J. Comp. Physiol. Psychol. **91**, 1–7.
- Weisleder, P., Lu, Y., and Park, T. J. (1996). "Anatomical basis of a congenital hearing impairment: basilar papilla dysplasia in the Belgian Waterslager canary," J. Comp. Neurol. **369**, 292–301.
- Weisleder, P., and Park, T. J. (1994). "Belgian Waterslager canaries are afflicted by Scheibe's-like dysplasia," Hear. Res. **80**, 64–70.
- Wright, T. F., Brittan-Powell, E. F., Dooling, R. J., and Mundinger, P. (2004). "Sex-linkage of deafness and song frequency spectrum in the Waterslager strain of domestic canary," Biol. Letters **271**, s409–s412.
- Wright, T. F., Cortopassi, K. A., Bradbury, J. W., and Dooling, R. J. (2003). "Hearing & vocalizations in conures," J. Comp. Psychol. **117**, 87–95.

Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled^{a)}

David R. R. Smith^{b)}

Centre for Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom and Department of Psychology, University of Hull, Cottingham Road, Hull HU6 7RX, United Kingdom

Thomas C. Walters and Roy D. Patterson

Centre for Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

(Received 26 September 2006; revised 17 September 2007; accepted 24 September 2007)

A recent study [Smith and Patterson, *J. Acoust. Soc. Am.* **118**, 3177–3186 (2005)] demonstrated that both the glottal-pulse rate (GPR) and the vocal-tract length (VTL) of vowel sounds have a large effect on the perceived sex and age (or size) of a speaker. The vowels for all of the “different” speakers in that study were synthesized from recordings of the sustained vowels of one, adult male speaker. This paper presents a follow-up study in which a range of vowels were synthesized from recordings of four different speakers—an adult man, an adult woman, a young boy, and a young girl—to determine whether the sex and age of the original speaker would have an effect upon listeners’ judgments of whether a vowel was spoken by a man, woman, boy, or girl, after they were equated for GPR and VTL. The sustained vowels of the four speakers were scaled to produce the same combinations of GPR and VTL, which covered the entire range normally encountered in every day life. The results show that listeners readily distinguish children from adults based on their sustained vowels but that they struggle to distinguish the sex of the speaker. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2799507]

PACS number(s): 43.66.Lj, 43.71.Bp, 43.71.An [RAL]

Pages: 3628–3639

I. INTRODUCTION

Much of the variability in the vowels of men, women, and children arises from characteristic differences in glottal-pulse rate (GPR) (Titze, 1989) and vocal-tract length (VTL) (Fant, 1970; Fitch and Giedd, 1999). GPR is perceived as voice pitch; VTL combines with GPR in the perception of speaker size (Smith and Patterson, 2005). Both GPR and VTL increase with age, and they increase disproportionately for males after puberty. Recent advances in auditory vocoders, such as STRAIGHT (Kawahara and Irino, 2004) and PRAAT (Boersma, 2001), have made it possible to vary the VTL of recorded speech without varying the GPR and vice versa. The ability to vary VTL and GPR while preserving the content of the speech and any other distinctive speaker characteristics has led to a series of studies on the role of GPR and VTL in the perception of vowels (e.g., Assmann and Neary, 2003; Smith *et al.*, 2005), syllables (e.g., Ives *et al.*, 2005), and sentences (e.g., Darwin *et al.*, 2003).

In a recent study, Smith and Patterson (2005) used sustained vowels to determine which of the four responses—man, woman, boy, or girl—would be assigned to vowels with

a wide range of combinations of GPR and VTL. The results showed, as expected, that shorter VTLs and higher GPRs lead to the perception that the speaker is a child, and longer VTLs and lower GPRs lead to the perception that the speaker is an adult. The authors drew attention to an apparent anomaly in the data, which was that the voices were heard as women less often than might have been expected, and they pointed out that all of the vowels for all of the “different” speakers in that study (that is, all of the different combinations of GPR and VTL) were synthesized from the speech of a single adult male. The current paper presents a follow-up study to determine how the characteristics of the original speaker’s voice affect judgments of the speaker’s sex and age, when the GPR and VTL of the voices are controlled. Specifically, the experiment of Smith and Patterson (2005) has been replicated with vowels from four different speakers: The vocoder STRAIGHT was used to synthesize vowels with the same wide range of GPR and VTL values for all four speakers. The vocoder preserves most of the information other than GPR and VTL and enables us to determine whether this extra information affects listeners’ ability to discriminate whether the original speaker was a man, woman, boy, or girl. The results show that it is possible to distinguish whether the original speaker was a child or an adult but it is difficult to discern the sex of the original speaker when GPR and VTL are equated. Portions of this work have been presented at several conferences (Smith *et al.*, 2006; Smith *et al.*, 2007a; Smith *et al.*, 2007b).

^{a)}Portions of this work were presented in “The role of input speaker upon judgements of speaker sex and age,” British Society of Audiology, Cambridge, UK, 2006, “Role of glottal-pulse rate, vocal-tract length and original talker upon judgements of speaker sex and age,” Acoustical Society of America, Salt Lake City, UT, 2007, and “Judging sex and age: Effect of glottal-pulse rate, vocal-tract length and original talker,” International Congress on Acoustics, Madrid, Spain, 2007.

^{b)}Electronic mail: d.r.smith@hull.ac.uk

The GPR and VTL information in sustained vowels. The length and mass of the vocal folds determine the rate at which the vocal folds open and close. The perceptual marker of GPR is voice pitch; the greater the GPR, the higher the perceived pitch of the voice. There is a strong link between speaker sex and GPR (Darwin, 1871; Morton, 1977). Men have pitches about an octave lower than women primarily because the vocal folds of men are about 60% longer than those of women (Titze, 1989). Voice pitch is a highly salient cue to sex and age because large men have low pitches, young children have high pitches, and adult women lie in the middle (averaging around 105, 260, and 220 Hz, respectively, cf. Huber *et al.*, 1999). Voice pitch can be derived from individual sustained vowels and people are highly sensitive to differences in voice pitch—the just noticeable difference is around 2% (Smith *et al.*, 2005). The sexual dimorphism in GPR is attributable to increased testosterone at puberty in males which stimulates growth in the laryngeal cartilages (Beckford *et al.*, 1985).

Although there are clear intergroup pitch differences between men and women, the correlation between GPR and body height within a group of adult men, or a group of adult women, is rarely statistically significant, e.g., Lass and Brown (1978), Künzel (1989), Hollien *et al.* (1994), and González (2004). The correlation between GPR and speaker size is also weakened by our use of GPR variation to make prosodic distinctions, such as the rising pitch contour of the interrogative sentence. Indeed, some individuals vary their pitch over an octave during conversation (Hudson and Holbrook, 1982). Thus, in everyday life, GPR provides a strong cue to speaker sex in adults (cf. Bachorowski and Owren, 1999), but it provides only a weak cue to speaker size within adult subgroups.

The length and the shape of the vocal tract (VT) causes certain frequencies to be reinforced and others attenuated. The length of the supra-laryngeal VT is highly correlated with speaker height, increasing with age in both sexes (Fitch and Giedd, 1999). The longer the VT, the more the prominent spectral peaks (formants) of speech shift toward lower frequencies (Fant, 1970). Recently, we have shown that small changes in the VTL of vocoded vowels (5%–7%) can be reliably discriminated (Smith *et al.*, 2005; Ives *et al.*, 2005), indicating that speaker size is potentially a perceptually salient aspect of speech. As a child grows between the ages of four and the onset of puberty (around 12), there is a steady increase in VTL with a concomitant decrease in formant frequency. The formant frequencies of adult males decrease by about 30% from their values at age four, while the formant frequencies of adult females decrease by about 20% (Huber *et al.*, 1999). VTL is an important cue to sex and age because it changes with physical size; large adult men have the longest VTLs, children have the shortest VTLs, and adult women have intermediate VTLs (Fitch and Giedd, 1999). The standard deviation for the height of adult men and women is just less than 5%,¹ which is a little less than the just noticeable difference for VTL in adult men. As a result, the correlation between speaker height and formant frequency is predictably weak within the relatively small groups of men or women in published studies (González,

2004; Rendall *et al.*, 2005). Nevertheless, formant-frequency differences clearly distinguish short children from tall adults, and the formant differences can be derived from individual sustained vowels.

In summary, it is clear that GPR and VTL provide potent cues to the main differences in speaker sex and age—that is, whether the speaker is a man, woman, boy, or girl. The question in this study is whether listeners are able to distinguish the sex and age of the original speaker when the study of Smith and Patterson (2005) is rerun with speakers having different sexes and ages.

II. METHOD

Listeners were presented sustained, isolated vowels recorded from four different speakers (adult man and woman, young boy and girl). The vowels were scaled to have the same GPR and VTL values over a large range of GPR and VTL values. Listeners were required to judge whether a boy, girl, man, or woman had spoken each scaled vowel.

A. Stimuli

Examples of the five English vowels (/a:/, /e:/, /i:/, /o:/, /u:/) of an adult man and woman, and a young boy and girl, were recorded using a high-quality microphone (Shure SM58-LCE), with a sampling rate of 48 kHz and a 16-bit amplitude resolution. The vowels were recorded in a sound-attenuating booth to avoid background noise; the microphone was held approximately 5 cm from the chin to maximize the signal to noise ratio. To avoid audible expiration noise, the microphone was held at a point 45° below the horizontal, and the speaker was instructed to pronounce the vowel sounds over, rather than into, the microphone. Speakers were required to utter a series of sustained vowels at a regular relaxed rate (i.e., fifteen /aaaa/ sounds), at a comfortable effort level and at a constant intensity. From these the best five examples were chosen for scaling; that is, vowels were rejected if they had a pitch wobble, jaw articulation noise, lip smacking, or a markedly different pitch from the other examples. For each speaker, five examples of each of the five vowels were selected, giving a total of twenty-five vowel sounds per speaker. Each vowel sound was cut out of the sequence of vowels with care being taken to retain the vowel's natural onset and offset. The age, weight, GPR, height, and estimated VTL (see the following) for each of the four speakers is shown in Table I.

The gain of all the vowel sounds for all speakers was adjusted up or down so that all the vowel sounds had the same rms level prior to scaling. Pilot listening indicated that the vowel sounds had similar loudness.

In order to scale the vowels of the four speakers to the same VTL, it is necessary to estimate the VTL of each speaker. This was done by analyzing the five recorded examples of each of the five vowels /a:/ to /u:/, as spoken by each speaker. The frequencies of formants F_1 to F_3 of each vowel were extracted, as a function of utterance time (formant tracks) using PRAAT (Boersma, 2001).² The values were found to largely agree with those reported in Hillenbrand *et al.* (1995). These formant tracks were fed into a physical

TABLE I. Physical variables for the four speakers.

Speaker	Age (yr)	Weight (kg)	GPR (Hz) ^a	Height (cm)	VTL (cm) ^a	Height (%) ^b	VTL (%) ^b
Man	24	69.6	108	183	17.6	100	100
Woman	41	68	226	175	14.9	96	85
Girl	9	36	239	143	13.2	78	75
Boy	6	22	256	121	12.5	66	71

^aAverage across all vowels.

^bExpressed as a percentage normalized to the value for the adult male speaker.

model of formant production tempered by statistical knowledge of VTL and shape variability, and knowledge concerning the error of measurement. The estimates were calibrated against the MRI estimates of vocal-tract length reported by Fitch and Giedd (1999) and the vowel database of Hillenbrand *et al.* (1995). This model performs a factor analysis with a single latent factor of speaker size (Turner *et al.*, 2004). Figure 1 shows estimates of VTL for each speaker, for each of the five vowels, using this model. The scale factors for the speakers were based on the average across all vowels for that speaker. The Appendix describes the model and a calibration test in more detail.

The final step was to create copies of all of the vowels for a wide range of GPR and VTL values for all of the speakers. The scaling of the vowels was performed by STRAIGHT (Kawahara *et al.*, 1999; Kawahara and Irino, 2004). STRAIGHT is a sophisticated vocoder that uses the classical source-filter theory of speech (Dudley, 1939) to segregate GPR information from the spectral-envelope information associated with the shape and length of the vocal tract. Liu and Kewley-Port (2004) have reviewed STRAIGHT and commented favorably on its ability to manipulate formant-related information. STRAIGHT produces a GPR-independent envelope that accurately tracks the motion of the spectral envelope throughout the utterance. Once STRAIGHT has seg-

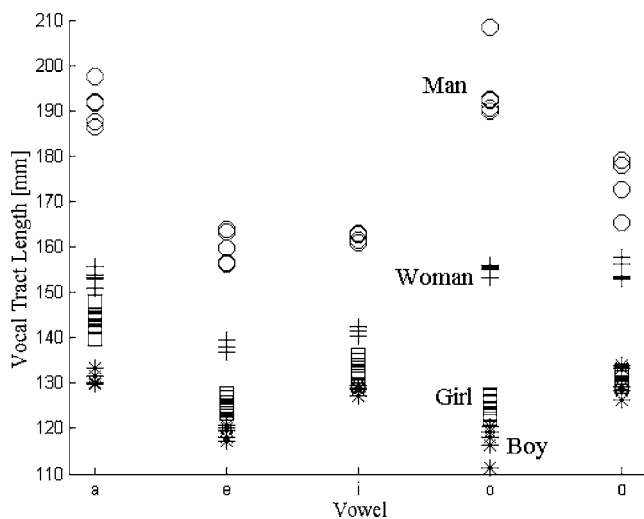


FIG. 1. Estimates of vocal-tract length from formant frequency data using a physical model and a latent variable factor analysis (Turner *et al.*, 2004). At least five examples of each of the vowels /a:/ to /u:/ were analyzed for an adult male (circle), an adult female (plus sign), a young girl (square), and a young boy (asterisk). Details for each of the speakers are provided in Table I.

regated a vowel into a GPR contour and a sequence of spectral-envelope frames, the vowel can be resynthesized with the spectral-envelope dilated or contracted to simulate a change in VTL; the change in VTL for a given scaled vowel is strictly, inversely proportional to the change in spectral-envelope ratio. The GPR dimension (time) can also be expanded or contracted. These operations are largely independent. Utterances recorded from a man can be transformed to sound like a woman or a child. The resynthesized utterances are of high quality even when the speech is resynthesized with GPR and VTL values beyond the normal range of speech (provided the GPR is not much greater than the frequency of the first formant, cf. Smith *et al.*, 2005). The duration of all vowels was adjusted to 850 ms within STRAIGHT, by stretching/expanding the signal without altering the pitch or spectral content. The use of STRAIGHT is described in Kawahara and Irino (2004).

The combinations of GPR and VTL used in the experiment are shown in Fig. 2. The values were chosen to encom-

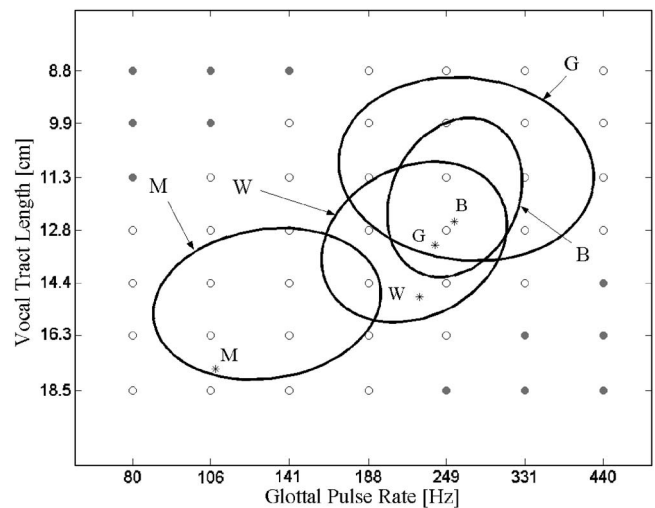


FIG. 2. The open circles show the GPR and VTL combinations of the stimuli used in the sex/age categorization experiment. The GPR values were 80, 106, 141, 188, 249, 331, and 440 Hz. The VTL values were 8.8, 9.9, 11.3, 12.8, 14.4, 16.3, and 18.5 cm. Six conditions in the top-left corner (low GPR and short VTL combinations), and six conditions in the bottom-right corner (high GPR and long VTL combinations), were not presented in the experiment because of their distance from the normal ellipses; these conditions are shown as filled gray circles. The four ellipses show the normal range of GPR and VTL values in speech for men (M), women (W), boys (B), and girls (G), derived from the data of Peterson and Barney (1952). Each ellipse contains 99% of the individuals from the respective category. The asterisks mark the coordinates in the GPR-VTL plane of the four input speakers; man (M), woman (W), boy (B), and girl (G).

pass the range of GPR and VTL encountered in the normal population and to include a large part of the GPR-VTL range employed in [Smith and Patterson \(2005\)](#); GPR varied from 80 to 440 in six, equal, logarithmic steps (seven sample points), and VTL ranged from 18.5 down to 8.8 cm in six, equal, logarithmic steps (seven sample points). The four ellipses show estimates of the normal range of GPR and VTL in speech for men, women, boys, and girls, derived from the [Peterson and Barney \(1952\)](#) vowel database. In each case, the ellipse encompasses 99% of the individuals in the Peterson and Barney data for that category of speaker.³ Six points in the top-left corner (low GPR and short VTL combinations) and six points in the bottom-right corner (high GPR and long VTL combinations) were not presented because these combinations are unusual and we wished to focus the listeners' attention on normal perception as far as possible.

Listeners were seated in a double-walled, IAC, sound-attenuating booth. The stimuli were played by a 24-bit sound card (Audigy 2, Sound Blaster), through a TDT anti-aliasing filter with a sharp cutoff at 10 kHz and a final attenuator (set at -18 dB), and presented diotically to the listener over AKG K240DF headphones. The sound level of the vowels at the headphones was ~ 60 dB SPL. The rms level of the vowels was 0.08 (relative to maximum ± 1).

B. Procedures

The experiments were performed using a single-interval, four-alternative, forced-choice (4AFC) paradigm. The listener heard scaled versions of five stationary English vowels (/a:/, /e:/, /i:/, /o:/, /u:/), and had to make a judgment about the sex/age of the speaker (man, woman, boy, girl). Sex/age judgments were made by selecting the appropriate button on a response box displayed on a monitor in the booth. The level of the vowel was roved in intensity over a 10 (± 5) dB range. Since the judgments are subjective there was no feedback.

A run of judgments consisted of one presentation of each GPR-VTL combination for all five vowels and all four input speakers, presented in a computer-randomized order (a total of 37 GPR-VTL combinations \times 5 vowels \times 4 input speakers, or 740 trials). For each trial, there were five possible examples of the single vowel that could be played (derived from the five examples of each input vowel for each speaker); the example that was presented was determined pseudorandomly by the computer. Each run took approximately 50 min to complete. Each listener completed five runs in three sessions over a week. Ten listeners participated in the experiments, five male and five female. They ranged in age from 20 to 53 years, and were paid volunteers. All had normal absolute thresholds at 0.5, 1, 2, 4, and 8 kHz.

III. RESULTS

The average results for all listeners are presented in [Fig. 3](#), as four groups of four surface plots. Each group of plots shows the probability of assigning one of the *responses* ("boy," "girl," "man," or "woman") to the sustained vowels of the four *speakers* (boy, girl, man, or woman). For the adult speakers, the distribution of sex and age judgments is very

similar across the GPR-VTL plane; that is, the sex of the speaker (man or woman) has relatively little effect on the judgments. Compare the plots for the man and the woman in the bottom row of each judgment group. Similarly, for the children, the distribution of sex and age judgments is largely unaffected by the sex of the speaker (boy or girl); compare the plots for the young boy and girl in the top row of each judgment group. It is also the case that the distribution of "man" responses (bottom-left judgment group) and the distribution of "girl" responses (top-right judgment group) are similar for all four speakers. The effects of original speaker appear mainly in the distribution of "woman" responses (bottom-right judgment group) and the distribution of "boy" responses (top-left judgment group), and they are largely associated with the age (or size) of the speaker. Compare the upper row (boy or girl speaker) with the lower row (man or woman speaker) in the top-left, and bottom-right, judgment groups. With regard to the main experimental question, the results show that when GPR and VTL are controlled, there remains at least one additional cue to the origin of the speaker in sustained vowels, and that cue is more closely associated with the age or size of the speaker (adult *versus* juvenile) than the sex of the speaker (male *versus* female).

Details of the Sex and Age Judgments. The two-dimensional (2D) surface plots in [Fig. 3](#), for the speaker sex and age judgments ("boy," "girl," "man," or "woman"), were constructed as follows: The responses were averaged over the five vowels and all ten listeners, since the pattern of responses was similar for all of the vowels and all of the listeners. Each *group* of four panels shows the probability of the listener assigning the response "boy," "girl," "woman," or "man" (as noted by the group header) to the stimulus vowel, as a function of GPR and VTL. The probability of classification is shown by color, ranging from 0 (dark-blue) meaning "never classified" to 1 (brown-red) meaning "always classified." Within each panel the abscissa is GPR and the ordinate is VTL, both on logarithmic axes. The open circles show the combinations of GPR and VTL presented to the listeners; between these data points, the surfaces have been generated by interpolation. The combinations in the top-left and bottom-right corners of the GPR-VTL plane were not presented because they rarely occur in the population, and as a result, they are omitted from the interpolated surface. The dotted black lines outline regions of the GPR-VTL plane where listeners consistently chose one category of response out of the four available to them. Within these regions, the probability of choosing the given combination of sex and age is greater than 0.5. The four ellipses show estimates of the normal range of GPR and VTL combinations in speech sounds for men, women, boys, and girls ([Peterson and Barney, 1952](#)), where each ellipse contains 99% of the individuals from the respective category.

There is one other aspect of the data to note before proceeding to detailed statistical analyses of the results, and that is the listeners' use of the response categories ("man," "woman," "boy," and "girl"), which differs considerably from the distribution of GPR-VTL combinations in the population. In the "man" response group (bottom-left group [Fig. 3](#)), the boundary beyond which listeners do not use the

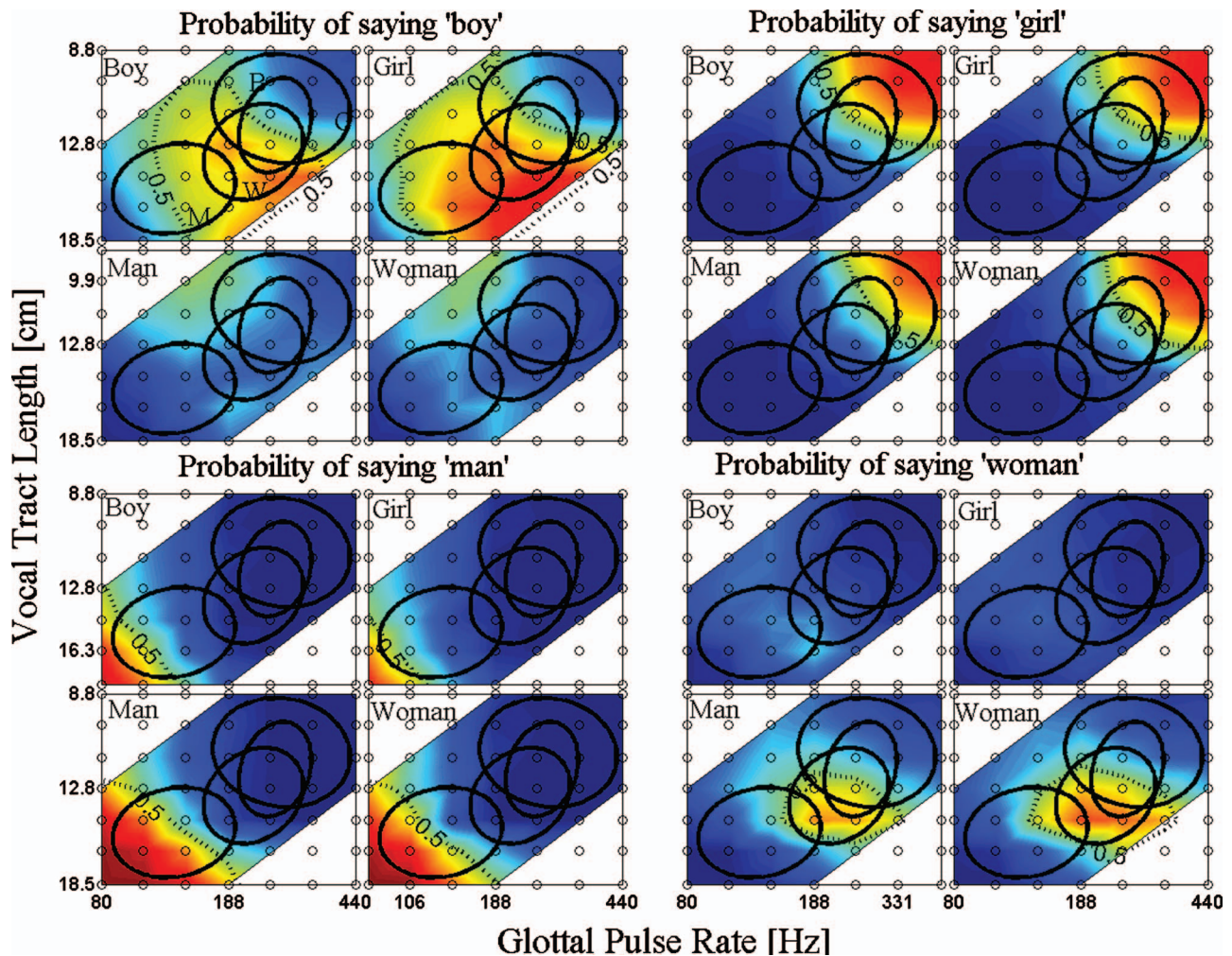


FIG. 3. Sex and age judgments for the four different speakers (young boy, young girl, adult man, and adult woman) averaged over all listeners ($n=10$). The data are presented as a series of 2D surface plots with color showing probability of assigning a given GPR-VTL combination to one of four perceptual categories (boy, girl, man, or woman). The probability of classification is shown by color, ranging from 0 (dark-blue) meaning “never classified” to 1 (brown-red) meaning “always classified.” The data for each perceptual response are shown separately as a group of four panels, where each panel is for a different speaker; top-left quadrant (probability of saying “boy”), top-right quadrant (probability of saying “girl”), bottom-left quadrant (probability of saying “man”), and bottom-right quadrant (probability of saying “woman”). The points in the plane where sex/age judgments were measured are shown by the open circles in each panel; between the data points the surface was derived by interpolation. Within each panel, GPR-VTL combinations in the top-left (low GPR and short VTL) corner and the bottom-right (high GPR and long VTL) corner were not presented. For each GPR-VTL combination, the probabilities from the four panels for a given speaker sum to 1. (Imagine four separate 2D maps stacked vertically and aligned over each other). The data were averaged across all five vowels and all ten listeners, so each sample-point probability is based on 250 trials. The dotted black contour line marks classification threshold, that is, a probability ≥ 0.5 of consistently choosing one category out of the four available. The region of GPR-VTL values enclosed by this line defines a region categorized as one particular sex or age. The four ellipses show the range of GPR and VTL in speech for men (M), women (W), boys (B), and girls (G), as derived from the data set of Peterson and Barney (1952).

“man” response is quite abrupt for each original speaker, and the boundary runs across the man ellipse of normal speakers at a point well short of where it might be expected to occur. If listeners were matching to the distribution of speakers in the population, the boundary might be expected to occur near the midline of the intersection of the man and woman ellipses. Similarly, in the “girl” response maps (top-right group Fig. 3), the boundary beyond which listeners do not use the “girl” response is quite abrupt, and it runs across the girl ellipse at a point short of the midline of the intersection of the girl and woman ellipses. So the listeners are not using the distribution of the GPR-VTL combinations in the population to assign their responses. They probably know, at some level, that the distributions for boys and girls overlap, but they assume that the experimenter wants them to be consistent in

their use of the responses, and so it does not occur to them that they might distribute their responses probabilistically, in accordance with the overlap in the distributions. As a result, they fail to assign the response “girl” to many combinations of GPR and VTL that might well be produced by girls. It is also the case that vowels with combinations of low GPR and long VTL are all perceived as coming from a large, or very large, person, even when the correct response is “girl,” “boy,” etc. Similarly, vowels with combinations of high GPR and short VTL are all perceived as coming from a small, or very small, person, independent of the original speaker. So, in these corners of the space, the size aspect of the perception is often at odds with the “correct” response. The listeners were aware that the vowels of the original speakers had all been scaled to all combinations of GPR and VTL.

The situation is different, however, in the central part of the space, where the response is typically “woman” or “boy.” Here, the speakers are heard as having sizes within the normal range for humans and there is nothing unusual about them. In this region, the listeners under use the categories “man” and “girl” somewhat, and they overuse the categories “woman” and “boy” somewhat, perhaps because they are, at some level, aware of overusing the “man” and “girl” response for the extreme combinations of GPR and VTL in the corners of the space. In any event, in the central region of the space, listeners use categories that differ in both size and sex (“woman” and “boy”), and there is sufficient ambiguity in the perception of the speaker to allow us to assess the relative effect of the size and sex of the original speaker on the perception. The statistical analyses were designed to quantify the main effects, and to determine whether the statistics confirm that listeners have relatively good information about whether the original speaker was an adult or a child, and at the same time, relatively poor information about the sex of the original speaker.

Three statistical analyses of the data were performed: First, there was an analysis to determine the spatial similarity of the response distributions (“boy,” “girl,” “man,” “woman”) between pairs of original speakers (1. *Quantifying the effects of GPR and VTL in sex and age judgments*). Then the details of the speaker effects were explored (2. *Details of speaker effect in sex and age judgments*). Finally, the effects of speaker sex and size were investigated (3. *Role of speaker size and speaker sex in judgments of sex and age*).

1. *Quantifying the Effects of GPR and VTL in Sex and Age Judgments*. The main effects of GPR and VTL on the distribution of sex and age judgments are shown in Fig. 3. We will begin with the distribution of “man” responses in the bottom-left quadrant; it is similar for all four speakers, inasmuch as vowels with a low GPR and long VTL tend to be categorized as being spoken by a man. The ellipse for adult men shows that this is the natural category to adopt for vowels scaled to these combinations of GPR and VTL. This result replicates Smith and Patterson (2005), who also found that sustained vowels in this region of the GPR-VTL plane are reported as being spoken by men. In Smith and Patterson (2005), the vowels were scaled from a single, adult-male speaker. The present data show that sustained vowels with a low GPR and a long VTL are categorized as being spoken by an adult man, regardless of the source, although the vowels from the boy and girl speakers have to be scaled to more extreme GPR and VTL values than those from the man or woman speakers to produce the same probability of “man” response.

Each of the sixteen panels in Fig. 3 defines a surface of perceptual probability in the GPR-VTL plane. A nonparametric test of spatial association (cf. Ramsden *et al.*, 1999) was used to compare pairs of maps to test whether there is significant overlap. The perceptual maps consist of 37 probability values, each of which represents the judgments of all ten listeners to all five vowels, for a given combination of GPR and VTL. Thus, for each pair of maps compared (map₁ and map₂), there are 37 pairs of probabilities on which to base the comparison. Before comparison, the probabilities were

subjected to a hard threshold, such that any point with probability $p \geq 0.5$ was classified as ON; otherwise it was classified as OFF. This results in two maps with binary values which can be compared point for point, using a simple procedure. A four-cell contingency table is constructed which counts corresponding points in each pair of maps which are (i) both ON, (ii) both OFF, (iii) ON in map₁ and OFF in map₂, and (iv) OFF in map₁ and ON in map₂. Each successful match increments the appropriate cell by one. The quantization of the data reduces the power of the test somewhat, but the purpose was just to distinguish the effect of age/size from the effect of sex and, for this purpose, the strong test with conservative criteria is entirely appropriate. “Details of speaker effect in sex and age judgments” to follow deals with the attempt to quantify more subtle effects.

For each pair of maps, the null hypothesis (H_0) that there is *no spatial association* between the two maps is tested. To test if the two maps are spatially associated, we calculate χ^2 from the contingency table using Yates’ correction for small cell counts. If the χ^2 value exceeds the critical value for the specified significance level, for one degree of freedom, then we can reject the null hypothesis, and conclude that there is a spatial association between the two maps. The degree of spatial association can be described by the degree of association in the contingency table, known as the contingency coefficient c . It ranges from 0 (meaning the two maps are not correlated at all), to a maximum (meaning the two maps are completely superimposed); the maximum is determined by the number of rows and columns in the contingency table, and for the current measure, the maximum is 0.707 ($1/\sqrt{2}$).

For the response “man,” the null hypothesis that there is no association between the perceptual maps generated by different original speakers can be rejected in the majority of cases, with $p < 0.001$ ($P(\text{man}|\text{b})$ vs $P(\text{man}|\text{g})$, $P(\text{man}|\text{b})$ vs $P(\text{man}|\text{m})$, $P(\text{man}|\text{b})$ vs $P(\text{man}|\text{w})$, $P(\text{man}|\text{m})$ vs $P(\text{man}|\text{w})$). A significant p value indicates the absence of a significant difference between two maps. $P(\text{man}|\text{b})$ refers to the conditional probability of responding “man,” defined as the matrix of probability responses across the GPR-VTL plane, *given* vowels scaled from a boy speaker. A conservative Bonferroni correction was adopted for the alpha value required for significance, since the four speaker maps were compared in six pair-wise tests; the resulting p value for significance is 0.008 (0.05/6). Using this correction, we *cannot* reject the null hypothesis (that there is *no* association between the response distributions) for two comparisons, the girl speaker *versus* the man speaker [$P(\text{man}|\text{g})$ vs $P(\text{man}|\text{m})$ is n.s.], and the girl speaker *versus* the woman speaker [$P(\text{man}|\text{g})$ vs $P(\text{man}|\text{w})$ is n.s.]. The “man” response distribution for vowels scaled from the *girl* speaker is different from the “man” response distributions for vowels scaled from the two adult speakers. Table II shows the results for all of the perceptual-map comparisons. For all speakers except the girl, the distribution of “man” responses seems to be largely determined by the combination of GPR and VTL of the vowel, rather than the sex and age of the original speaker.

The perceptual maps associated with different speakers are also very similar for the “girl” response (top-right quad-

TABLE II. χ^2 test of significance on data collapsed across all five vowels and ten listeners. The first two columns denote the maps being compared; the third column presents the χ^2 value; the fourth column, the significance (or not); and the fifth column, the contingency coefficient. Note that the p value for significance is taken to be 0.008 (=0.05/6); this adopts a conservative Bonferroni correction to compensate for the comparison of four maps in six pair-wise tests for each perceptual response category.

Map ₁	Map ₂	χ^2	p	c
P(boy b)	P(boy g)	9.7097	n.s. (0.01)	0.46
P(boy b)	P(boy m)	0.0646	n.s.	0.04
P(boy b)	P(boy w)	0.0646	n.s.	0.04
P(boy g)	P(boy m)	0.0191	n.s.	0.02
P(boy g)	P(boy w)	0.0191	n.s.	0.02
P(boy m)	P(boy w)	8.7432	n.s. (0.01)	0.44
P(girl b)	P(girl g)	27.4075	<0.001	0.65
P(girl b)	P(girl m)	27.4075	<0.001	0.65
P(girl b)	P(girl w)	27.4075	<0.001	0.65
P(girl g)	P(girl m)	22.4971	<0.001	0.61
P(girl g)	P(girl w)	22.4971	<0.001	0.61
P(girl m)	P(girl w)	31.7669	<0.001	0.68
P(man b)	P(man g)	16.7733	0.001	0.56
P(man b)	P(man m)	15.1715	0.001	0.54
P(man b)	P(man w)	15.1715	0.001	0.54
P(man g)	P(man m)	8.3159	n.s. (0.01)	0.43
P(man g)	P(man w)	8.3159	n.s. (0.01)	0.43
P(man m)	P(man w)	32.1033	<0.001	0.68
P(woman b)	P(woman g)	a	a	a
P(woman b)	P(woman m)	a	a	a
P(woman b)	P(woman w)	a	a	a
P(woman g)	P(woman m)	a	a	a
P(woman g)	P(woman w)	a	a	a
P(woman m)	P(woman w)	19.0421	0.001	0.58

^aThere are very few “woman” responses for the boy and girl speakers, so in the marked cases the comparison is between two essentially flat planes of no response. It is meaningless to report χ^2 and contingency coefficient values in such cases.

rant group of Fig. 3). Vowels with combinations of high GPRs and short VTLs, which appear in the upper-right corner of each of the GPR-VTL planes, are consistently categorized as being spoken by girls, and there is little effect of original speaker upon the distribution of girl responses. This corner contains the ellipse for girls and the ellipse for boys, but the ellipse for girls extends to higher GPRs and shorter VTLs, so it is arguably the natural category to adopt for the extreme values of GPR and VTL. This result was also reported in Smith and Patterson (2005). We can reject the null hypothesis that there is no spatial association between perceptual maps for the response “girl” (P(girl|b) vs P(girl|g) etc.) with $p < 0.001$ (cf. Table II).

However, the null hypothesis for speaker effects cannot be rejected in the case when the response is “boy” or “woman;” the distributions in the upper row are different from those in the lower row for both of these response groups. We will return to this below.

The perceptual maps are very similar for the man and woman speakers (bottom row of each judgment group of Fig. 3). The null hypothesis, that there is no association between corresponding perceptual maps for the man and woman speakers, can be rejected with $p < 0.001$ for all perceptual

categories except P(boy|m) vs P(boy|w) which is n.s. (0.01) following Bonferroni correction (cf. Table II bottom row of each perceptual group).

The perceptual maps are also very similar for the boy and girl speakers (top row of each perceptual response group of Fig. 3). The null hypothesis, that there is no association between corresponding perceptual maps for the boy and girl, can be rejected with $p < 0.001$ for all perceptual categories except P(boy|b) vs P(boy|g) which is n.s. (0.01) following Bonferroni correction (cf. Table II top row of each perceptual group).

In summary, the statistics support the effects observed in Fig. 3; although GPR and VTL have a major effect on the perception of speaker sex and age, they are not the sole determinants of sex and age discriminations based on sustained vowels. Specifically, it is hard to make the sustained vowels of adults sound like those of a boy (top-left quadrant group of Fig. 3), and it is hard to make the sustained vowels of children sound like those of a woman (bottom-right quadrant group of Fig. 3).

2. Details of Speaker Effect in Sex and Age Judgments.

The statistical test described above measures *spatial association* between the two perceptual maps. This is a strong test which not only requires that the two maps have similar spread across the GPR-VTL plane, but also requires that the maps are aligned spatially with each other. However, it performs this test globally, which has drawbacks if we wish to investigate more subtle differences between two distributions.

The effect of original speaker upon the judgments can be revealed by comparing the volume of response enclosed by the dotted-line in each panel, across the sixteen panels (Fig. 3), i.e., summing over all sample points with a “probability of response” value ≥ 0.5 . Summing only over points with a response ≥ 0.5 means our analysis is focused on parts of the GPR-VTL plane where responses are strong, and where we can be fairly confident about the category to which listeners ascribe the scaled vowel. We do not sum over the *entire* surface within the thresholded region, but only at the sampled points; the response volume is thus discretely rather than continuously defined. This measure gives an accurate and fairer assessment than visual comparisons, because interpolation between sample points in the 2D surface plots sometimes gives a falsely heightened visual impression of the similarities between plots. Table III tabulates these perceptual volumes as a function of the four original speakers. Differences between perceptual responses, as measured by response volume, reveal the effect of speaker upon sex and age judgments. These effects are secondary to those produced by GPR and VTL inasmuch as the simple measure of response volume does not take into consideration the 2D spatial association of the map. Differences between response volumes across subjects can be tested statistically using the t-test for repeated samples.

P(boy): The girl speaker is more likely to elicit the response “boy” than is the boy speaker [$t(9) = -5.028$, $p < 0.001$, two-tailed, P(boy|b) vs P(boy|g)]. The p value is less than the Bonferroni-corrected significance level of 0.008 (=0.05/6). Obvious differences in “boy” response volume

TABLE III. Response volumes (expressed as percentage of maximum of 37), averaged across all listeners ($n=10$), with standard error of mean (percentage). P(b) means probability of responding “boy” over the sampled GPR-VTL plane, P(g) means probability of responding “girl,” etc.

	Boy speaker				Girl speaker				Man speaker				Woman speaker			
	P(b)	P(g)	P(m)	P(w)	P(b)	P(g)	P(m)	P(w)	P(b)	P(g)	P(m)	P(w)	P(b)	P(g)	P(m)	P(w)
x	30.86	21.28	12.99	3.70	41.11	19.78	10.12	2.30	8.41	16.89	26.19	15.66	7.05	20.63	22.00	18.13
s	4.96	2.76	2.59	2.05	4.35	2.78	2.52	1.65	2.66	1.52	1.82	3.36	2.26	1.58	2.08	3.71

between juvenile and adult speakers are subsequently discussed in the paragraph entitled “ $P(\text{child})$ vs $P(\text{adult})$.” There are no significant differences between adult speakers ($P(\text{boy}|\text{m})$ vs $P(\text{boy}|\text{w})$).

$P(\text{girl})$: There are no statistical differences in the volumes of the regions that elicit the response “girl” across the different speakers.

$P(\text{woman})$: There are no statistical differences in the volume of the regions that elicit the response “woman” across the different speakers. Obvious differences in “woman” response volume between juvenile and adult speakers are subsequently discussed in the paragraph entitled “ $P(\text{child})$ vs $P(\text{adult})$.”

$P(\text{man})$: The volumes for vowels that elicit the response “man” are greatest for the adult man and woman speakers, and they are substantially smaller for the boy and girl speakers. The difference between adults is significant [$t(9) = 4.701$, $p < 0.001$, two-tailed, $P(\text{man}|\text{m})$ vs $P(\text{man}|\text{w})$], with the adult man speaker being more likely to elicit the response “man” than the woman speaker. The difference between children and adults for $P[\text{man}]$ is significant [$t(9) = -5.992$, $p < 0.001$, two-tailed, $\{P(\text{man}|\text{b}) + P(\text{man}|\text{g})\}$ vs $\{P(\text{man}|\text{m}) + P(\text{man}|\text{w})\}$], with adult speakers being more likely to elicit an adult response (“man” or “woman”) than juvenile speakers.

$P(\text{child})$ vs $P(\text{adult})$: It is possible to scale the vowels of juvenile speakers so that they elicit the response “man” (bottom-left Fig. 3), and to scale the vowels of adult speakers so that they elicit the response “girl” (top-right Fig. 3), provided the vowels are scaled to extreme GPR and VTL values. However, for the “boy” and “woman” responses (top-left and bottom-right groups in Fig. 3, respectively), juvenile and adult vowels cannot be equivalently scaled. To quantify this, we compared the response volumes $P(\text{child})$ vs $P(\text{adult})$, where $P(\text{child})$ is defined as $P(\text{boy}) + P(\text{girl})$, and $P(\text{adult})$ is defined as $P(\text{man}) + P(\text{woman})$, given vowels scaled from juvenile and adult speakers. The null hypothesis is that there is no difference between response volumes $P(\text{child})$ and $P(\text{adult})$ for juvenile and adult speakers. We can reject this for $P(\text{child}|\text{child})$ vs $P(\text{child}|\text{adult})$ at $p < 0.001$ ($t(9) = 5.988$, two-tailed), and for $P(\text{adult}|\text{child})$ vs $P(\text{adult}|\text{adult})$ at $p < 0.001$ ($t(9) = -5.561$, two-tailed). Accordingly, it is reasonable to conclude that the response volumes for $P(\text{child})$ and $P(\text{adult})$ differ.

3. Role of Speaker Size and Speaker Sex in Judgments of Sex and Age. The purpose of the following is to explicitly test the hypothesis that speaker size has greater power than speaker sex in explaining the distributions of responses. To test the power of speaker size as an explanatory factor, we collapsed the volume-of-response values (sample points

≥ 0.5) across sex (boy and girl speakers versus man and woman speakers), and tested the significance across the four perceptual responses. To test the power of speaker sex as an explanatory factor, we collapsed the volume-of-response values across size (boy and man speakers versus girl and woman speakers), and tested the significance across the four perceptual responses. We tested this using a one-way analysis of variance (ANOVA), where the independent categorical variable was speaker type and the single dependent variable was volume of response.

First, we performed a simple one-way ANOVA with original speaker as four categories of the independent variable, thereby conflating sex and size as factors. The results across the four different perceptual responses are: $P(\text{boy})$ $F(3, 36) = 20.4$, $p < 0.001$, $\eta^2 = 0.63$; $P(\text{girl})$ $F(3, 36) = 0.74$, $p = \text{n.s.}$, $\eta^2 = 0.06$; $P(\text{man})$ $F(3, 36) = 10.99$, $p < 0.001$, $\eta^2 = 0.48$; $P(\text{woman})$ $F(3, 36) = 8.23$, $p < 0.001$, $\eta^2 = 0.41$. The correlation ratio, η^2 , is a measure of the size of the effect. It is calculated as a proportion of the between sum of squares (i.e., that part of the variation in the data attributable to the independent variable), to the total sum of squares (i.e., that part of the variation in the data attributable to the independent variable plus all other factors), giving $\eta^2 = (SS_{\text{bet}}/SS_{\text{tot}})$. Having original speaker as four categories of the independent variable, thereby conflating sex and size as factors, accounts for between $\sim 40\%$ and 65% of the variance in the data when the data are expressed as response volumes.

For speaker size as a factor, we performed a one-way ANOVA with speaker collapsed over sex, i.e., boy and girl speaker versus man and woman speaker. The results across the four different perceptual responses are: $P(\text{boy})$ $F(1, 18) = 30.66$, $p < 0.001$, $\eta^2 = 0.63$; $P(\text{girl})$ $F(1, 18) = 0.34$, $p = \text{n.s.}$, $\eta^2 = 0.02$; $P(\text{man})$ $F(1, 18) = 15.99$, $p < 0.001$, $\eta^2 = 0.47$; $P(\text{woman})$ $F(1, 18) = 12.59$, $p = 0.002$, $\eta^2 = 0.41$. Size as an independent variable (disregarding sex) still accounts for between $\sim 40\%$ and 65% of the variance in the data expressed as response volumes.

For speaker sex as a factor, we performed a one-way ANOVA with speaker collapsed over size, i.e., boy and man speaker versus girl and woman speaker. The results across the four perceptual responses are: $P(\text{boy})$ $F(1, 18) = 1.16$, $p = \text{n.s.}$, $\eta^2 = 0.06$; $P(\text{girl})$ $F(1, 18) = 0.25$, $p = \text{n.s.}$, $\eta^2 = 0.01$; $P(\text{man})$ $F(1, 18) = 1.61$, $p = \text{n.s.}$, $\eta^2 = 0.08$; $P(\text{woman})$ $F(1, 18) = 0.03$, $p = \text{n.s.}$, $\eta^2 = 0.00$. Sex as an independent variable (disregarding size) does not have the power necessary to explain any differences in the data expressed as response volumes.

In summary, speaker size has greater power than speaker sex in accounting for the variance in our data. This can be appreciated informally by inspecting the figure; if one merges, by eye, the upper and lower panels by *column* for each response group in Fig. 3, then the remaining two panels in each response group would be very similar. Having collapsed over size, comparison between sexes shows little difference. Alternatively, if one merges, by eye, the upper and lower panels in each response group by *row*, the remaining two panels in each response group are very different. Having collapsed over sex, the comparison between sizes reveals substantial differences.

IV. DISCUSSION

Previous research intended to identify the acoustic properties of male and female voices responsible for the perception of gender has used a variety of methods such as statistical clustering and perceptual categorization (e.g., Childers and Wu, 1991; Coleman, 1976; Whiteside, 1998; Wu and Childers, 1991; Bachorowski and Owren, 1999; Schwartz and Rine, 1968; Ingemann, 1968; Lass *et al.*, 1976). The general conclusion from these studies is that the main acoustic variables affecting perception of gender are GPR and VTL, although the relative importance of the two factors is moot in previous research. Motivated by recent work on the perception of auditory size (Smith *et al.*, 2005; Ives *et al.*, 2005; Turner *et al.*, 2006), Smith and Patterson (2005) measured the interaction of GPR and VTL in judgments of age/size, as compared with judgements of sex. In that study, however, all of the “different” (different GPR and VTL combinations) speakers were created from vowels recorded from a single, adult male speaker. The current study was intended to determine how the pattern of response observed in Smith and Patterson (2005) would vary if the original speaker were an adult woman, a young boy, or a young girl, as opposed to an adult man,

The distributions of sex and age judgments across the GPR-VTL plane are very similar for the man and woman speakers (bottom row of each judgment group of Fig. 3), and separately, very similar for the boy and girl speakers (top row of each judgment group of Fig. 3). The “man” distributions (bottom-left quadrant of Fig. 3) and the “girl” distributions (top-right quadrant of Fig. 3) are largely independent of the sex and age of the original speaker. There are, however, clear differences in the use of the “boy” response (top-left quadrant group of Fig. 3) and the use of the “women” response (bottom-right quadrant group of Fig. 3). The results suggest that speaker *size* is more important than speaker *sex* in determining whether vowels are judged to come from a boy, girl, man, or woman (cf. the last section of Sec. III).

In a recent paper, Assmann *et al.* (2006) presented perceptual judgments of voice gender where the speech of an adult male and an adult female speaker had been manipulated to have the same fundamental and formant frequencies. They found that adult female voices can be made to sound like adult male voices and vice versa. However, they did not use children’s voices so there are no data concerning the child/adult distinction.

After controlling for GPR and VTL, there remain several interesting effects of original speaker. First, the girl speaker is more likely to be assigned the response “boy” than is the boy speaker, especially for the more extreme values of GPR and VTL. In point of fact, the girl was a little taller and older than the boy, and her GPR and VTL values (239 Hz, 13.2 cm) indicate a slightly larger person than those of the boy (256 Hz, 12.5 cm). The results of the current experiment suggest that, at least for sustained vowels, it is largely size that determines how the voice of a juvenile will be heard. In the current experiment, when listeners can hear that the voice comes from a juvenile speaker, and there is no contextual information to indicate the sex of the speaker, they assign the response “boy” to voices with longer vocal tracts and lower pitches and “girl” to shorter vocal tracts and higher pitches. It might well be that this is a general bias in the population responses. Moreover, it may be the case that, in the absence of contextual information, there is a bias in the perception of speaker sex, with listeners actually hearing the voices of larger children as boys and the voices of smaller children as girls. However, it would take further experiments to confirm such a hypothesis.

Second, the results show that it is hard to make the sustained vowels of children sound enough like a woman to elicit the response “woman,” yet if the vowels of children are scaled to extreme values of GPR and VTL, as occurs in the bottom-left corner of the GPR-VTL plane, they are assigned the response “man.” Informal listening indicates that this is a true perceptual effect and that the extreme GPR and VTL values override more subtle cues in this region, causing us to conclude not only that the voice comes from a very large person, but also that the person is male. Again, however, it would take further experiments to confirm such a hypothesis. Note, however, that the vowels of the young boy and girl have to be driven to more extreme GPR and VTL values than those of the adult speakers before listeners assign them the response “man.”

Beyond GPR and VTL. Scaling all of the voices from one original speaker, as in Smith and Patterson (2005), meant that certain characteristics of the vowels were fixed regardless of VTL. This is probably not typical of the population of human voices as a whole (Fant, 1966, 1975; Diehl *et al.*, 1996). There is an important anatomical difference between children and adults; children have proportionately larger heads relative to their body size than do adults. As a result, the ratio of oral cavity length to pharyngeal cavity length is greater in children than it is in adults (Fant, 1966). Figure 4 shows how oral and pharyngeal length grow as a proportion of VTL as children mature into adults (redrawn from Turner *et al.*, 2004). Since oral/pharyngeal length ratio (OPR) changes markedly with age, it is reasonable to hypothesize that the changes in OPR might produce changes in formant ratios that could account for the effects of size observed in our data; that is, the formant ratios for a given vowel might be somewhat different in children and adults.

Accordingly, we calculated the $F2/F1$ and $F3/F1$ formant ratios for the sustained vowels of our four speakers; the average ratios are plotted, as a function of speaker height, in Fig. 5. The $F2/F1$ ratio increases marginally as speaker

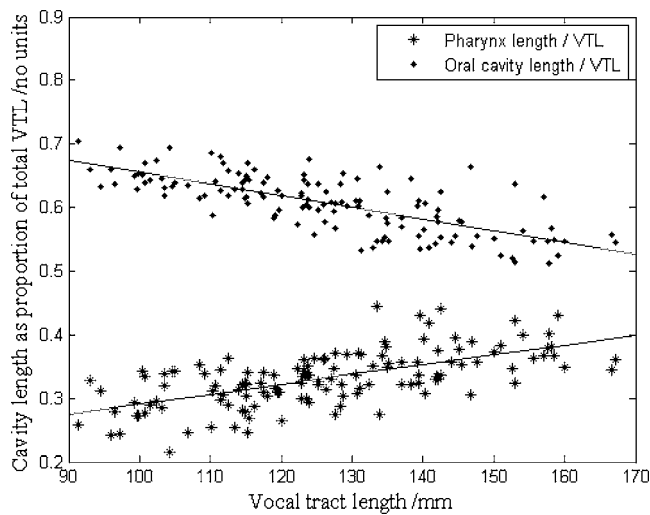


FIG. 4. Oral and pharyngeal cavity lengths, expressed as a proportion of total VTL, and plotted as a function of speaker VTL (redrawn from Turner *et al.*, 2004). The cavity lengths were derived from MRI measurements made by Fitch and Giedd (1999). The lines show the best-fitting linear regression.

height increases while the $F3/F1$ ratio increases somewhat more. The error bars show the standard deviations across vowel; the standard deviations for the individual vowels are much smaller—on the order of the size of the symbols for the means. Since it is not obvious which measure of variability determines discrimination performance, and since there are only four individuals in this study, the data have to be interpreted with caution, but if the results are representative, it is clear that the change in the $F2/F1$ ratio is far less than would be predicted by the nonuniform growth of the OPR, and that it would be difficult to use the information given the variability. It seems more likely that, in attempting to preserve vowel identity, humans vary the position of the tongue constriction to counteract the potential effects of nonlinear growth, and so maintain the $F2/F1$ ratio for individual vow-

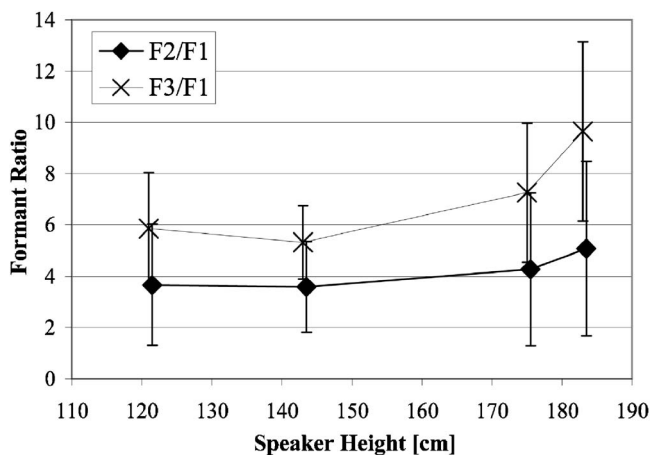


FIG. 5. $F2/F1$ and $F3/F1$ formant ratios of the vowels of the four speakers in the study. The speakers are boy, girl, woman, and man in ascending height along the abscissa. The formant ratios are the average of all five examples of each of the five vowels for each speaker. The formant frequencies were calculated from the middle 10% of each formant track, as extracted by PRAAT (Boersma, 2001). The error bars show ± 1 s.d. across vowels for each speaker.

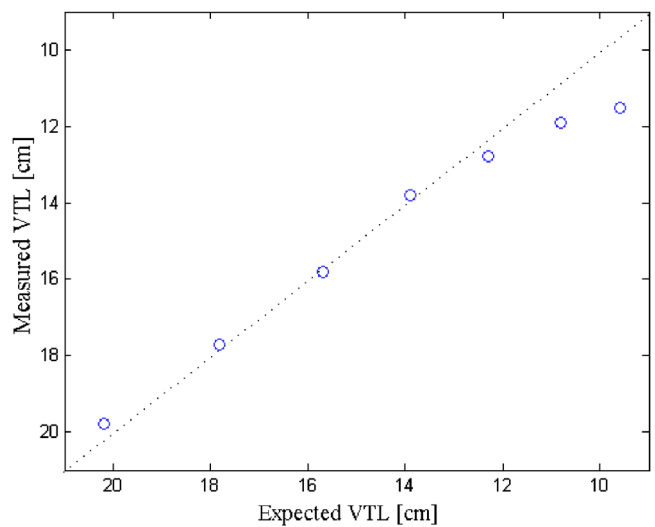


FIG. 6. (Color online) Results of a calibration test involving the scaling applied by STRAIGHT and the physical model (cf. Turner *et al.*, 2004) used to derive VTL from a vowel sound.

els. It is also possible that there is a limit to the adaptation process, and that it is not possible to preserve both the $F2/F1$ ratio and the $F3/F1$ ratio, which would account for the larger changes in $F3/F1$ with age. In any event, it seems that it is more likely that it is the $F3/F1$ ratio that is the basis of the age effect observed in our data, and if this is the result of the nonuniform growth of OPR the connection is complex.

Finally, it is important to remember that the research discussed in this paper deals with acoustic information about speaker sex and age in *sustained vowels*, rather than in sentences or running speech. It may well be that other cues in natural speech, such as increased articulation, would lead to a somewhat different pattern of results.

V. SUMMARY AND CONCLUSIONS

Listeners were presented isolated, sustained vowels recorded from four different speakers (young boy and girl, adult man and woman). The vowels were scaled to produce the same range of GPR and VTL values in each case. Listeners were required to discriminate whether the original speaker was a boy, girl, man, or woman. The results show that, for adult speakers, the distribution of responses across the GPR-VTL plane is largely independent of the sex and the age of the original speaker (top row in each group of Fig. 3). The results also show that, for juvenile speakers, the distribution of responses across the GPR-VTL plane is largely independent of the sex and age of the original speaker (bottom row in each group of Fig. 3). Where differences exist, as in the distribution of “boy” responses (top-left response group of Fig. 3), and “woman” responses (bottom-right response group of Fig. 3), they arise when scaling from adult speakers to juvenile speakers and vice versa. The results show that listeners readily distinguish whether the original speaker was a child or an adult, based on their sustained vowels, but they find it difficult to distinguish the sex of the original speaker.

ACKNOWLEDGMENTS

This research was supported by the UK Medical Research Council (G9900369 and G0500221), the German Volkswagen Foundation (VWF 1/79 783), and the European Office of Aerospace Research and Development (FA8655-05-1-3043).

APPENDIX

Estimates of vocal-tract length were derived from vowel sounds using a physical model and latent variable factor analysis (Turner *et al.*, 2004; cf. Fig. 1 this paper). These estimates were subsequently used to scale different-sized speakers with STRAIGHT (Kawahara *et al.*, 1999; Kawahara and Irino, 2004) to produce matching GPR and VTL values for the four speakers. We report an additional calibration test that feeds the scaled values after manipulation in STRAIGHT back into our physical model to gain an estimate of VTL for the new scaled vowels (Fig. 6). VTL values are plotted as expected VTL (abscissa) versus measured VTL (ordinate). Expected VTL is the value we expect after manipulation in STRAIGHT. Measured VTL is the value measured by our physical model (Turner *et al.*, 2004). For the most part, values fall along the positive diagonal, indicating that VTL values have been correctly scaled. Errors arise for very short VTLs; these are attributable to errors in the ability to accurately extract formant frequencies from these sounds.

¹Health Survey for England 2004, representative sample of 2436 adult men and 3311 adult women. Average height adult men 1750 mm with a standard deviation of 89.93 mm, and average height adult women 1612 mm with a standard deviation of 69.05 mm [<http://www.ic.nhs.uk/pubs/hlthsvyeng2004upd>] (last viewed 17 September 2007).

²Boersma, P., and Weenik, D. (2005). "Praat: doing phonetics by computer" (Version 4.4.30) [Computer program from <http://www.praat.org/>] (last viewed 17 September 2007).

³The set of formant values for each of the 76 speakers in the classic study of Peterson and Barney (1952) were converted to VTL values using the VTL data that Fitch and Giedd (1999) extracted from magnetic resonance images of a large population of subjects. Each ellipse represents the mean ± 3 s.d. for each category of speaker. The calibration details are presented in Turner *et al.* (2004); this poster can be found at http://www.pdn.cam.ac.uk/groups/cnbh/research/posters_talks/BSA2004/TWPBSA04.pdf. (last viewed 17 September 2007).

Assmann, P. F., Dembling, S., and Nearey, T. M. (2006). "Effects of frequency shifts on perceived naturalness and gender information in speech," in *Proceedings of the Ninth International Conference on Spoken Language Processing*, Pittsburgh, PA.

Assmann, P. F., and Neary, T. M. (2003). "Frequency shifts and vowel identification," in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain.

Bachorowski, J., and Owren, M. J. (1999). "Acoustic correlates of talker sex and individual talker sex identity are present in a short vowel segment produced in running speech," *J. Acoust. Soc. Am.* **106**, 1054–1063.

Beckford, N. S., Rood, S. R., and Schaid, D. (1985). "Androgen stimulation and laryngeal development," *Ann. Otol. Rhinol. Laryngol.* **94**, 634–640.

Boersma, P. (2001). "Praat, a system for doing phonetics by computer," *Glot International* **5**, 9/10, 341–345.

Childers, D. G., and Wu, K. (1991). "Gender recognition from speech. II Fine analysis," *J. Acoust. Soc. Am.* **90**, 1841–1856.

Coleman, R. O. (1976). "A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice," *J. Speech Hear. Res.* **19**, 168–180.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex* (Murray, London).

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (2003). "Effects of

fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," *J. Acoust. Soc. Am.* **114**, 2913–2922.

Diehl, R. L., Lindholm, 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.

Dudley, H. (1939). "Remaking speech," *J. Acoust. Soc. Am.* **11**, 169–177.

Fant, G. (1966). "A note on vocal tract size factors and non-uniform F-pattern scalings," *STL-QPSR* **4**, 22–30.

Fant, G. (1970). *Acoustic Theory of Speech Production*, 2nd ed. (Mouton, Paris).

Fant, G. (1975). "Non-uniform vowel normalization," *STL-QPSR* **2-3**, 1–19.

Fitch, W. T., and Giedd, J. (1999). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," *J. Acoust. Soc. Am.* **106**, 1511–1522.

González, J. (2004). "Formant frequencies and body size of speaker: A weak relationship in adult humans," *J. Phonetics* **32**, 277–287.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.

Hollien, H., Green, R., and Massey, K. (1994). "Longitudinal research on adolescent voice change in males," *J. Acoust. Soc. Am.* **96**, 3099–3111.

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T., and Johnson, K. (1999). "Formants of children, women and men: The effects of vocal intensity variation," *J. Acoust. Soc. Am.* **106**, 1532–1542.

Hudson, A., and Holbrook, A. (1982). "Fundamental frequency characteristics of young black adults: Spontaneous speaking and oral reading," *J. Speech Hear. Res.* **25**, 25–28.

Ingemann, F. (1968). "Identification of the speaker's sex from voiceless fricatives," *J. Acoust. Soc. Am.* **44**, 1142–1144.

Ives, D. T., Smith, D. R. R., and Patterson, R. D. (2005). "Discrimination of speaker size from syllable phrases," *J. Acoust. Soc. Am.* **118**, 3816–3822.

Kawahara, H., and Irino, T. (2004). "Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Boston, MA), pp. 167–180.

Kawahara, H., Masuda-Kasuse, I., and de Cheveigne, A. (1999). "Restructuring speech representations using pitch-adaptive time-frequency smoothing and instantaneous-frequency-based F0 extraction: Possible role of repetitive structure in sounds," *Speech Commun.* **27**, 187–207.

Künzel, H. J. (1989). "How well does average fundamental frequency correlate with speaker height and weight?," *Phonetica* **46**, 117–125.

Lass, N. J., and Brown, W. S. (1978). "Correlational study of speakers' heights, weights, body surface areas and speaking fundamental frequencies," *J. Acoust. Soc. Am.* **63**, 1218–1220.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., and Bourne, V. T. (1976). "Speaker sex identification from voiced, whispered, and filtered isolated vowels," *J. Acoust. Soc. Am.* **59**, 675–678.

Liu, C., and Kewley-Port, D. (2004). "STRAIGHT: A new speech synthesizer for vowel formant discrimination," *ARLO* **5**, 31–36.

Morton, E. S. (1977). "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds," *Am. Nat.* **111**, 855–869.

Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.

Ramsden, B. M., Hung, C. P., and Roe, A. W. (1999). "Activation of illusory contour domains in macaque area V2 is accompanied by relative suppression of real contour domains in area V1," *Abstr. Soc. Neurosci.* **25**, 2060.

Rendall, D., Vokey, J. R., Nemeth, C., and Ney, C. (2005). "Reliable but weak voice-formant cues to body size in men but not women," *J. Acoust. Soc. Am.* **117**, 2372.

Schwartz, M. F., and Rine, H. E. (1968). "Identification of speaker sex from isolated, whispered vowels," *J. Acoust. Soc. Am.* **44**, 1736–1737.

Smith, D. R. R., and Patterson, R. D. (2005). "The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age," *J. Acoust. Soc. Am.* **118**, 3177–3186.

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.

Smith, D. R. R., Walters, T. C., and Patterson, R. D. (2007a). "Role of glottal-pulse rate, vocal-tract length and original talker upon judgements of speaker sex and age," *J. Acoust. Soc. Am.* **121**, 3135–3136.

Smith, D. R. R., Walters, T. C., and Patterson, R. D. (2007b). "Judging sex and age: Effect of glottal-pulse rate, vocal-tract length and original

- speaker," in *Proceedings of the 19th International Congress on Acoustics ICA2007*, Madrid, Spain. Special issue *Revista de Acústica* **38**, PPA-090-010.
- Smith, D. R. R., Walters, T. C., Walland, K., and Patterson, R. D. (2006). "The role of input speaker upon judgements of speaker sex and age," paper presented at British Society of Audiology, Cambridge, p. 43.
- Titze, I. R. (1989). "Physiologic and acoustic differences between male and female voices," *J. Acoust. Soc. Am.* **85**, 1699–1707.
- Turner, R. E., Al-Hames, M. A., Smith, D. R. R., Kawahara, H., Irino, T., and Patterson, R. D. (2006). "Vowel normalisation: Time-domain processing of the internal dynamics of speech," in *Dynamics of Speech Production and Perception*, edited by P. Divenyi, S. Greenberg, and G. Meyer (IOS Press, Amsterdam), pp. 153–170.
- Turner, R. E., Walters, T. C., and Patterson, R. D. (2004). "Estimating vocal tract length from formant frequency data using a physical model and a latent variable factor analysis," paper presented at British Society of Audiology, UCL, London, p. 61, http://www.pdn.cam.ac.uk/groups/cnbh/research/posters_talks/BSA2004/TWPBSA04.pdf (Last accessed 9/17/07).
- Whiteside, S. P. (1998). "Identification of a speaker's sex from synthesized vowels," *Percept. Mot. Skills* **86**, 595–600.
- Wu, K., and Childers, D. G. (1991). "Gender recognition from speech. P I. Coarse analysis," *J. Acoust. Soc. Am.* **90**, 1828–1840.

Perceptual thresholds for detecting modifications applied to the acoustical properties of a violin

Claudia Fritz^{a)} and Ian Cross

Centre for Music and Science, Music Faculty, University of Cambridge, West Road, Cambridge CB3 9DP, United Kingdom

Brian C. J. Moore

Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

Jim Woodhouse

Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

(Received 1 November 2006; revised 9 August 2007; accepted 17 September 2007)

This study is the first step in the psychoacoustic exploration of perceptual differences between the sounds of different violins. A method was used which enabled the same performance to be replayed on different “virtual violins,” so that the relationships between acoustical characteristics of violins and perceived qualities could be explored. Recordings of real performances were made using a bridge-mounted force transducer, giving an accurate representation of the signal from the violin string. These were then played through filters corresponding to the admittance curves of different violins. Initially, limits of listener performance in detecting changes in acoustical characteristics were characterized. These consisted of shifts in frequency or increases in amplitude of single modes or frequency bands that have been proposed previously to be significant in the perception of violin sound quality. Thresholds were significantly lower for musically trained than for nontrained subjects but were not significantly affected by the violin used as a baseline. Thresholds for the musicians typically ranged from 3 to 6 dB for amplitude changes and 1.5%–20% for frequency changes. Interpretation of the results using excitation patterns showed that thresholds for the best subjects were quite well predicted by a multichannel model based on optimal processing.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799475]

PACS number(s): 43.66.Lj, 43.75.Cd, 43.75.De, 43.66.Jh [DD]

Pages: 3640–3650

I. INTRODUCTION

There is an extensive literature on the acoustics of the violin, and an even more extensive literature on human perception of sounds in general, and of musical sounds in particular. However, there is virtually no published research on the combined problem of the human capability for perception, discrimination, and judgment of the sounds of violins with particular measurable acoustical properties. This is a very significant gap, since perceptual judgments must define what makes a violin different from other bowed-string instruments, and one violin different from another. To the best of the authors' knowledge this paper represents the first attempt to apply rigorous psychoacoustical techniques to a question of this nature.

The ultimate aim underlying the research presented here is to answer the typical question that a violin maker will ask: “Why does this violin sound better than this one,” or more specifically “What will happen to the sound if I change such-and-such a constructional detail?” This paper starts the process of attacking that broad aim with a more modest target—to establish the just-noticeable difference for certain particular acoustical changes to the mechanical frequency re-

sponse of a violin. These changes all relate to quantities previously proposed as significant to the sound quality of a violin. This initial investigation is of admittedly limited scope but is already of interest to instrument makers, telling them, for example, how far they need to move an individual low body resonance to have an audible effect. Since a violinist will explore a very wide range of bowing as well as types of musical input to a violin when judging its quality, the underlying philosophy of this study is to seek the input which results in the lowest perceptual threshold for each given acoustical change.

There are two stages necessary to such a study—to relate a constructional change to an acoustical change, i.e., a mechanical change of the vibrational properties, and to evaluate the perceptual effect of that change. There is already a significant literature concerned with the first stage (Cremer, 1985; Durup and Jansson, 2005). The experiments reported here concentrate on the second stage, that of establishing quantitative links between acoustical parameters of the instrument body and the perceptions of a listener.

The methodology of the study relies on the large impedance jump between the strings and the bridge of the instrument. The player manipulates the string to vibrate in certain ways, the vibrating string applies a force to the bridge, and the body vibrates in response to this force, radiating sound to

^{a)}Electronic mail: cf291@cam.ac.uk

the listener. To a first approximation, the body motion has little backward influence on the string motion. There are exceptions—most obviously the wolf note (Cremer, 1985; Woodhouse, 1993). More generally, if the topic of interest were the “playability” of the violin rather than its sound, then it would certainly not be admissible to ignore this back reaction. Similarly, if the study were concerned with the guitar or the piano, then string/body coupling would be crucial because it determines the decay rates of the various overtones of the string motion. However, for a bowed string such coupling effects can be ignored in the first instance. If strings of the same type are fitted to two different violins, a skilled player will adjust bowing to coerce the vibration into the standard Helmholtz motion with an acceptably short transient (Guettler and Askenfelt, 1997). The force waveforms acting at the bridge in the two cases will be very similar, and one would expect that the major differences in sound between the two instruments could be captured by driving them both with identical forcing waveforms. With this in view, representative force waveforms can be recorded using normal playing on a violin whose bridge is equipped with a piezoelectric force sensor under each string. These predetermined force functions can then be applied to different violins, so that sound differences can be compared with no complications arising from variations in playing.

Such a test could be carried out using different physical violins, applying the force at the bridge with a vibration shaker of some kind. However, for this study a different approach was taken. The mechanical frequency response function of the violin was mimicked using a digital filter, and the output signal for listening tests was generated by applying this filter to the recorded bridge force signal. Once the violin response is represented in digital filter form, it becomes very easy to make controlled variations of a kind which would be virtually impossible to achieve by physical changes to a violin.

Langhoff *et al.* (1995) conducted experiments in which violin performances were filtered digitally in a similar way. However, they equipped the violin with a velocity sensor on the bridge and could not accurately derive the force signal from the velocity signal. They used one violin as a baseline and then modified its frequency response curve (and therefore its impulse response) in several ways, to give enhancement of the Helmholtz resonance and of midrange frequencies (around 1.7 kHz), and creation of a smoother decay toward higher frequencies. This experiment did show that it is possible to compare violin spectra by listening to digitally filtered signals but it did not address the question of how people perceived the different sounds created. Langhoff *et al.* (1995) only report the “subjective impressions” of one of the authors and no other participant was involved. Here we report the results of psychoacoustic measures of the ability of musically and nonmusically trained subjects to discriminate changes in frequency and amplitude of single and multiple resonances.

II. STIMULI

A. Generation principle

In order to create the stimuli, we first recorded input signals (i.e., the force applied by the bowed strings on the bridge) during a live performance on a violin whose bridge was instrumented with a piezoelectric force sensor under each string. Second, we measured a suitable frequency response function for the two chosen violins (a modern instrument of good quality made by David Rubio, and a student-quality instrument used for comparison in the final stage of the work).

The choice of frequency response function raises some important technical issues. The most natural choice would be some kind of pressure response measured by a microphone, in response to force applied to the violin bridge. However this choice would be unsuitable for two reasons. First, there is the question of variation with microphone position and the influence of room acoustics: No single measurement can be regarded as giving a “typical” sound. More serious is a theoretical issue. The essence of the experimental methodology is to make controlled changes in the frequency response function to generate the set of stimuli for testing. To make such changes, an explicit mathematical formula is needed for the frequency response in question in which the parameters to be varied appear explicitly. No such formula is known for radiated sound from a complex structure like a violin body. However, such a formula, in terms of mode shapes and natural frequencies, is available for a frequency response describing structural vibration rather than sound radiation (e.g., Skudrzyk, 1981). To take advantage of this, we have chosen to work with a mechanical frequency response, the input admittance function of the violin: This is defined as the ratio between the velocity at the string position on the bridge and the force applied at the same point. This function governs the energy transfer from string to body, and it is the most appropriate structural response for the present purpose. Of course, this raises important questions about whether different results might be obtained using radiated-sound response functions: These questions are currently being considered in ongoing research. The approach used here can be considered as being comparable to listening to a stationary violin with one ear from a fixed position in an unchanging room. These synthesized sounds are realistic enough for listeners to clearly recognize the sound as coming from a violin [for sound examples see Fritz (2007)] and for us to investigate relative changes.

The measurement procedure for input admittance was standard (e.g., Jansson, 1997): The bridge was excited with a miniature force hammer (PCB 086D80) at the G-string corner, and at the E-string corner velocity was measured using a laser-Doppler vibrometer (Polytec OFV056/OFV3001). This calibrated input admittance was then processed by modal identification techniques (e.g., Ewins, 2000), and resynthesized from the fitted parameters in the frequency range up to 7000 Hz, to allow parametric modification. To cover this range, 54 modes were needed for the Rubio violin. The frequency response could be modified by manipulating the modal parameters (amplitude, frequency, and Q factor), and

then used to construct finite impulse response (FIR) digital filters to obtain the “sound” of modified virtual violins. The filtering (i.e., the convolution of the input signal with the inverse Fourier transform of the frequency response of the violin in the time domain) was carried out using MATLAB. It was found that the phase of the frequency response had no audible influence on the output sound, so for improved noise performance, noncausal zero-phase filters were used.

B. Acoustical modifications to be tested

Informal tests showed only a very slight perceptual influence of the Q factors of the major modes, for changes up to 40%. Hence, it was decided to limit this initial study to measurements of thresholds for detecting increases in amplitude and frequency of one or several modes of the original admittance function, as described in the following.

1. Modes A0, B1–, and B1+

At low frequencies, the sound of a violin is dominated by three strongly radiating modes. Several authors have suggested that these modes are important for sound quality (e.g., Hutchins, 1962). A0 is a modified Helmholtz resonance (“air mode”), which usually falls around 280 Hz. The two other modes are “plate modes,” which arise primarily from the bending and stretching of the front and back plates: B1– is usually centered in the range 470–490 Hz and B1+ between 530 and 570 Hz. Collectively B1– and B1+ account for what early researchers on the violin called the “main body resonance.”

In the admittance, the A0 mode has a very small amplitude compared to the other low modes, B1– and B1+. However, it plays a very significant role in the radiated sound. To represent this effect approximately, its amplitude was artificially increased by a factor of 5 (14 dB), so that its amplitude was similar to that of B1–, as observed in radiated sound measurement in far field (e.g., Dünwald, 1991).

The amplitude and the frequency of each of these three modes were altered individually, as illustrated in Fig. 1 for an increase of frequency for B1+.

2. All modes in each of the four Dünwald bands

Based on his measurement of the acoustical properties of a large range of violins that had previously been classified as of very good or moderate quality, Dünwald (1991) proposed four frequency bands, which he suggested were important for the judgment of sound quality: 190–650, 650–1300, 1300–4200, and 4200–6400 Hz. The first range includes the lower overtones and may be related to “richness,” the second he associated with “nasality,” the third with “brilliance,” and the fourth with “clarity.” We measured detection thresholds for a change, either in amplitude or in frequency, of all the modes within each of these bands. For a shift in amplitude, as a few modes were close to the boundary between two bands, it was decided, instead of using a rectangular pass-band, to apply a gain function with a flat top and sloping edges (half of a symmetric Hanning window, of total width equal to 100 Hz, outside the nominal range of the band), as illustrated in Fig. 2.

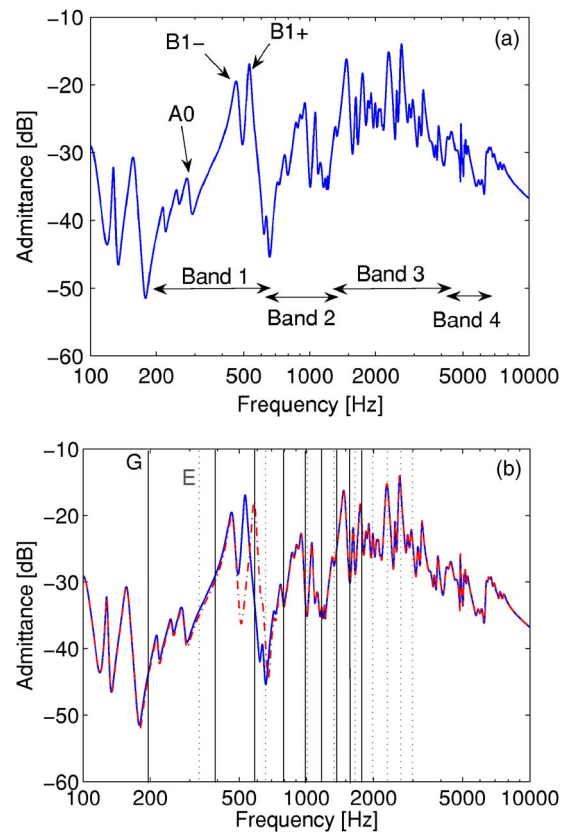


FIG. 1. (a) (Color online) The resynthesized admittance of the Rubio violin, indicating the modes and the Dünwald bands which were modified; (b) an example of a shift in frequency of a single mode: Mode B1+ of the original admittance (solid curve) is shifted upwards by 14% (dash-dot curve). The positions of the harmonics are shown for each note by the solid vertical lines for G and dotted lines for E.

For the frequency modification, the frequencies of all the modes within a given band were shifted by the same factor. However, it should be noted that, for a shift larger than about 20%, the modified modes overlap so much with the modes of the adjacent upper band that the change becomes very artificial.

3. All modes

In a further modification, the frequencies of all modes were simultaneously shifted by the same factor. This simulates, roughly, a change in the size of the violin body. This modification was not done in amplitude as this would only simulate a violin played harder, not a change in the violin body. It would in any case be nullified by the equalization of loudness, to be described shortly.

C. Input signal

In a preliminary study (Fritz *et al.*, 2006), performances of single notes and short phrases were played through filter sets corresponding to the measured frequency response curves of three violins, found from blind listening tests to be of very different sound quality. The phrases and single notes were tested for both discrimination and preference. Results showed that all listeners had lower thresholds for single notes than for musical phrases. Since the purpose of the

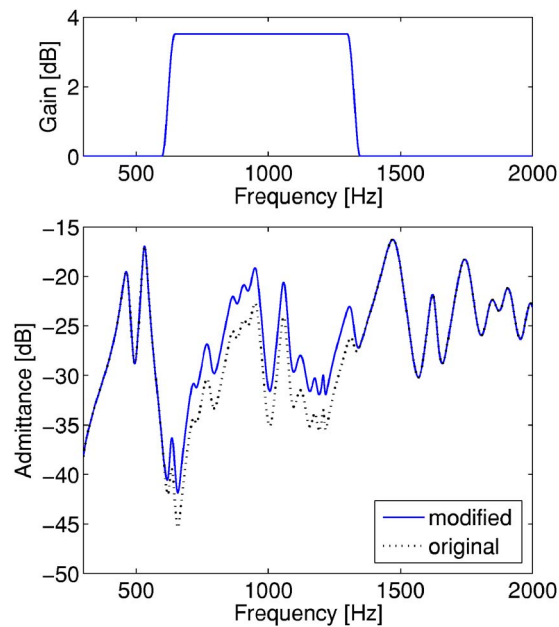


FIG. 2. (Color online) Illustration of the method for modifying the amplitude of a single Dünwald band. The gain function shown in the top panel applied to the original admittance function (dashed curve in the bottom panel) gives the admittance function illustrated by the solid curve in the bottom panel: All the modes of the second Dünwald band are shifted in level by 3.5 dB (amplitude increased by 50%).

present tests was to establish the thresholds for discrimination of changes which are the lowest that can be achieved under optimal test conditions, it was decided to use two single notes: G3 at 196 Hz and E4 at 330 Hz. The choice of these two notes results from the distribution of their harmonics. In particular, G3 has its second and third harmonics close to the center frequencies of modes B1⁻ and B1⁺, whereas E4 has no harmonics near these modes. We wished to assess whether this would lead to poorer discrimination of changes in B1⁻ and B1⁺ when E4 was used. The duration of the single notes was chosen to be relatively short (300 ms) because of echoic memory effects (see Sec. III A). With such short notes, it is very difficult to tell how the violin was bowed and, thus, the influence of the way of bowing is reduced.

D. Control of loudness

Large modifications of the modes, in particular of their amplitude, can lead to a change of loudness. One procedure for removing loudness cues is to randomly vary (rove) the overall level from one stimulus to the next. However, to completely eliminate such cues, a large rove range is required (Green, 1988), and this would have made the experience of listening to the violin sounds very unnatural. As an alternative method for ensuring that subjects would discriminate between the sounds on the basis of their spectral shape and not of their loudness, the overall level of each sound file was adjusted to keep the loudness level approximately constant at a value of 93 phons over the whole range of modifications. For each sound file (corresponding to a certain modification), this was achieved by first calculating the maximum value, M , over the duration of the sound of the

short-term average loudness level, which was updated every millisecond, using the loudness model developed by Glasberg and Moore (2002). The overall loudness level did not change much over the duration of the sound so working with the maximum was similar to working with the mean. The sound amplitude was then multiplied by $10^{(93-M)/20}$. The result was that each sound file used in the experiment had a calculated loudness level of 93 ± 1 phons. This may appear loud but it should be remembered that the sounds were broadband, and such sounds have a greater loudness than a narrow-band sound of the same level (Moore, 2003). The sound levels used were below those typically experienced by violinists when playing (Royster *et al.*, 1991).

III. EXPERIMENT

A. Procedure

Thresholds were estimated using a three-alternative forced-choice procedure. A three-down one-up adaptive tracking rule was used which estimated the 79% correct point on the psychometric function (Levitt, 1971). Three sounds—two the same (the reference violin sound), one different (the modified violin sound)—were played in a sequence, and the subject was asked to choose which one was different. In order to allow echoic memory to operate effectively (Darwin *et al.*, 1972), the sounds and the interstimulus intervals were each 300 ms in duration. The amount of modification (either in frequency or in linear amplitude) between the stimuli in a given trial was changed by a certain factor (step size). Eight turnpoints were obtained. A relatively large initial step size of $2^{1/2}$ was applied until the second turnpoint was reached, in order to allow rapid convergence toward the threshold region. After the second turnpoint, the step size was reduced to $2^{1/4}$. Threshold was taken as the mean of the values of the amount of modification at the last six turnpoints.

For trials involving frequency shifts, the modification of the single modes was done by moving symmetrically from the original frequency; for instance, a shift of 10% was achieved by moving the center frequency of the mode by +5% and by -5%. This was done to reduce incidences of a shifted mode merging with an adjacent higher or lower mode, especially given that B1⁻ and B1⁺ are quite close. During the listening test, the reference sound was thus not kept constant. In contrast, for the amplitude test, the reference sound was kept constant and equal to the sound of the original violin, in order to increase discriminability (subjects “learnt” to recognize the reference sound, as it was always the same) and the modification was an upwards shift in amplitude. This last method was used for all the other modifications.

Subjects were given visual feedback during the experiment but did not get any practice or any training beforehand. However, if they performed erratically on the first run, they were asked to do the run again. This applied only to a few subjects, and they were all able to perform the task at the second attempt. However, a few subjects had difficulties in performing certain conditions in the middle of the experiment and were given the opportunity to retry such conditions

up to three times. Some succeeded, but not all. The conditions which could not be performed were not always the same among the subjects and this problem could happen even for subjects with low thresholds for other conditions. To avoid this problem, the initial amount of modification would have needed to be significantly larger. However, increasing all the initial values because of one or two people would have considerably increased the duration of the test for the others, and so a compromise was chosen.

The sounds were presented diotically via Sennheiser HD580 headphones, chosen because of their diffuse field response, in a relatively quiet environment. The sampling rate was 44 100 Hz and the number of bits was 16.

B. Subjects

Three groups each of 18 subjects were selected according to their musical background. The first group had relatively little musical training (less than 6 years of formal training) and did not practice regularly: This group will be termed “nonmusicians” in the following. The two other groups both had considerable musical training (more than 8 years of formal training) and practiced at least weekly. These last two groups were differentiated according to the instrument played: The violinists, viola players, and cellists were in one group, and the remaining musicians in the other. Fifty subjects were between 18 and 40 years old, and the four others were between 50 and 60 years old. All subjects reported having normal hearing, although this was not checked. No systematic effect of age was observed in the results. Subjects were paid for their participation. These subjects were available to undertake the tests involving modifications in frequency and amplitude of the Dünwald bands as well as the modification in amplitude of the single modes.

A different group of subjects was employed for the tests of frequency modification of the three single modes A0, B1+, and B1-. This group consisted of 9 nonmusicians, 18 string players, and 17 other musicians.

IV. RESULTS

The threshold results for all tests are summarized in Figs. 3 and 4. They are presented only for two categories of subjects, as initial analyses of variance (ANOVAs) did not show any significant difference between string players and other musicians, so the results for these two groups were combined. The average thresholds were calculated as the geometric mean since the standard deviation of the thresholds across subjects tended to increase with the mean value of the threshold. The variability of results among subjects was large, as is evident in some large error bars, representing ± 1 s.d. This variability was probably partly due to a lack of training. Individual differences could also be observed in the standard deviation of the turnpoints within a run. For the best subjects, the tracking variable went down directly to the region of the threshold and then oscillated closely around it, which gave a low standard deviation. For other subjects (with higher thresholds), the tracking variable fluctuated much more, giving a large standard deviation. Some people

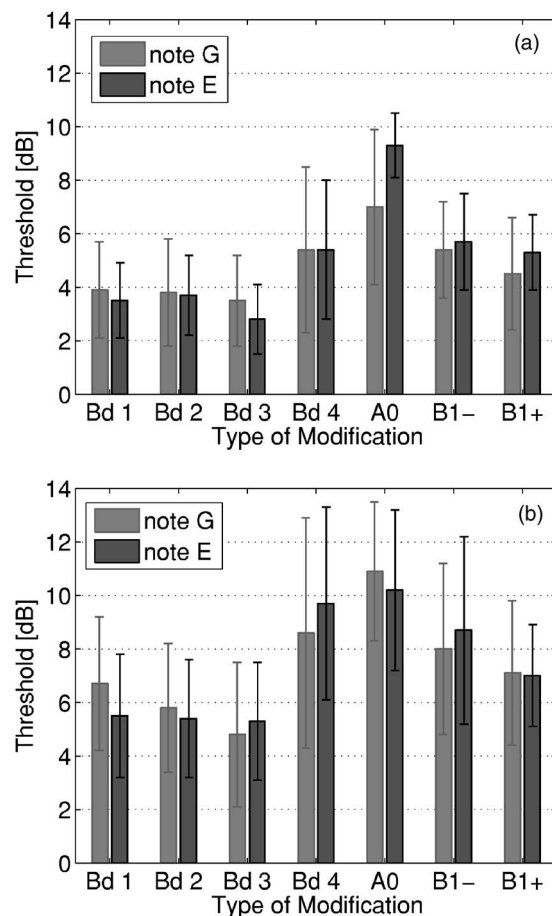


FIG. 3. Mean thresholds for detecting a modification in amplitude, expressed in decibels. The upper (a) and lower (b) panels show results for musicians and nonmusicians, respectively. The type of modification is indicated by the label under each pair of bars. Light and dark bars show results for the notes G and E, respectively. Error bars represent ± 1 s.d. across subjects for each category of subject.

had a much lower threshold than average for one modification and a much higher threshold for another.

In the following, first the mean results will be presented and discussed. Then, to reduce the effect of the lack of training, the thresholds of the five best subjects will be presented and interpreted. We propose that these thresholds are close to the best that can be achieved, representing the limits of perceptual performance.

A. General comments on the statistical analysis

ANOVAs were run separately for the amplitude and the frequency modifications. Moreover, separate ANOVAs were performed for each type of modification, i.e., for the three single modes and for the four bands. The ANOVA based on data for the frequency modification of each of the four bands also incorporated, as an additional condition, the simultaneous modification of all modes.

For each of the analyses, the type of modification (the band or the single mode which was modified) and the note (G or E) were used as the dependent variables in a mixed (one between, two within) repeated-measures ANOVA. The between-group variable, the subject type, had three values, nonmusicians, musicians, and string players. The first of the

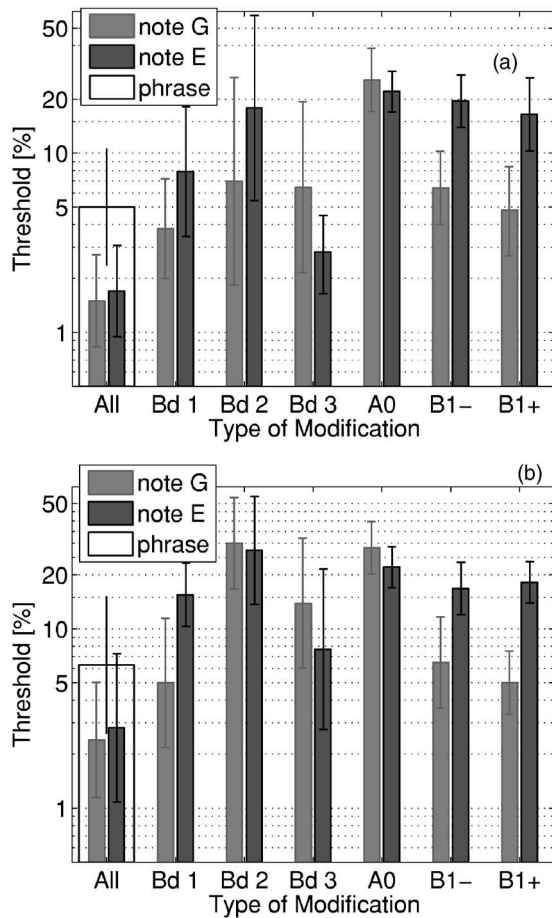


FIG. 4. As in Fig. 3, but showing thresholds [(a) for musicians and (b) for nonmusicians] for detecting a modification in frequency, expressed as a percentage of the center frequency and plotted on a logarithmic scale. The large open bars show thresholds when a musical phrase was used rather than a single note.

two within-group variables, the modification, had either three or four values, while the second, the note, had two. In cases where the condition of sphericity was violated, we report values using the Huynh-Feldt correction.

The variance was not homogeneous for the modification in frequency, either for the Dünwald bands or for the single modes. Therefore, the ANOVAs were conducted using the logarithms of the thresholds, which kept the violation of homogeneity at a level where ANOVA was still reasonably robust (for a given within variable, the ratio of the standard deviations across the three between-groups did not exceed 4).

For each of the four analyses, up to three subjects were sometimes removed as their data were incomplete: Either they could not finish in the arranged time or could not do one of the tasks, even after several attempts.

B. Influence of subjects' musical training

As mentioned earlier, the ANOVAs showed no significant difference between string players and other musicians, so these two groups are treated as a single group for subsequent analyses. However, musicians performed significantly

better than nonmusicians (all p values less than 0.001), except for the test involving modification in frequency of the single modes [$F(2, 38)=1.7, p=0.19$].

C. Thresholds

1. Modification in amplitude

Thresholds for both categories of subjects and for the two single notes G and E are given in Fig. 3. The x axis represents the various conditions—amplitude modification of all modes of the i th Dünwald band “Bd i ” or of one single mode A0, B1–, or B1+. The mean thresholds range from about 3 dB (musicians, modification of band 3, note E) to over 10 dB (nonmusicians, modification of A0, both notes).

The ANOVA of the results for the bands showed a main effect of group, with musicians performing significantly better than nonmusicians [$F(1, 50)=18.5, p<0.001$]. There was a main effect of modification [$F(2, 100)=6.3, p=0.003$] but no main effect of note [$F(1, 50)=2.6, p=0.110$]. There was no significant interaction between modification and note [$p>0.1$] but there was a significant interaction among group, modification, and note [$F(2, 100)=3.6, p=0.03$]. This three-way interaction arises from the fact that nonmusicians consistently performed poorly (compared to musicians) for each band with the note E, while musicians' performance with that note was better for band 3 than for either band 1 or band 2.

In the analysis of results for modes, there was a main effect of modification [$F(2, 98)=83.9, p<0.001$], and a main effect of note [$F(1, 49)=4.7, p=0.035$]. There was no significant interaction between modification and note [$p>0.1$] but there was a significant interaction between note and group [$F(1, 49)=14.8, p<0.5$]. Nonmusicians performed consistently poorly for both notes whereas musicians performed rather better (1 dB) for note G than for note E.

2. Modification in frequency

Figure 4 shows corresponding results to Fig. 3 for the frequency-modification tests. A shift in frequency of the modes in Dünwald band 4 was not detectable at all, so no threshold is given for this case nor is it included in the analyses to follow. “All” means that all modes were shifted in frequency. For this particular modification a 2-s-long musical phrase (the first two notes of the third theme of the Glazunov *Concerto for violin in A minor op.82*) was also tested (with the interstimulus interval kept to 300 ms) and the threshold obtained (also shown in Fig. 4) confirms what was shown in the preliminary study: People are less sensitive to subtle changes with a musical phrase than with single notes. This is consistent with the finding that the threshold for detecting a change in center frequency of a single formant in speech sounds (a formant corresponds to a resonance in the vocal tract) is higher when sentences are used than when isolated vowels are used (Liu and Kewley-Port, 2004).

Remember that thresholds above 20% have no real meaning (see Sec. II B 2). A high threshold usually indicates that the corresponding modification was not perceptible.

The ANOVA for bands was based on the results obtained for Dünwald bands 1, 2, and 3 and for the case when all modes were shifted simultaneously. There was a main effect

of modification [$F(2.0, 97.7)=118.1$, $\epsilon=0.67$, $p<0.001$], and a main effect of note [$F(1, 49)=6.2$, $p=0.016$]. There was a significant interaction between modification and note [$F(2.4, 118.4)=26.4$, $\epsilon=0.81$, $p<0.001$] due to the fact that the average performance for musicians and nonmusicians was better for G than E when all modes were shifted and for bands 1 and 2, but this pattern reversed for band 3. There was also a significant interaction among note, modification, and group [$F(2.4, 118.4)=4.9$, $\epsilon=0.81$, $p=0.006$]. Musicians and nonmusicians exhibited the same pattern of responses with respect to note when all modes were shifted and for bands 1 and 3, but not for band 2, for which musicians performed better for G than for E, while the results for nonmusicians showed a slight trend in the opposite direction.

The ANOVA of the results for single modes showed a significant main effect of modification [$F(2, 78)=130.4$, $p<0.001$] and a main effect of note [$F(1, 39)=124.7$, $p<0.001$]. There was a significant interaction between modification and note [$F(2, 78)=134.3$, $p<0.001$], reflecting the fact that performance was better for note G than for note E for modes B1- and B1+ but not for mode A0.

The significant influence of note on the thresholds can be explained as follows. Note G has more harmonics in Dünwald bands 1 and 2 than E, and G has its second and third harmonics close to B1- and B1+, which makes a slight change in the frequency of the corresponding modes much more noticeable for note G than for note E. Regarding the manipulation of band 3, it is not clear at first why there is such a difference between thresholds for E and G. However, some insight can be gained by calculation of excitation patterns, which can be defined as the relative response of the auditory filters plotted as a function of the filter center frequency (Moore and Glasberg, 1983). Excitation patterns calculated according to the procedure described by Glasberg and Moore (2002) (see Fig. 5) show that the difference between excitation patterns for the modified sound and the reference sound is much larger for E than for G for a given amount of modification, which explains why the threshold is larger for G. Note that the difference in excitation level between the reference and the modified sounds outside band 3 is due to the control of loudness. If the amplitude is lower in band 3, the amplitude has to be higher for other frequencies to give a constant overall loudness. This compensation effect is of course bigger when the difference in band 3 is large (which is why it is more noticeable for E than for G). A more systematic and quantitative study of excitation patterns is presented in Sec. IV E.

D. Results based on the five best subjects

The results for each type of modification varied markedly across subjects. At least some of this variability arose from differences in musical experience. It probably depended also on whether or not subjects had previously taken part in auditory discrimination tasks, especially tasks requiring analytical listening. The subjects used here were given very little training before testing began. Previous research has shown that training can lead to improved performance for many aspects of auditory discrimination (Irvine and Wright, 2005).

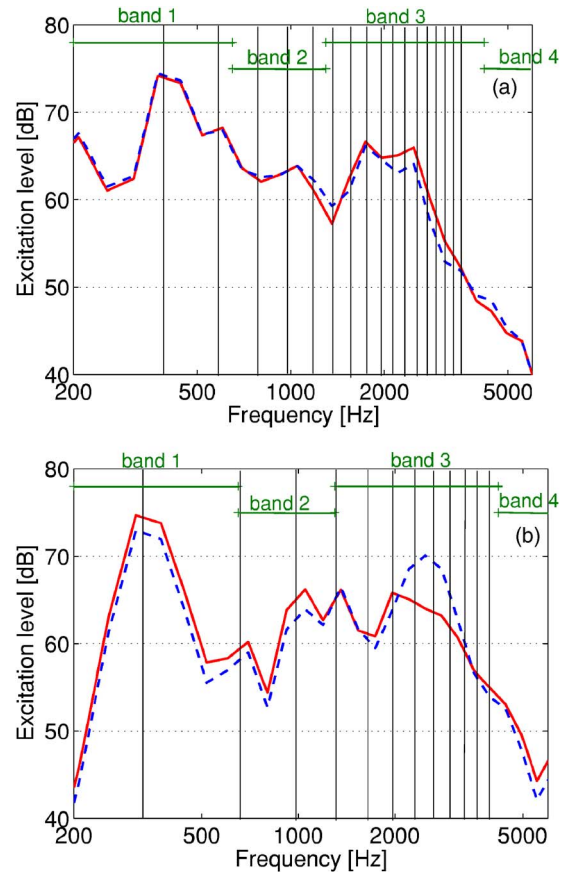


FIG. 5. (Color online) Excitation patterns for notes G (a) and E (b), for the reference sound (dashed line) and the modified sound (solid line) when the third band was shifted in frequency by 5%. The vertical lines show the positions of the harmonics for each note.

However, those subjects who initially show relatively good performance tend to show only small improvements with practice, while those who initially show relatively poor performance tend to improve markedly with practice (Fitzgerald and Wright, 2005; Irvine and Wright, 2005; Micheyl *et al.*, 2006; Moore, 1976). Hence it seems likely that the performance of the “best” subjects tested here would improve little with practice and would be representative of the thresholds that can be achieved by trained subjects attending to the optimal detection cues. To assess what this best performance was, we selected the five subjects who had the lowest average thresholds for the amplitude modification and the five who had the lowest average thresholds for the frequency modification (recall that different subjects were used for the two types of modification), and we determined mean scores for those subjects only. The results are shown in Fig. 6, together with predicted results which will be explained in Sec. IV E.

For the amplitude modification, the pattern of results is generally similar to that obtained for all subjects (Fig. 3), except that thresholds are lower. However, for the modification to A0 using note E, the mean threshold remained relatively high (about 8 dB) even for the five best subjects. The reason was found by looking more carefully at the spectra of the stimuli: The fundamental component of E corresponds to the trough lying above the A0 resonance, which is shifted in

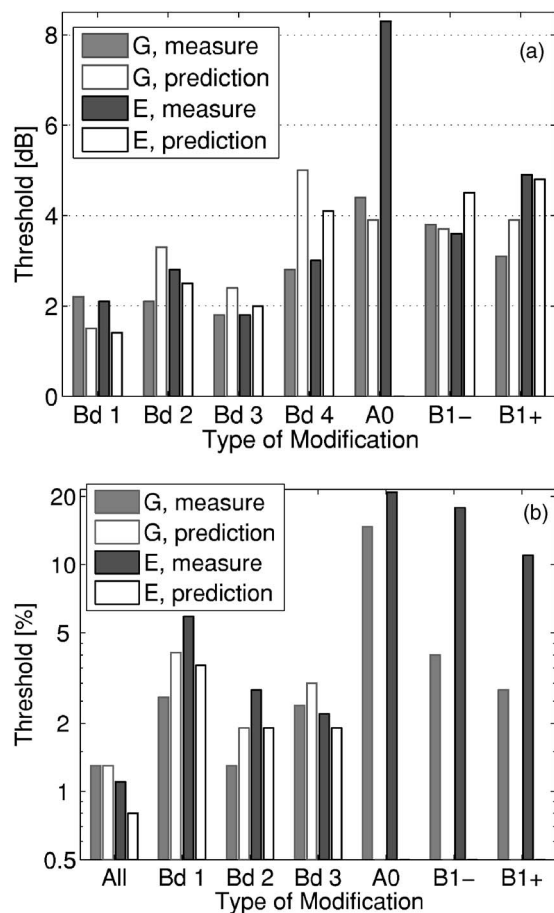


FIG. 6. Comparison between obtained mean thresholds for the five “best” subjects for each type of modification—(a) in amplitude and (b) in frequency—and the thresholds predicted by model 2. The upper and lower panels show results for the modifications in amplitude and frequency, respectively.

frequency when the amplitude of the resonance increases. Thus, the amplitude of the fundamental of E does not vary monotonically when the amplitude of A0 increases: The frequency of the trough moves from below the fundamental frequency of E to above it. Therefore, two different modifications can have the same effect on the amplitude of the fundamental component of E, if the corresponding troughs lie on either side of fundamental frequency of E. For modifications to bands 1–4, thresholds for the five best subjects were relatively small, in the range 2 to 3 dB, while thresholds for the single modes were mostly around 4 dB, again with the notable exception of A0 with the note E.

The thresholds for the amplitude modification obtained here can be compared to those obtained in experiments on “profile analysis” (Green, 1988). In such experiments, thresholds are measured for the detection of a change in level of a single “target” frequency component relative to the levels of other components, which form a kind of “background” or “profile.” To prevent subjects from using the change in level of the target component as a cue, the overall level of the whole stimulus is randomly varied from one stimulus to the next. When the background contains a large number of equal-amplitude components, the threshold for detecting a change in relative level of the target is typically only about

1 to 2 dB (Green, 1988). However, if the components in the background do not have equal amplitudes, i.e., the profile is irregular in some way, the thresholds increase to 2–4 dB (Kidd *et al.*, 1986), values comparable to those found here. Our thresholds are also similar to thresholds for detecting a change in amplitude of a single formant in synthetic speech sounds (Pols, 1999).

For the frequency modification, the pattern of results for the five best subjects is generally similar to that obtained for all subjects (Fig. 4), except that the thresholds for the five best subjects are not markedly higher for band 2 than for the other Dünwald bands. Thresholds are lowest (about 1%) for the modification to all bands.

The thresholds for the modifications to a single Dünwald band are comparable to thresholds for detecting changes in the center frequency of a single formant in synthetic speech sounds (Kewley-Port and Watson, 1994; Lyzenga and Horst, 1997). The relatively low thresholds found here for discrimination of a change in all modes simultaneously are consistent with the finding that, for synthetic speech sounds, thresholds are lower when all formants are shifted together in the same direction than when only a single formant is shifted (Hawks, 1994).

E. Interpretation using excitation patterns

It is of interest to explore the extent to which existing auditory models of intensity and frequency discrimination might be capable of accounting for our results. Accordingly, we have attempted to model the results using three different, empirically grounded, auditory models.

Excitation patterns have been used as the basis of models for predicting the ability to detect changes in frequency and/or level of sounds (Zwicker, 1956; Florentine and Buus, 1981; Moore and Sek, 1992, 1994). Here, we compare how well three different excitation-pattern models can account for the results obtained in the present experiment. For the analysis presented here, the excitation patterns were calculated with filter center frequencies spaced at 1- ERB_N intervals, using the ERB_N -number scale given by Glasberg and Moore (1990). To calculate excitation patterns from the waveforms of the sounds, we used the method described by Glasberg and Moore (2002). The excitation patterns were calculated at 50-ms intervals. The analysis that follows is based on the results of the five best subjects for each type of modification, as described in Sec. IV D.

The models make use of the detectability index, d' (Green and Swets, 1974; Macmillan and Creelman, 1991). We assume here that the contribution to detectability in the i th frequency channel, d'_i , is proportional to the excitation level difference ΔL_i in the i th ERB_N between the reference sound and the modified sound, when the modified sound is at the threshold value. Because of this assumption we actually calculated a quantity for each model, D_1 , D_2 , and D_3 , which was based on the ΔL_i values and was proportional to the value of d' for each model. The models differ in whether and how “information” is combined across channels, and therefore with our assumption, whether and how excitation level differences are combined.

The single-channel model (Zwicker, 1956, 1970) is based on the assumption that detection depends on monitoring the single place on the excitation pattern that changes the most:

$$D_1 = \max(\Delta L_i). \quad (1)$$

For this model, excitation patterns were determined every 50 ms and then averaged over time. The largest average excitation level difference given by a single channel was chosen as an estimate of D_1 .

The multichannel model with optimal processing (Florentine and Buus, 1981; Moore and Sek, 1992) is based on the assumption that information from different parts of the excitation pattern can be combined in an optimal manner:

$$D_2 = \sqrt{\sum_{i=1}^n (\Delta L_i^2)}. \quad (2)$$

For this model, D_2 was calculated every 50 ms and all quantities were then averaged to give the final estimated value of D_2 .

The multichannel model without optimal processing (Moore and Sek, 1992) assumes that observers base their decision on an unweighted sum of decision variables (d'_i values) across all channels:

$$D_3 = \frac{\sum_{i=1}^n (\Delta L_i)}{\sqrt{n}}. \quad (3)$$

For this model, the averaging across time was done in the same way as for model 2.

For all three models, only “active channels,” i.e., channels with excitation level above an assumed absolute threshold, were considered. The threshold excitation level was chosen to be 5 dB.

If any of the models worked perfectly, and there were no errors of measurement, then the calculated D values would be constant across all conditions except for the frequency modification of the single modes. This condition was conducted differently as the reference sound was not kept constant (see Sec. III A) so this condition is not included in the following analysis. The modification in amplitude of mode A0 is excluded as well, as this modification did not have a monotonic effect. The calculated D values were not constant for any of the models. However, we can assess which model is the best by evaluating, for each model, the coefficient of variation (Cov) of the values of D , which is defined as the ratio of the standard deviation to the mean, across all conditions. The Cov was 0.4, 0.3, and 0.4 for models 1, 2, and 3, respectively, so model 2 was marginally the best.

Using the mean value of D given by model 2, $D_2=6.1$, we can predict the threshold for each task that would give a value of D equal to this average value D_2 . The predictions were generated by successive iterations to find the amount of modification which would give a value of D equal to D_2 . Predictions are shown in Fig. 6. The rms deviation of the data from the predicted values is slightly less than 1 dB for the modifications in amplitude and 1% for the modifications

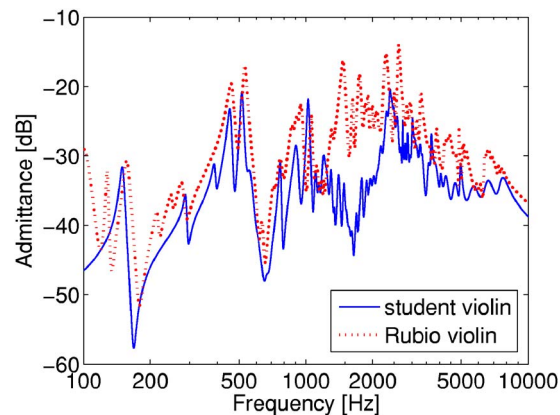


FIG. 7. (Color online) Comparison of the resynthesized input admittances of the Rubio violin and the student violin of relatively poor quality.

in frequency, so the model is quite accurate and is equally good for G and E. This is particularly clear for the predictions obtained when all modes are shifted where, on average, both thresholds and predictions are lower than any that could be predicted on the basis of results obtained in response to shifts in individual bands. This suggests that listeners did indeed integrate information across a wide range of frequencies in a way which was close to optimal.

V. EXPERIMENTS WITH ANOTHER VIOLIN

All of the preceding results were obtained using only one violin. It is a possibility that the acoustical particularities of that instrument had a significant impact on listeners’ responses. In order to test this, and also to test the robustness of the excitation-pattern model, a subset of the listening tests was repeated using a student violin judged to be of relatively poor quality, and which had acoustical characteristics that differed considerably from those of the Rubio violin. Its resynthesized input admittance is compared to that of the Rubio violin in Fig. 7.

It was decided to restrict this additional experiment to the shift in amplitude of all modes in each of the Dünwald bands. The subjects for this second experiment were 15 musicians, among whom 5 had participated in the experiment with the Rubio violin.

Figure 8 compares thresholds for the two violins. For the Rubio violin, thresholds correspond to the average across 15 musicians, including the 5 musicians who also did the second experiment, and 10 others chosen arbitrarily from the 31 other musicians. Although the thresholds tend to be lower for the “bad” violin (except for the modification of band 1 for note E, for which thresholds are equal), an ANOVA showed no significant difference between thresholds for the two violins [$F(1,28)=2.3$, $p=0.145$], as the standard deviations were quite large.

Figure 9 shows a comparison between the predictions obtained with model 2, using the same value for D_2 and the average results for the five best subjects. The rms deviation is 0.8 dB so the fit is quite good.

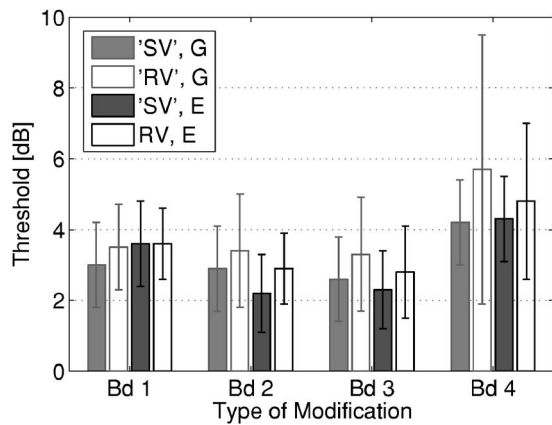


FIG. 8. Comparison of the thresholds obtained by musicians for a manipulation in amplitude of the modes within each Dünwald band, for the Rubio violin (RV) and the student violin (SV). See the key for details of the conditions. The error bars represent ± 1 s.d. across subjects.

VI. SUMMARY AND CONCLUSIONS

The work described in this paper represents the first stage of a project to provide quantitative information about the discriminability of and perceptual preferences between violins. The eventual aim of the project is to make direct links between the perceptual results and parameters relevant to instrument makers—materials choices, constructional geometry, and set-up details. The results will also inform efforts to improve the quality of computer-synthesized string sounds.

This initial study explored two aspects of violin acoustics which have received great prominence in the earlier literature as possible indicators of aspects of “quality:” The three individual low-frequency modes of vibration (below 700 Hz), which dominate the sound of a violin and are usually labeled A0 (a modified Helmholtz resonance), B1– and B1+ (two strong “wood modes”); and a set of four frequency bands proposed by Dünwald (190–650, 650–1300, 1300–4200, and 4200–6400 Hz) on the basis of measurements of a large number of violins of varying quality. Tests were conducted to establish thresholds for the perception of a change

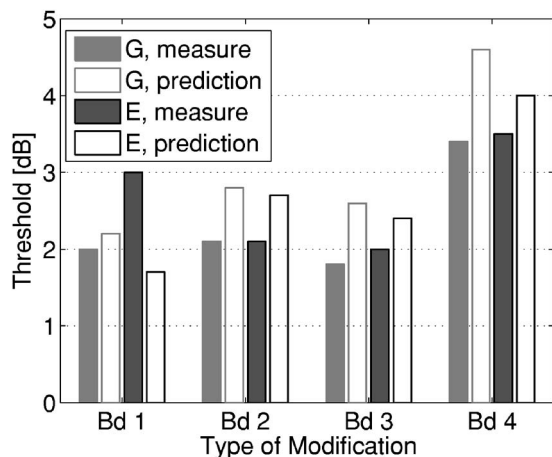


FIG. 9. Comparison between results and predictions of model 2, for the five best subjects, for a modification in amplitude of the student violin’s modes. See the key for details of the conditions.

in frequency or amplitude of each of the three modes separately and for blocks of modes lying in the four “Dünwald bands.” Finally, a test was conducted in which the frequencies of all modes were varied simultaneously.

Results were presented for two groups of listeners—with and without extensive musical training. As might have been anticipated, the musically trained listeners had consistently lower thresholds. A series of ANOVAs was performed to investigate the significance of different main effects and interactions, in particular the influence of the type of modification and of the single note used as input on the thresholds. To obtain an estimate of the discrimination thresholds attainable by trained listeners attending to the optimal detection cues, results were calculated for the best five subjects in each group of tests. For modifications of amplitude, these best thresholds were in the range 3–5 dB for individual modes and 1–3 dB for the Dünwald bands. For modifications in frequency, the best listeners had thresholds around 3%–5% for individual modes, 1%–3% for the first three Dünwald bands, and around 1% when all frequencies were varied simultaneously. Frequency changes in the fourth Dünwald band were not detectable.

Predictions of threshold were made using three different models based on excitation patterns. The best performance was obtained using a multichannel model based on optimal combination of information across channels. This model reproduced the main results well, including the remarkably low threshold for detection of a simultaneous frequency shift of all modes. This success allows tentative predictions to be made for any combination of input signal and filter modification. This may allow a more systematic design of future tests, since input signals could be optimally chosen to allow a listener to best discriminate a given type of filter modification.

The best choice of stimulus sounds for these threshold tests proved to be very short single notes, probably partly because they allowed echoic memory to assist discrimination. When tests were repeated using a short musical phrase, higher thresholds were obtained. This finding is in some respects counterintuitive, but is consistent with what is known about “informational masking;” it is difficult to detect a subtle change in a sound when the sound itself is varying strongly (Watson, 1987).

There is strong anecdotal evidence that certain subtle differences between violins can be perceived by violinists, and have great importance to them. It is sufficient to note that the market values of superficially similar violins range over some four orders of magnitude: In round numbers, from about \$100 to \$1 000 000. The authors are conscious of the fact that the tests described here are based on sounds which are very unmusical—our short single notes are barely recognizable as violin sounds. It would surely not be possible to obtain subtle judgments of quality and preference from such sounds. The likely conclusion is that the thresholds obtained here only tell part of the story of violin discrimination, and that higher-level perceptual processes are brought into play when a trained violinist compares instruments in a musical setting—for example, during the process of choosing a new instrument.

There are many directions for possible further work along the lines explored in the study reported here. There are many more parameters which could be varied to establish perceptual thresholds. Of particular interest might be parameters with a direct interpretation in terms of a physical modification to a violin: For example, the parameters influencing the “bridge hill” (e.g., Woodhouse, 2005), or material properties of the wood used to build the violin body. Another important way of extending the study would be to use radiated-sound transfer functions rather than structural response (input admittance) as in this study.

ACKNOWLEDGMENTS

This research was supported by the Leverhulme Trust. We thank Brian Glasberg and Hugh Greenish for their help in the design of the experiment, and Emily Smith for having conducted part of the experiment. We are also grateful to all the subjects who took part in this study. We thank two anonymous reviewers for detailed constructive comments.

- Cremer, L. (1985). *The Physics of the Violin* (MIT, Boston).
- Darwin, C. J., Turvey, M., and Crowder, R. G. (1972). “An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage,” *Cogn. Psychol.* **3**, 255–267.
- Dünnwald, H. (1991). “Deduction of objective quality parameters on old and new violins,” *Catgut Acoust. Soc. J. Series II* **1**, 1–5.
- Durup, F., and Jansson, E. V. (2005). “The quest of the violin bridge-hill,” *Acta. Acust. Acust.* **91**, 206–213.
- Ewins, D. (2000). *Modal Testing: Theory, Practice and Application* (Research Studies Press, Baldock, England).
- Fitzgerald, M. B., and Wright, B. A. (2005). “A perceptual learning investigation of the pitch elicited by amplitude-modulated noise,” *J. Acoust. Soc. Am.* **118**, 3794–3803.
- Florentine, M., and Buus, S. (1981). “An excitation-pattern model for intensity discrimination,” *J. Acoust. Soc. Am.* **70**, 1646–1654.
- Fritz, C. (2007). “Experimental investigation of the perceptual correlates of violin acoustics,” <http://www2.eng.cam.ac.uk/~cf291/project.htm>. (Last accessed October 2007).
- Fritz, C., Cross, I., Smith, E., Weaver, K., Petersen, U., Woodhouse, J., and Moore, B. C. J. (2006). “Perceptual correlates of violin acoustics,” in *Proceedings of the Ninth International Conference on Music Perception and Cognition*, Bologna, Italy.
- Glasberg, B. R., and Moore, B. C. J. (1990). “Derivation of auditory filter shapes from notched-noise data,” *Hear. Res.* **47**, 103–138.
- Glasberg, B. R., and Moore, B. C. J. (2002). “A model of loudness applicable to time-varying sounds,” *J. Audio Eng. Soc.* **50**, 331–342.
- Green, D. M. (1988). *Profile Analysis* (Oxford University Press, Oxford).
- Green, D. M., and Swets, J. A. (1974). *Signal Detection Theory and Psychophysics* (Krieger, New York).
- Guettler, K., and Askenfelt, A. (1997). “Acceptance limits for the duration of pre-Helmholtz transients in bowed string attacks,” *J. Acoust. Soc. Am.* **101**, 2903–2913.
- Hawks, J. W. (1994). “Difference limens for formant patterns of vowel sounds,” *J. Acoust. Soc. Am.* **95**, 1074–1084.
- Hutchins, C. M. (1962). “The physics of violins,” *Sci. Am.* **207**, 79–93.
- Irvine, D. R., and Wright, B. A. (2005). “Plasticity of spectral processing,” *Int. Rev. Neurobiol.* **70**, 435–472.
- Jansson, E. V. (1997). “Admittance measurements of 25 high quality violins,” *Acust. Acta Acust.* **83**, 337–341.
- Kewley-Port, D., and Watson, C. S. (1994). “Formant-frequency discrimination for isolated English vowels,” *J. Acoust. Soc. Am.* **95**, 485–496.
- Kidd, G., Mason, C. R., and Green, D. M. (1986). “Auditory profile analysis of irregular sound spectra,” *J. Acoust. Soc. Am.* **79**, 1045–1053.
- Langhoff, A., Farina, A., and Tronchin, L. (1995). “Comparison of violin impulse responses by listening to convoluted signals,” in *Proceedings of the International Symposium on Musical Acoustics*, Paris.
- Levitt, H. (1971). “Transformed up-down methods in psychoacoustics,” *J. Acoust. Soc. Am.* **49**, 467–477.
- Liu, C., and Kewley-Port, D. (2004). “Vowel formant discrimination for high-fidelity speech,” *J. Acoust. Soc. Am.* **116**, 1224–1233.
- Lyzenga, J., and Horst, J. W. (1997). “Frequency discrimination of stylized synthetic vowels with a single formant,” *J. Acoust. Soc. Am.* **102**, 1755–1767.
- Macmillan, N. A., and Creelman, C. D. (1991). *Detection Theory: A User’s Guide* (Cambridge University Press, Cambridge).
- Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). “Influence of musical and psychoacoustical training on pitch discrimination,” *Hear. Res.* **219**, 36–47.
- Moore, B. C. J. (1976). “Comparison of frequency DLs for pulsed tones and modulated tones,” *Br. J. Audiol.* **10**, 17–20.
- Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).
- Moore, B. C. J., and Glasberg, B. R. (1983). “Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,” *J. Acoust. Soc. Am.* **74**, 750–753.
- Moore, B. C. J., and Sek, A. (1992). “Detection of combined frequency and amplitude modulation,” *J. Acoust. Soc. Am.* **92**, 3119–3131.
- Moore, B. C. J., and Sek, A. (1994). “Effects of carrier frequency and background noise on the detection of mixed modulation,” *J. Acoust. Soc. Am.* **96**, 741–751.
- Pols, L. C. W. (1999). “Flexible, robust, and efficient human speech processing versus present-day speech technology,” in *Proceedings of the International Congress of Phonetic Sciences ICPHS’99*, San Francisco, Vol. **1**, pp. 9–16.
- Royster, J. D., Royster, L. H., and Killion, M. C. (1991). “Sound exposures and hearing thresholds of symphony orchestra musicians,” *J. Acoust. Soc. Am.* **89**, 2793–2803.
- Skudrzyk, E. (1981). *Simple and Complex Vibratory Systems* (Pennsylvania State University Press, University Park, PA).
- Watson, C. S. (1987). “Uncertainty, informational masking, and the capacity of immediate auditory memory,” in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ).
- Woodhouse, J. (1993). “On the playability of violins. Part 2. Minimum bow force and transients,” *Acustica* **78**, 137–153.
- Woodhouse, J. (2005). “On the ‘bridge hill’ of the violin,” *Acta. Acust. Acust.* **91**, 155–165.
- Zwicker, E. (1956). “Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs (The foundations for determining the information capacity of the auditory system),” *Acustica* **6**, 356–381.
- Zwicker, E. (1970). “Masking and psychological excitation as consequences of the ear’s frequency analysis,” in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 376–396.

Within- and across-channel gap detection in cochlear implant listeners

John H. Grose^{a)} and Emily Buss

Department of Otolaryngology - Head & Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7070

(Received 31 January 2007; revised 30 August 2007; accepted 1 September 2007)

This study examined within- and across-electrode-channel processing of temporal gaps in successful users of MED-EL COMBI 40+ cochlear implants. The first experiment tested across-ear gap duration discrimination (GDD) in four listeners with bilateral implants. The results demonstrated that across-ear GDD thresholds are elevated relative to monaural, within-electrode-channel thresholds; the size of the threshold shift was approximately the same as for monaural, across-electrode-channel configurations. Experiment 1 also demonstrated a decline in GDD performance for channel-asymmetric markers. The second experiment tested the effect of envelope fluctuation on gap detection (GD) for monaural markers carried on a single electrode channel. Results from five cochlear implant listeners indicated that envelopes associated with 50-Hz wide bands of noise resulted in poorer GD thresholds than envelopes associated with 300-Hz wide bands of noise. In both cases GD thresholds improved when envelope fluctuations were compressed by an exponent of 0.2. The results of both experiments parallel those found for acoustic hearing, therefore suggesting that temporal processing of gaps is largely limited by factors central to the cochlea.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2793608]

PACS number(s): 43.66.Mk, 43.66.Ts [RYL]

Pages: 3651–3658

I. INTRODUCTION

Gap detection (GD) measures sensitivity to an interruption in an otherwise ongoing sound. In the absence of the imposed gap, the sound is perceived as continuous throughout the listening interval. Gap duration discrimination (GDD), on the other hand, measures sensitivity to a change in the duration of an interval between two perceptually discontinuous markers. The discontinuity may be a prolonged silent interval between two similar markers, or it may be the transition inherent to the juxtaposition of two dissimilar markers. In the latter case, a discontinuity is perceived even if no temporal gap is interjected between the two markers; here, GDD measures sensitivity to the duration of the interjected silent interval.

In acoustic hearing, thresholds for both GD and GDD are lowest when the leading and trailing markers of the gap are spectrally similar. Thresholds are usually elevated when any of the following conditions apply: (1) The leading marker [M1] and the trailing marker [M2] do not overlap in frequency (e.g., Perrott and Williams, 1971; Divenyi and Danner, 1977); (2) M1 and M2 overlap in frequency content but are otherwise spectrally dissimilar (e.g., Formby *et al.*, 1993; Phillips and Hall, 2002); (3) M1 and M2 are spectrally similar but are perceptually dissimilar along some other dimension such as modulation type/rate or pitch (e.g., Oxenham, 2000; Grose *et al.*, 2001); (4) M1 and M2 are spectrally identical but are presented to opposite ears (e.g., Penner, 1977; Phillips *et al.*, 1997); and (5) M1 and M2 are spectrally identical but are characterized by perceptually salient fluctuations (e.g., Glasberg and Moore, 1992; Grose *et al.*, 2007a).

Although GD and GDD have not been studied extensively in listeners with cochlear implants, parallels to normal acoustic hearing can be seen in existing data. In electric hearing, the detection and discrimination of gaps is most acute when M1 and M2 are carried on the same electrode channel. Thresholds are usually elevated when any of the following conditions apply: (1) M1 and M2 are carried on different electrode channels (Hanekom and Shannon, 1998; van Wieringen and Wouters, 1999); (2) M1 and M2 have some, but not all, electrode channels in common (van Wieringen and Wouters, 1999); (3) M1 and M2 share the same electrode channel but differ in some dimension such as pulse rate or level (Chatterjee *et al.*, 1998); and (4) M1 and M2 share the same electrode channel but are characterized by amplitude fluctuations (Dobie and Dillier, 1985; Hochmair-Desoyer *et al.*, 1985).

These parallels between acoustic and electric hearing are informative in terms of mechanisms underlying temporal resolution. One particular area where comparison across stimulation domains provides insight is the issue of within-versus across-channel processing. In acoustic hearing, the inverse relationship between frequency specificity and stimulus rise/fall time (or duration in general) confounds the assessment of temporal resolution for stimuli confined to a narrow frequency region. In addition, the need for relatively long rise/fall times to restrict gap cues to a single peripheral auditory filter, or frequency channel, can result in an imposed “modulation” that is perceptually similar to the modulations inherent to the fluctuation pattern of the narrow-band stimulus. In electric hearing, assessment of temporal resolution for stimuli restricted to a single electrode channel is largely free from confounding effects of frequency specificity (from the stimulation perspective) and therefore the distinction be-

^{a)}Electronic mail: jhg@med.unc.edu

tween within-electrode-channel processing and across-electrode-channel processing can be assessed in a more straightforward manner. The focus of this study is on within- and across-channel temporal processing in electric hearing using both GD and GDD tasks. The purpose is to examine two specific issues: (1) the relative performance of across-ear (bilateral) GDD in comparison to monaural, across-electrode-channel GDD; (2) the effect of within-electrode-channel amplitude envelope fluctuations on gap threshold. The literature pertinent to each of these issues will be briefly reviewed.

In acoustic hearing, a plethora of studies have shown that GDD thresholds become progressively elevated as the frequency separation between the leading and trailing markers increases. This deterioration in GDD performance reaches an asymptote at frequency separations greater than about one octave (Divenyi and Danner, 1977; Heinz *et al.*, 1996; Formby *et al.*, 1998a). For these separations, it is unlikely that the listener is monitoring gap duration in a single frequency region activated by spread of excitation from both markers (i.e., a within-frequency-channel temporal judgment) (Formby *et al.*, 1998a). Rather, GDD likely reflects the monitoring of elapsed time between the offset of M1 in one frequency channel and the onset of M2 in an independent frequency channel (i.e., an across-frequency-channel temporal judgment). Two studies that compared monaural across-frequency-channel GDD with across-ear GDD noted that gap thresholds for monaural markers widely separated in frequency are similar to those measured across ears for isofrequency markers (Phillips *et al.*, 1997; Formby *et al.*, 1998a). This correspondence in performance for across-ear GDD and monaural across-frequency-channel GDD prompted Phillips *et al.* (1997) to suggest that both configurations were examples of across-*perceptual*-channel temporal processing. The notion here is that temporal judgments are inherently poor when they must be made across perceptual channels irrespective of whether those channels correspond physically to different frequency bands in the same ear or to similar frequency bands but in different ears.

In terms of electric hearing, a similar elevation in gap threshold has been noted as the separation between electrode channels carrying M1 and M2 increases (Hanekom and Shannon, 1998). Hanekom and Shannon argued that small gap thresholds for relatively proximal electrode channels represented direct current spread (i.e., channel interactions), but the asymptotic values seen for wider channel separations represented central temporal processing limits. van Wieringen and Wouters (1999) also found that gap thresholds were smaller in some implanted listeners when both markers were carried on the same electrode channel than when they were carried on separate electrode channels. This elevation in threshold for markers carried by well-separated electrodes also strongly suggests that the two markers are stimulating different neural populations. Overall, analogous findings therefore exist for *monaural* temporal processing measured across frequency channels (acoustic hearing) and electrode channels (electric hearing). However, no across-ear GDD data exist for electric hearing, primarily because bilateral cochlear implantation has tended to be the exception rather

than the norm. Moreover, even in cases of bilateral implantation the implants are driven by independent speech processors, making precise temporal measurements technically difficult to implement. The availability of synchronized laboratory-based implant interfaces has now overcome this limitation. The primary purpose of experiment 1, therefore, is to measure monaural across-electrode-channel GDD and across-ear GDD in the same listeners with bilateral cochlear implants.

A subsidiary purpose of experiment 1 was to measure GDD for electrode configurations where M1 and M2 are asymmetric in terms of electrode channel/ear combinations. In acoustic hearing, frequency asymmetry between markers is detrimental to GDD. That is, when the spectral content of one marker is subsumed within the spectral content of the other marker GDD performance tends to be poor—particularly in the case where the additional spectral components are present in the trailing marker (Formby *et al.*, 1998b). This finding is remarkable because it demonstrates that complete definition of the temporal gap within a single frequency channel is not in itself a sufficient condition for optimal gap detection performance. Grose *et al.* (2007b) suggested that the detrimental effect of a spectrally complex trailing marker is due to informational masking associated with the occurrence of novel spectral events in the trailing marker. In their study of electric hearing, van Wieringen and Wouters (1999) report findings analogous to those in acoustic hearing; when M1 was carried on a single electrode channel and M2 was carried simultaneously on two or three electrode channels, some listeners exhibited elevated gap thresholds relative to the baseline condition where both M1 and M2 were carried on the same single electrode channel. The subsidiary purpose of experiment 1 was to take this exploration of marker asymmetry a step further to test novel conditions that combined across-ear and across-electrode-channel asymmetries.

The purpose of experiment 2 was to assess the effect of amplitude fluctuations on gap thresholds in electric hearing. The focus here was entirely on within-electrode-channel processing since the stimuli were constrained to a single electrode channel. In acoustic hearing, studies have shown that gap thresholds vary as a function of the degree of envelope fluctuation such that the greater the fluctuation, the higher the gap detection threshold (Glasberg and Moore, 1992; Moore *et al.*, 2001; Grose *et al.*, 2007a). It is generally thought that this is due to confusion between the imposed gap and the ongoing envelope fluctuations. One manifestation of this effect is that GD thresholds are inversely proportional to the bandwidth of the stimulus; this is because the inherent fluctuations of narrow bands of noise tend to be highly salient. Some limited findings suggest that amplitude modulation degrades gap detection in electric hearing as well (Dobie and Dillier, 1985; Hochmair-Desoyer *et al.*, 1985). Hochmair-Desoyer *et al.* (1985), using a single-channel implant, found that GD thresholds tended to be higher for an input stimulus consisting of a 750-Hz wide band of noise compared to a 2000-Hz wide band of noise. Although this shows a dependency on stimulus bandwidth in the expected direction, bandwidths of this extent are usually not associ-

ated with salient fluctuations. [These bandwidths correspond to average modulation rates of about 480 and 1280 Hz, respectively (Rice, 1954), and measures of temporal modulation transfer functions in both acoustic hearing (Viemeister, 1979) and electric hearing (Shannon, 1992) predict poor sensitivity to these rates of modulation.] In one patient with a single-channel extracochlear implant, Dobie and Dillier (1985) found that GD thresholds for a 316-Hz wide noise band stimulus were markedly worse than for an unmodulated stimulus. Because of the limited, almost anecdotal, nature of these data concerning GD for modulated stimuli in electric hearing, the purpose of experiment 2 was to provide a parametric study of the effects of envelope fluctuation on the temporal processing of gaps in listeners with cochlear implants.

In summary, this investigation examined within- and across-electrode-channel temporal processing in cochlear implant listeners using GD and GDD tasks. The investigation consisted of two related experiments. Experiment 1 measured monaural across-electrode-channel GDD and across-ear GDD in the same listeners with bilateral cochlear implants. Subsidiary conditions measured GDD for monaural and bilateral channel-asymmetric markers. Experiment 2 measured GD as a function of envelope fluctuation on a single electrode channel. The general methodology was the same across the two experiments, and this will be described next.

II. GENERAL METHODOLOGY

A. Stimulus generation and presentation

All listeners in this study were implanted with the MED-EL COMBI 40+ device, and stimuli were delivered to the implant through a Research Interface Box (RIB, University of Innsbruck, Austria), bypassing the speech processor. In conditions of bilateral presentation, two RIB units were linked in hardware such that one unit controlled the timing for both units (master and slave), allowing fine control of the pulse trains presented to each side. The COMBI 40+ device incorporates a 12-channel electrode array, and the stimulation mode is monopolar such that an electrode channel is identified by the active electrode on the array referenced to a common return electrode embedded in the temporalis muscle. The electrode numbering convention for this device places the highest-numbered electrode in the base of the cochlea.

Stimuli were generated in MATLAB (Mathworks), and each marker consisted of a biphasic pulse train presented to either one or two electrode channels. Each biphasic pulse had a phase duration of $26.7 \mu\text{s}$ and was presented at 3 kHz—a rate approximately twice that of the typical pulse rate used in a clinical map. This high rate was chosen based on previous data suggesting that the best gap detection performance is obtained with a high-rate carrier (Preece and Tyler, 1989). When two electrodes carried a marker, the pulse trains were interleaved with a $167\text{-}\mu\text{s}$ delay between each successive pulse onset so as to maximize the interval between pulses, thus minimizing channel interactions. Three channels were identified for stimulus presentation, channels

3, 6 and 9. These channels were associated with unique and discriminable pitch percepts in all patients, and all incorporated these channels in their every-day clinical map. The one exception to this was Observer 2 for whom channel 6 activation elicited facial stimulation in both the clinical map and the experimental map; for this observer, channels 2, 5, and 9 were used instead. However, for ease of description, the data from this observer will be treated as if channels 3, 6, and 9 had been stimulated.

B. Procedures

Stimuli were presented in a two-alternative forced-choice paradigm, with a 500-ms inter-stimulus interval. Responses were entered on a handheld response box, and lights visually indicated listening intervals and provided feedback. Gap duration between the two markers was adjusted based on the observer's responses: Gap duration was reduced following two successive correct responses and increased following each incorrect response, a procedure that estimates the 70.7% correct point on the psychometric function. Adjustments in gap duration were made in factors of 1.2. A track continued until six reversals in gap duration were obtained, and the threshold estimate for that track was computed as the geometric mean of the duration at the last four reversals. Because the gap duration was rounded to the nearest integer number of periods of the 3-kHz pulse train, thresholds below 1 ms were judged to be at the floor of sensitivity and could not be estimated with greater precision by this procedure. For the purposes of data analysis, a value of 1 ms was entered for these floor threshold estimates. Thresholds reported below are the geometric mean of at least two, but typically three or four, threshold estimates.

III. EXPERIMENT 1. ACROSS-ELECTRODE-CHANNEL AND ACROSS-EAR GAP DURATION DISCRIMINATION

A. Method

1. Observers

Four adult bilateral cochlear implant listeners participated (Observers 1–4). All four were bilaterally implanted with MED-EL COMBI 40+ devices, and in all cases the surgical notes indicate that a full insertion of the electrode array was achieved. In addition, two listeners unilaterally implanted with MED-EL COMBI 40+ devices provided data in the monaural conditions (Observers 5 and 6). Demographics for all observers are provided in Table I. All observers could be described as good performers, as supported by their open-set CNC scores.

2. Loudness balancing

At the initial test session, the maximum comfortable listening level (M-level) was identified for all of the channels used in the experiment, and measured independently in each ear for bilateral observers. This was accomplished using clinical software and a standard clinical interface (Diagnostic Interface Box, MED-EL Corporation, Austria). Stimuli used in this procedure were 50-ms samples of a 3-kHz biphasic

TABLE I. Demographics of all CI observers, including open set speech testing score using the CNC test (audio only). Duration of deafness is the time between onset of profound hearing loss and receipt of a cochlear implant.

Observer No.	Gender	Age (yr)	Duration of deafness (yr)	Implant	Duration of use (yr)	CNC (%)
1	F	43.8	2	Bilateral	3.1	74
2	F	52.9	18	Bilateral	3.5	88
3	M	59.8	4	Bilateral	3.9	54
4	M	29.3	6	Bilateral	3.1	62
5	M	27.2	27	Left	6.0	89
6	F	26.9	21	Right	4.3	90
7	M	69.1	26	Left	3.5	62
8	M	37.1	23	Left	2.3	56
9	F	48.8	43	Right	2.9	86

pulse train with 26.7- μ s duration per phase—similar stimulus parameters as used in the body of the experiment. After making an initial estimate of M-level for each electrode, the observer listened to a series of stimulus sequences, ascending through the three channels, and was asked to make any necessary adjustments in level in order to achieve equal loudness across channels. Bilateral patients were presented with an analogous opportunity to compare loudness across ears with a series of stimulus sequences incorporating both left and right side stimuli, a manipulation carried out using a pair of linked RIB interfaces. (One bilateral patient did not undergo this comparison.) The goal of these procedures was to identify the stimulus level associated with a relatively high and uniform sensation of loudness across all the channels used in the experiment. A map based on these data was saved for each observer/ear and served as the reference in the experimental software to prevent inadvertent overstimulation. For patients who attended multiple sessions, the map was checked prior to each session. In the majority of cases, the map was stable; in a few instances, minor adjustments were made. The amplitude of each pulse train used in the experiment was set to the M-level stored for that particular electrode irrespective of the number of electrodes (1 or 2) stimulated.

3. Stimuli and conditions

The duration of each marker (M1 and M2) was nominally 60 ms, but the actual duration varied randomly on a presentation-by-presentation basis from a uniform distribution ranging between 40 and 80 ms (i.e., 60 ± 20 ms). The randomization of marker duration was designed to make overall stimulus duration (two markers plus gap) an unreliable cue to the presence/duration of a gap between markers. Experimental conditions are described in terms of the channel(s) associated with each marker, with the number referring to the electrode channel and the letter referring to the ear of presentation. For example, the notation [M1=6L, M2=3L] indicates that both the leading marker (M1) and the trailing marker (M2) were presented on the left side (L), but that M1 was presented to channel 6 whereas M2 was presented to channel 3. Using these conventions, the baseline conditions were the within-electrode-channel configurations for each side: [M1=6L, M2=6L] and [M1=6R, M2=6R]. Monaural

across-electrode-channel conditions were: [M1=6L, M2=3L], [M1=6L, M2=9L], and [M1=3L, M2=9L]. The iso-channel across-ear condition was [M1=6L, M2=6R]. Finally, the across-ear, across-electrode-channel condition was [M1=3L, M2=9R]. Note that same-numbered channels in opposite ears were not explicitly pitch matched. Pitch as a function of electrode number for bilaterally implanted subjects has been shown to vary across ears, a fact that has been noted primarily in conjunction with studies of localization (van Hoesel and Clark, 1997; Long *et al.* 2003, 2006). Nevertheless, it was hypothesized that same-numbered channels across ears would be associated with more nearly matched pitches than different-numbered channels. This hypothesis was supported by informal observer reports and is consistent with the pattern of gap detection data obtained. Subsidiary conditions examining bilateral channel-asymmetric markers ([M1=6L, M2=6R+9L]) and monaural channel-asymmetric markers ([M1=6L, M2=6L+9L], [M1=6R, M2=6R+9R]) were also tested in two observers.

B. Results and discussion

The results of the main conditions of experiment 1 are shown in Fig. 1. GDD thresholds are plotted as a function of the channel number and ear of presentation of M1 and M2. For each marker configuration, individual data are shown clustering around the group geometric mean (grey diamond). The baseline data, shown in Fig. 1(A), represent monaural within-electrode-channel performance. For each observer, performance was optimal in this condition and was almost the same in each ear (closed symbols indicate right-ear presentation; open symbols indicate left-ear presentation). Two of the observers (Observers 1 and 2) performed at floor level (≤ 1 ms) for this condition, as shown by the downward pointing arrows. In contrast, when the two markers were presented to different electrode channels in the same ear, thresholds were markedly elevated, as shown in Fig. 1(B). Thresholds continued to be elevated when the markers were presented across ears [Fig. 1(C)] either on the same-numbered electrode channel or across different electrode channels. Note that Observer 4 did not participate in all conditions. In order to assess the significance of this data pattern, the data from the three representative conditions that all four listeners participated in were submitted to a repeated-

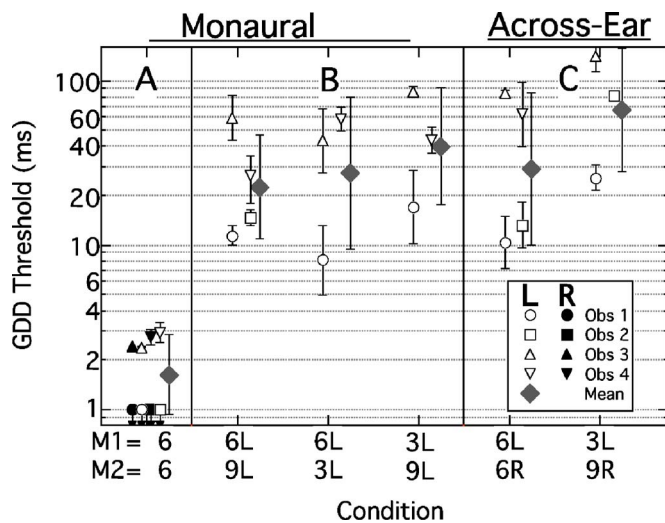


FIG. 1. (Color online) Individual and group mean GDD thresholds for bilaterally implanted CI listener. The abscissa gives the channel number and ear of presentation for the leading marker (M1) and the trailing marker (M2). Panel A shows the baseline data where both markers were presented monaurally to the same channel. Here, closed symbols indicate right-ear presentation and open symbols indicate left-ear presentation. Data points with downward-pointing arrows at 1 ms represent floor performance. The mean is for left-ear data only. Panel B shows monaural, across-electrode-channel data for three marker configurations. Panel C shows across-ear data for two marker configurations. Error bars are ± 1 s.d.

measures analysis of variance (RANOVA). The three conditions were: (1) Monaural, within-electrode-channel [M1=6L, M2=6L]; (2) monaural, across-electrode-channel [M1=6L, M2=9L]; and (3) across-ear, same-numbered electrode channel [M1=6L, M2=6R]. In this, and all analyses in this study, the log transforms of the data were used to ensure homogeneity of variance. The results of the analysis showed a significant effect of condition ($F_{2,6}=87.363$; $p < 0.001$), and post-hoc contrasts showed that thresholds in the baseline, within-electrode-channel, condition were significantly lower than either the monaural across-electrode-channel condition ($p=0.002$) or the across-ear condition ($p=0.001$), but that the latter two thresholds did not differ from each other ($p=0.347$).

This pattern of results confirms earlier findings in cochlear implant listeners that monaural gap thresholds are lowest when the markers are presented on the same electrode channel and decline when they are presented on different electrode channels (Hanekom and Shannon, 1998; van Wieringen and Wouters, 1999). The elevation in threshold for widely separated electrodes suggests that different neural populations were responding to the two markers. This suggestion of distinct neural populations is supported in the present study by the fact that all observers reported increasing pitch percepts with increasing electrode number. For proximal electrode pairs, current spread can result in a common population of neurons being stimulated (Hanekom and Shannon, 1998). The novel finding here is that gap thresholds for markers presented on the same-numbered electrode, but across ears, also appear to be elevated to a level similar to that of monaural, across-electrode-channel thresholds. This finding complements acoustic hearing results that show iso-frequency, across-ear thresholds to be elevated to similar

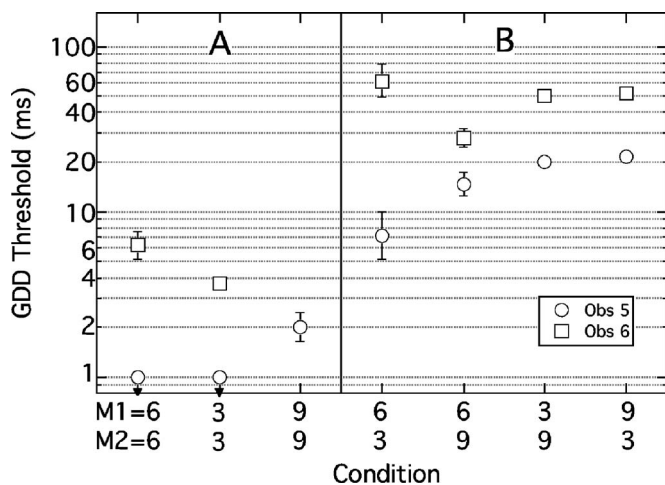


FIG. 2. Individual GDD thresholds for two unilaterally implanted CI listeners. Axis labeling is the same as Fig. 1. Panel A shows the baseline data where both markers were presented to the same channels. Panel B shows across-electrode-channel data. Error bar ± 1 s.d.

levels as monaural, across-frequency thresholds (Phillips *et al.*, 1997; Formby *et al.*, 1998a). This supports the contention that, for cochlear implant listeners, across-ear gap processing constitutes a form of across-perceptual-channel temporal processing, just as it does for acoustic hearing.

One limitation of this data set is that baseline measurements (monaural, within-electrode-channel) were not obtained for all of the electrode channels that were used in the across-electrode-channel conditions. That is, baseline measurements were obtained for only [M1=6L, M2=6L] even though electrode channels 3 and 9 were also used in the across-electrode-channel conditions. It could be argued that the elevated thresholds in these latter conditions might have been due to poor temporal resolution on channels 3 and 9. Although the original observers were not available to test this, two additional observers were recruited (Observers 5 and 6), both with unilateral implants (see Table I). These two observers were tested on two within-electrode-channel conditions ([M1=3, M2=3], [M1=6, M2=6]), and four across-electrode-channel conditions ([M1=6, M2=3], [M1=6, M2=9], [M1=3, M2=9], [M1=9, M2=3]). Observer 5 was also tested on the within-electrode-channel condition [M1=9, M2=9]. The results are shown in Fig. 2. Although the two observers differed in their overall GDD sensitivity, it is clear that the across-electrode-channel thresholds are uniformly higher than any of the within-electrode-channel thresholds. This makes it unlikely that the data pattern for the four bilateral observers shown in Fig. 1 is due to poor within-electrode-channel temporal processing on channels 3 or 9. Rather, the parsimonious interpretation is that monaural, across-electrode-channel temporal processing is inherently poorer than within-electrode-channel temporal processing.

Experiment 1 also included subsidiary conditions that were designed to assess the effects of marker asymmetries. The results for the two observers tested are shown in Fig. 3. For comparison purposes, monaural within-electrode-channel, across-electrode-channel, and across-ear thresholds for these two observers are replotted from Fig. 1. For the

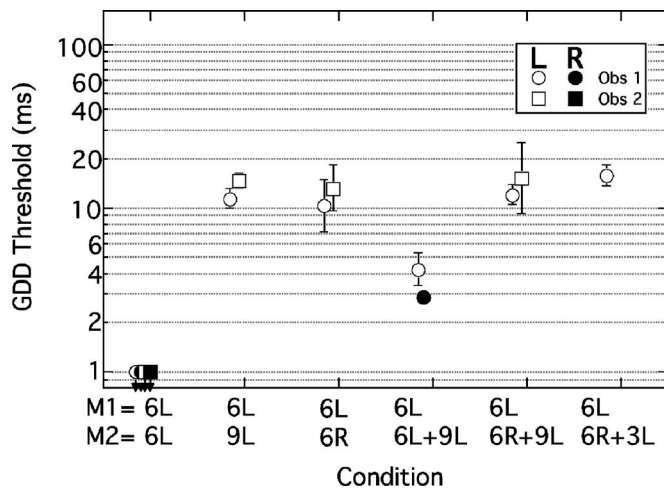


FIG. 3. Channel-asymmetric data from two observers. GDD thresholds are plotted as a function of M1 and M2 configuration. For comparison, monaural within-electrode channel, across-electrode-channel, and across-ear thresholds for these observers are replotted from Fig. 1. Where monaural data were collected separately for each ear, open symbols indicated left-ear data and closed symbols indicate right-ear data. Error bars indicate ± 1 s.d.

monaural channel-asymmetric condition [M1=6, M2=6+9], Observer 1 exhibited a relatively modest threshold elevation for both left- and right-ear stimulation that was not as great as the respective across-electrode-channel threshold [M1=6L, M2=9L]. This result is analogous to the acoustic hearing finding of Grose *et al.* (2007b), who showed that thresholds for frequency-asymmetric markers were intermediate between within-frequency-channel performance and across-frequency-channel performance. For the bilateral channel-asymmetric condition [M1=6L, M2=6R+9L], both observers exhibited elevated thresholds relative to the baseline within-electrode-channel condition. Although the bilateral channel-asymmetric condition can be viewed as a combination of the monaural, across-electrode-channel configuration [M1=6L, M2=9L] and the across-ear, same-numbered electrode configuration [M1=6L, M2=6R], the thresholds for all three of these conditions were very similar for the two observers. If sensitivity to each of the across-perceptual-channel configurations represented by [M1=6L, M2=9L] and [M1=6L, M2=6R] was independent of the other, then it might be conjectured that sensitivity to the combined configuration should lead to an improved threshold, i.e., the d -primes associated with the independent cues should combine. Failure to find such an improvement could be due to one of several possibilities: The cues associated with the two across-perceptual-channel configurations are not independent; the cues associated with the two configurations are independent but sensitivity to these cues at the given settings is not the same (unequal d -primes), and performance in the combined case is dominated by the more salient configuration; the cues associated with the two configurations are independent but the slopes of the psychometric function are very steep, rendering it difficult to demonstrate the integration. Without a careful mapping of the psychometric functions, it is not possible to distinguish between these possibilities.

In summary, the results of experiment 1 have shown that

across-ear GDD is relatively poor in electric hearing just as it is for acoustic hearing. Limited data from the subsidiary channel-asymmetric conditions confirmed that monaural asymmetries, where M1 and M2 have some—but not all—electrode channels in common, are detrimental to performance. However the decline in performance for monaural channel asymmetries is not as marked as it is for across-electrode-channel configurations where there is no overlap between M1 and M2. A bilateral asymmetry resulting from a combination of monaural across-electrode-channel and across-ear configurations also resulted in poor performance.

IV. EXPERIMENT 2. EFFECT OF ENVELOPE FLUCTUATION ON GAP DETECTION

The purpose of experiment 2 was to determine the effect of amplitude envelope fluctuations on gap thresholds in electric hearing. In acoustic hearing, gap thresholds vary as a function of the degree of envelope fluctuation (Glasberg and Moore, 1992; Moore *et al.*, 2001; Grose *et al.*, 2007a), but there has been no systematic evaluation of this parameter in electric hearing.

A. Method

1. Observers

Five adult cochlear implant listeners, aged 27–69 years, participated (Observers 1, 6–9). All were implanted with the MED-EL COMBI 40+ device. Two had participated in experiment 1. Demographics for these observers are provided in Table I.

2. Stimuli and conditions

All stimuli were presented unilaterally to electrode channel 6. The duration of each marker (M1 and M2) was nominally 125 ms, but varied randomly from 100 to 150 ms on a presentation-by-presentation basis. The pulse phase duration (26.7 μ s) and rate (3 kHz) were the same as in experiment 1. In the baseline condition, no amplitude modulation (AM) was applied to the pulse train. In the remaining four conditions, the AM pattern associated with the envelope of a narrow band of noise was applied to the pulse train. First, the Hilbert envelope was extracted from a narrow band of Gaussian noise, either 50 or 300 Hz wide. These two bandwidths were selected because, in acoustic hearing, the fluctuations associated with a 50-Hz wide band of noise are more perceptually salient than those of a 300-Hz wide band, leading to poorer GD performance in the narrow band (Eddins *et al.*, 1992). Next, the extracted envelope was raised to an exponent of either 1.0 (retaining the original) or 0.2 (compressing the modulation peaks). The resulting envelope was then applied to the pulse train. These manipulations resulted in four AM conditions associated with the 2-by-2 matrix of bandwidth (BW) and exponent (exp): [BW=50, exp=1.0], [BW=50, exp=0.2], [BW=300, exp=1] and [BW=300, exp=0.2]. For each condition, the envelope was assigned to M1 and M2 as if the ongoing envelope had been zeroed out during the gap (i.e., the envelope at the onset of M2 resumed

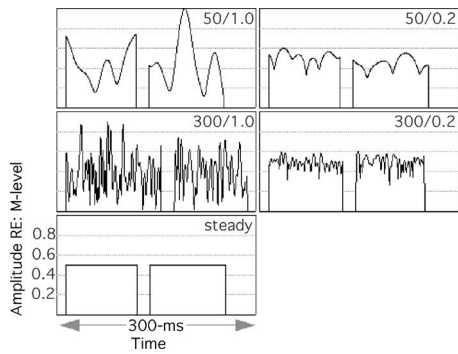


FIG. 4. Examples of marker envelopes for the five conditions of experiment 2. Each panel shows the envelope for a combination of bandwidth/exponent.

where it would have been had the envelope of MI continued throughout the gap). Examples of the marker envelopes are shown in Fig. 4.

The stimulus level was determined by first measuring the maximum comfortable listening level (M-level) for an equal-amplitude pulse train presented on electrode channel 6. Detection threshold (T-level) for this stimulus was also measured. The M-level was used to fix the *peak* value of markers in the [BW=50, exp=1] condition, and the envelope minimum for this condition was set to the T level. This ensured that the full dynamic range of the stimulus was audible. Loudness balancing was then performed between the [BW=50, exp=1] condition and each of the remaining three conditions.

B. Results and discussion

The results of experiment 2 are shown in Fig. 5. Group mean gap thresholds are shown for each condition (open squares), with results for individual observers clustering around the respective mean. In the baseline (unmodulated) condition, thresholds ranged from 1 ms (floor) to just over 10 ms. Threshold was not measured for Observer 1 in this condition.¹ In the two conditions where envelope compression

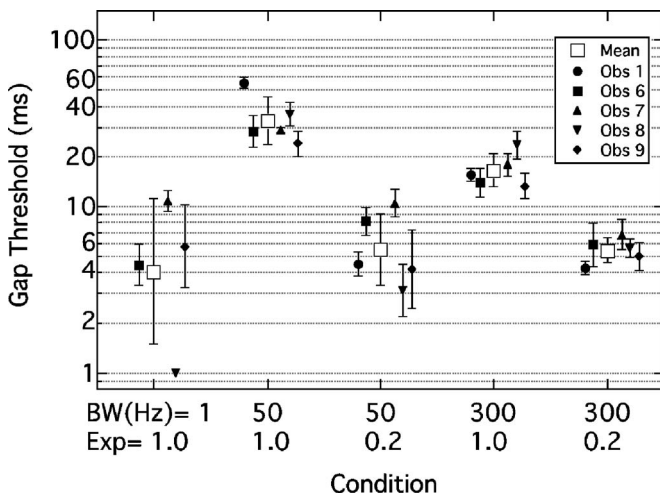


FIG. 5. Individual and group mean GD thresholds for experiment 2. Abscissa gives envelope characteristics imposed on the carrier pulse train: BW=bandwidth of narrow band of noise; Exp=exponent of envelope. Error bars are ± 1 s.d.

was applied [exp=0.2], thresholds were similar to baseline irrespective of the original noise bandwidth. However, thresholds were elevated for both conditions where uncompressed envelopes were applied, and were more elevated for the 50-Hz bandwidth envelope than for the 300-Hz bandwidth envelope. The significance of this data pattern was confirmed using a pair of RANOVAs. The first RANOVA was a single-factor analysis that looked simply at the effect of treatment level (five conditions) on GD threshold. The effect was significant ($F_{4,16}=17.429$; $p < 0.001$), and post-hoc simple contrasts indicated that thresholds in the two conditions with uncompressed envelopes were significantly elevated from baseline ($F_{1,4}=20.726$; $p=0.01$ and $F_{1,4}=9.154$; $p=0.039$ for the 50 and 300-Hz envelopes, respectively); thresholds in the two conditions with compressed envelopes were not elevated. A second two-factor RANOVA assessing the effects of bandwidth and exponent indicated a marginally insignificant effect of bandwidth ($F_{1,4}=7.459$; $p=0.052$), a significant effect of exponent ($F_{1,4}=51.473$; $p=0.002$), and a significant interaction between these two factors ($F_{1,4}=10,312$; $p=0.033$). The interaction was due to the effect of bandwidth being pronounced for the uncompressed envelopes but not for the compressed envelopes.

This pattern of results parallels that seen for acoustic hearing, where GD thresholds are higher for 50-Hz wide bands of noise than for 300-Hz wide bands of noise when no envelope compression is introduced (Eddins *et al.*, 1992). Unlike acoustic hearing, this finding for electric hearing is free of the inherent confound between stimulus fluctuation and stimulus bandwidth. That is, the envelopes associated with different bandwidths of noise are constrained here to a single electrode channel. In this sense, the result is more analogous to the situation in acoustic hearing where the degree of envelope fluctuation is manipulated without appreciably changing the spectrum of the stimulus. By applying envelope compression, or using algorithms to generate noise bands with minimized AM excursions, thresholds can be shown to be lower for acoustic markers with “low-fluctuation” envelopes than for uncompressed, or regular, envelopes (Glasberg and Moore, 1992; Moore *et al.*, 2001; Grose *et al.*, 2007a). Listeners with cochlear implants, therefore, appear to be subject to the same limitations in the processing of temporal gaps as are normal-hearing listeners in terms of the effects of perceptually salient AM.

V. SUMMARY AND CONCLUSIONS

Two experiments were undertaken to examine within- and across-electrode-channel temporal processing in cochlear implant listeners. The first experiment demonstrated that, in bilaterally implanted listeners, gap discrimination for markers carried on same-numbered electrodes, but across ears, is much poorer than for markers carried on the same electrode, but within an ear. The elevated across-ear thresholds are on a par with monaural, across-electrode-channel gap discrimination. This pattern of results is analogous to that seen for across-ear/across-frequency-channel findings in acoustic hearing. The commonality of performance pattern for electric and acoustic hearing supports the notion that the

decline in temporal processing for gap markers that are moved apart in frequency or space is associated with the analysis of temporal discontinuities that occur across perceptual channels rather than across cochlear frequency channels or electrode channels. Subsidiary conditions of experiment 1 also demonstrated that listeners with cochlear implants exhibited a decline in gap discrimination when some form of asymmetry (e.g., number of electrode channels) existed between the gap markers.

The second experiment parametrically varied the degree of envelope fluctuation applied to the marker pulse trains. Envelopes associated with 50-Hz wide bands of noise resulted in higher gap thresholds than envelopes associated with 300-Hz wide bands of noise. However, in both cases gap detection improved to baseline when the envelope fluctuations were compressed. Because all stimuli were restricted to a single electrode channel, these effects of marker AM are not confounded with spread of excitation.

The results of this study as a whole demonstrate that temporal processing of gaps in electric hearing parallels that in acoustic hearing. This supports the interpretation that the limitations of temporal processing for this class of stimuli are determined by mechanisms largely central to the cochlea.

ACKNOWLEDGMENTS

We gratefully acknowledge the comments of Joseph W. Hall III, Associate Editor Ruth Litovsky, and two anonymous reviewers on a previous version of this manuscript. The assistance of Reinhold Schatzer and Marian Zerbi in implementing the RIB interface is also noted with gratitude. This work was supported by NIDCD R01 DC001507.

¹Based on the fact that Observer 1 performed at floor level in the within-electrode-channel condition of experiment 1, which used nominally 60-ms duration markers, it would be expected that floor performance would also have been observed for the baseline condition of experiment 2, which used nominally 125-ms duration markers. Nevertheless, for the purposes of statistical analysis the threshold estimate used for Observer 1 in the baseline condition was the geometric mean of the other four observers; i.e., 4.06 ms.

- Chatterjee, M., Fu, Q.-J., and Shannon, R. V. (1998). "Within-channel gap detection using dissimilar markers in cochlear implant listeners," *J. Acoust. Soc. Am.* **103**, 2515–2519.
- Divenyi, P. L., and Danner, W. F. (1977). "Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra," *Percept. Psychophys.* **21**, 125–142.
- Dobie, R. A., and Dillier, N. (1985). "Some aspects of temporal coding for single-channel electrical stimulation for the cochlea," *Hear. Res.* **18**, 1–55.
- Eddins, D. A., Hall, J. W., and Grose, J. H. (1992). "The detection of temporal gaps as a function of frequency region and absolute noise bandwidth," *J. Acoust. Soc. Am.* **91**, 1069–1077.
- Formby, C., Barker, C., Abbey, H., and Raney, J. J. (1993). "Detection of silent temporal gaps between narrow-band noise markers having second-formantlike properties of voiceless stop/vowel combinations," *J. Acoust. Soc. Am.* **93**, 1023–1027.
- Formby, C., Gerber, M. J., Sherlock, L. P., and Magder, L. S. (1998a). "Evidence for an across-frequency, between-channel process in asymptotic monaural temporal gap detection," *J. Acoust. Soc. Am.* **103**, 3554–3560.
- Formby, C., Sherlock, L. P., and Li, S. (1998b). "Temporal gap detection measured with multiple sinusoidal markers: Effects of marker number, frequency, and temporal position," *J. Acoust. Soc. Am.* **104**, 984–998.
- Glasberg, B. R., and Moore, B. C. J. (1992). "Effects of envelope fluctuations on gap detection," *Hear. Res.* **64**, 81–92.
- Grose, J. H., Buss, E., and Hall, J. W. (2007a). "Gap detection in modulated noise: Across-frequency facilitation and interference," *J. Acoust. Soc. Am.* in press.
- Grose, J. H., Hall, J. W., and Buss, E. (2007b). "Gap duration discrimination for frequency-asymmetric gap markers: Psychophysical and electrophysiological findings," *J. Acoust. Soc. Am.* **122**, 446–457.
- Grose, J. H., Hall, J. W., III, Buss, E., and Hatch, D. (2001). "Gap detection for similar and dissimilar gap markers," *J. Acoust. Soc. Am.* **109**, 1587–1595.
- Hanekom, J. J., and Shannon, R. V. (1998). "Gap detection as a measure of electrode interaction in cochlear implants," *J. Acoust. Soc. Am.* **104**, 2372–2384.
- Heinz, M. G., Goldstein, M. H., and Formby, C. (1996). "Temporal gap detection thresholds in sinusoidal markers simulated with a multi-channel, multi-resolution model of the auditory periphery," *Aud. Neurosci.* **3**, 35–56.
- Hochmair-Desoyer, I. J., Hochmair, E. S., and Stiglbanner, H. K. (1985). "Psychoacoustic temporal processing and speech understanding in cochlear implant patients," in *Cochlear Implants*, edited by R. A. Schindler and M. M. Merzenich (Raven, New York).
- Long, C. J., Carlyon, R. P., Litovsky, R. Y., and Downs, D. H. (2006). "Binaural unmasking with bilateral cochlear implants," *J. Assoc. Res. Otolaryngol.* **7**, 352–360.
- Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (2003). "Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user," *J. Acoust. Soc. Am.* **114**, 1565–1574.
- Moore, B. C., Glasberg, B. R., Alcantara, J. I., Launer, S., and Kuehnel, V. (2001). "Effects of slow- and fast-acting compression on the detection of gaps in narrow bands of noise," *Br. J. Audiol.* **35**, 365–374.
- Oxenham, A. J. (2000). "Influence of spatial and temporal coding on auditory gap detection," *J. Acoust. Soc. Am.* **107**, 2215–2223.
- Penner, M. J. (1977). "Detection of temporal gaps in noise as a measure of the decay of auditory sensation," *J. Acoust. Soc. Am.* **61**, 552–557.
- Perrott, D. R., and Williams, K. L. (1971). "Auditory temporal resolution: Gap detection as a function of interpulse frequency disparity," *Psychonomic Sci.* **25**, 73–74.
- Phillips, D. P., and Hall, S. E. (2002). "Auditory temporal gap detection for noise markers with partially overlapping and non-overlapping spectra," *Hear. Res.* **174**, 133–141.
- Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (1997). "Detection of silent intervals between noises activating different perceptual channels: Some properties of "central" auditory gap detection," *J. Acoust. Soc. Am.* **101**, 3694–3705.
- Preece, J. P., and Tyler, R. S. (1989). "Temporal-gap detection by cochlear prosthesis users," *J. Speech Hear. Res.* **32**, 849–856.
- Rice, S. O. (1954). "Mathematical analysis of random noise," in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York).
- Shannon, R. V. (1992). "Temporal modulation transfer functions in patients with cochlear implants," *J. Acoust. Soc. Am.* **91**, 2156–2164.
- van Hoesel, R. J., and Clark, G. M. (1997). "Psychophysical studies with two binaural cochlear implant subjects," *J. Acoust. Soc. Am.* **102**, 495–507.
- van Wieringen, A., and Wouters, J. (1999). "Gap detection in single- and multi-channel stimuli by LAURA cochlear implantees," *J. Acoust. Soc. Am.* **106**, 1925–1939.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.

Unsteady behavior of flow in a scaled-up vocal folds model

Michael Krane^{a)}

¹Center for Advanced Information Processing, Rutgers University, Piscataway, New Jersey 08854

Michael Barry^{b)} and Timothy Wei^{c)}

Department of Mechanical Engineering, Rutgers University, Piscataway, New Jersey 08854

(Received 29 April 2005; revised 16 October 2006; accepted 18 November 2006)

Measurements of the fluid flow through a scaled-up model of the human glottis are presented to determine whether glottal flow may be approximated as unsteady. Time- and space-resolved velocity vector fields from digital particle image velocimetry (DPIV) measurements of the flow through the gap between two moving, rigid walls are presented in four cases, over a range of Strouhal numbers: 0.010, 0.018, 0.035, 0.040, corresponding to life-scale f_0 of 30, 58, 109, and 126 Hz, respectively, at a Reynolds number of 8000. It is observed that (1) glottal flow onset is delayed after glottal opening and (2) glottal flow shutoff occurs prior to closure. A comparison between flow through a fully open, nonmoving glottis and that through the moving vocal folds shows a marked difference in spatial structure of the glottal jet. The following features of the flow are seen to exhibit strong dependence on cycle frequency: (a) glottal exit plane velocity, (b) volume flow, (c) vortex shedding rates, and (d) vortex amplitude. Vortex shedding appears to be a factor both in controlling flow resistance and in cycle-to-cycle volume flow variations. All these observations strongly suggest that glottal flow is inherently unsteady.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2409485]

PACS number(s): 43.70.Bk, 43.28.Ra, 43.28.Py [AL]

Pages: 3659–3670

I. INTRODUCTION

This paper presents an experimental assessment of the assumption that glottal flow is quasisteady. The study presents temporally and spatially resolved *in vitro* measurements of the velocity field through a model of the time-varying glottis. The focus is on (1) a description of the test rig and how the flow measurements were made, (2) observations concerning the structure and cycle frequency dependence of the glottal jet, and (3) how this behavior indicates unsteady, as opposed to quasi-steady, glottal flow.

The question of whether glottal flow is inherently unsteady has arisen largely from modeling considerations (see, e.g., [Ishizaka and Flanagan, 1972](#); [Story and Titze, 1995](#); [Liljencrantz, 1996](#)). The issue may be stated essentially thus: Does the flow in the glottis respond more or less instantaneously to a change in the transglottal pressure gradient? In other words, is the inertia of air in the glottis negligible compared to that of the airflow in the trachea and the pharynx? Observations of flow through glottis like ducts ([Hofmans *et al.*, 2003](#)) show that the steady-state jet structure downstream of the vocal folds is established during a period longer than a typical vocal fold vibration cycle. This observation strongly suggests that glottal flow never reaches a quasisteady state during phonation, and that the inertia of the glottal jet flow is a controlling factor. [Hofmans *et al.* \(2003\)](#) did caution, however, that, because their model glottis walls did not move, the

conclusions may not fully apply to phonation. On the other hand, other experimental studies in ducts with moving walls [[Mongeau *et al.* \(1997\)](#), [Zhang *et al.* \(2002\)](#), and [Vilain *et al.* \(2004\)](#)—the follow-on to [Hofmans *et al.* \(2003\)](#)] have so far seemed to support the notion that glottal flow inertia is indeed quasisteady. However, as [Krane and Wei \(2006\)](#) pointed out, for normal phonation, glottal flow may be approximated as quasisteady if (1) the separation point does not move and (2) if the inertia of the glottal jet is neglected. They also suggested that the aforementioned studies seem to demonstrate quasisteady, flow because the data reduction was performed implicitly assuming the pressure is uniform downstream of the separation point, which is equivalent to neglecting the inertia of the glottal jet. [Krueger and Gharib \(2003, 2005\)](#) and [Krueger \(2005\)](#) have shown that the pressure downstream of the formation point is indeed not uniform for an unsteady jet, so that its inertia is clearly not negligible. Clearly glottal jet speed changes continuously though the entire cycle. Is its inertia small enough to be neglected? Answering this question is best accomplished by a study of the glottal jet in a model with moving walls, in which the jet flow is characterized with sufficient spatial and temporal resolution.

In vitro experiments are described here, in order to build upon previous work. This approach to studying glottal flow has a long history, having the distinct advantages [compared to *in vivo* experiments like those of [Cranen and Boves \(1985\)](#) or of [Teager \(1980\)](#)] of repeatability, controllability, and accessibility. The pressure measurements made by [van den Berg *et al.* \(1957\)](#) in static models of the glottis established the foundations of the notion that the Bernoulli effect plays a role in vocal fold vibration. [Coker *et al.* \(1996\)](#) and [Mongeau *et al.* \(1997\)](#), using similar life-scale models, measured both

^{a)}Current address: Fluids Research Dept., Applied Research Laboratory, Penn State University, State College, PA 16804.

^{b)}Current address: Sunset Park Preparatory School, Brooklyn, NY 11232.

^{c)}Current address: Department of Mechanical, Aerospace, and Nuclear Engineering, Rensselaer Polytechnic University, Troy, NY 12180.

transglottal pressure and hot-wire measurements of the glottal jet issuing from a cast silicone model of the vocal folds that were driven into motion patterns mimicking those observed in phonation. The latter study demonstrated that for low-frequency vibrations studied (not higher than 80 Hz), flow was quasisteady for roughly 60% of the open portion of the cycle. The unsteadiness in the first and last phases of the cycle were attributed to the effects of starting and stopping vortices near the glottis. [Barney *et al.* \(1999\)](#) also conducted hot-wire and pressure measurements of the flow in a tube downstream of a time-varying shutter, but their interest was less in glottal flow *per se* than in the aeroacoustic effect of interaction between the tube exit and the residue of the glottal jet.

Where air velocity was measured in the studies mentioned, hot-wire anemometers were used. While this technique has many advantages, such as ample time resolution, it also has several drawbacks in the study of a spatially and temporally evolving glottal flow. First, it is an inherently invasive procedure, thus modifying the flow it measures, especially in confined spaces. Second, the measurement is limited to the small region occupied by the wire. For these reasons, it is not possible for a hot-wire to measure snapshots of the entire flow; its spatial structure must be inferred by performing multiple measurements in many locations and ensemble averaging the results. If the jet's structure is unsteady and does not develop repeatably in phase with the wall motion, ensemble averaging must be done carefully to ensure realizations of like behavior are averaged together. In addition, the use of ensemble averaging precludes study of cycle-to-cycle variations in volume flow.

More recently, several different approaches have been taken to characterize glottal flow and assess the degree to which it is quasisteady. [Zhang *et al.* \(2002\)](#) measured the acoustic volume flow far away from the model glottis while transglottal pressure was also measured. The results of this study were interpreted to indicate that the flow through the model glottis is in fact quasisteady, but neither this nor any of the previous studies presented velocity field measurements with sufficient temporal and spatial resolution to address whether glottal jet inertia can be neglected. [Deverge *et al.* \(2003\)](#) and [Vilain *et al.* \(2004\)](#) studied the dynamics of glottal flow using measurements of pressure in a scaled-up model of rigid vocal folds while one of the folds was in motion. These measurements were compared to predictions of the pressure using approximate aerodynamic models. These comparisons showed that, when the glottal gap is small, viscous forces and the unsteady motion of the vocal fold walls predominate. In other words, just after opening and just before closure, the steady Bernoulli equation is not sufficient to predict glottal volume flow. These results are consistent with those of [Mongeau *et al.* \(1997\)](#), which advocated the validity of quasisteady flow in the middle of the cycle, but not during flow initiation and shutoff. [Krane and Wei \(2006\)](#) showed that these intervals of inherently unsteady flow comprised a larger fraction of the cycle as cycle frequency increased. Taken together, these results imply that

frequency dependence of the jet flow in the middle interval of the cycle would indicate that glottal jet inertia is non-negligible.

For the current study, a series of experiments was designed in order to address the above mentioned issues using precise measurements of the velocity field of a flow through a time-varying duct, which mimics the motion of the vocal folds. Digital particle image velocimetry (DPIV) measurements were made of the flow, providing a series of snapshots of the velocity field in the plane of symmetry. The model was scaled up and used water as the working fluid in order to reduce the aerodynamic and vocal fold wall motion time scales so that these motions can be resolved using standard video framing rates. These measurements were taken over several realizations of a single oscillation cycle. Four oscillation frequencies were studied in the expectation that as the oscillation frequency increases, the relative importance of unsteady effects also increases. Three types of information are presented. First, inspection of the velocity fields provides information about the evolution of the flow's spatial structure. Second, the waveform of peak velocity at the glottal exit plane allows an assessment of the relevant dynamics during each phase of the cycle, identification of the relevant aerodynamic time scales, and how these vary with cycle frequency. Finally, the relevance of this information to voice production can be obtained from the volume flow at the glottal exit plane. Specifically, the correlation of the volume flow waveform shape to wall motion and particle velocity, and how these vary with frequency, allows further identification of the aerodynamic mechanisms that are likely to most affect the production of voiced speech sounds.

II. EXPERIMENTAL DESIGN

The experimental setup described here was designed to obtain temporally and spatially resolved measurements of the glottal flow velocity field. Digital particle image velocimetry (DPIV), a noninvasive optical technique, was used to obtain these measurements on the glottal midplane while the model vocal fold walls are in motion. In order to obtain temporal resolution at standard video framing rates (<15 image pairs per second) used by DPIV, it was necessary to use a scaled-up model. The test model was sized using dynamic similarity by matching the Reynolds and Strouhal numbers that occur in real speech. In addition, the model was designed for maximum optical access, as described below.

Two steps were taken to slow the model flow down from that which occurs at life scale. First, the model was scaled up approximately ten times life size, which also improves spatial resolution. Thus, the ratio of model length scale d_M to life length scale d_L is 10. Second, water was used as the working fluid, so the ratio of model kinematic viscosity ν_M to that of air (ν_L) is roughly 1/15. By matching the Reynolds number Ud/ν , where U is the velocity scale and d the minimum glottal width, the model flow speed is 150 times smaller than the life scale flow. By further matching the Strouhal number $L/(UT)$, where T is the time scale of the wall motion and L the glottis axial length, the model time scale is 1500 times that of life scale. For example, in the

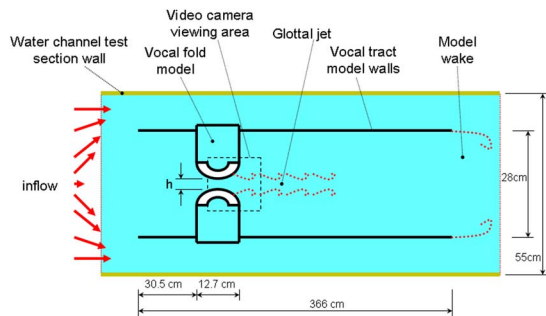


FIG. 1. (Color online) Schematic of test model in water channel, showing real dimensions. Channel flow is set to a steady flow rate and moves from left to right, flowing both around and through the model when the model glottis is open. The dotted line box shows a typical camera viewing area for measurements presented in this article.

scaled up, water model flow equivalent to a 100-Hz life-scale vocal fold vibration (period 0.01 s), the model vocal fold vibration period is 15 s. At a 15 Hz sampling rate, it is then possible to obtain 150 snapshots of the fluid motion during an oscillation cycle.

A. Test model

The test rig consists of an optically accessible model vocal tract, through which two rigid model vocal folds insert and are made to move perpendicular to the tract axis by stepper motors. A sketch of the test rig is shown in Fig. 1. The entire vocal tract model is immersed in the Open Water Channel Facility at Rutgers University (see [Shah et al., 1999](#); [Hsu et al., 2000](#)). The side and bottom walls of the rectangular cross-section water channel test section (6.1 m long \times 55 cm wide \times 1.22 m deep) are 1.27-cm-thick glass panes, allowing optical access. The water channel pumps are set to a constant speed, and the water flows around the test rig—not only around the sides, but also the top and bottom.

By placing the model vocal tract in the steady flow so that the water flows around it, an essentially constant pressure gradient boundary condition is obtained. In other words, the pressures at the model vocal tract inlet and outlet are not only uniform, but a positive pressure difference is also maintained throughout the cycle. The pressure difference arises because the vocal tract model is essentially a bluff body. When the glottis is closed, the flow is deflected around the inlet and a wake is formed. The upstream deflection results in a high pressure at the inlet, while the wake flow ensures a low pressure at the outlet. As a result, a pressure difference between the fluid on either side of the model vocal folds is produced when the glottis is closed, just as in the real system. When the model glottis is opened, not more than 4% of the incident flow passes through the glottis. The glottal flow is then but a small perturbation to the bulk flow around the model, ensuring that the pressure distribution is not significantly altered. What alterations to the pressure distribution do occur are consistent with those that occur in the vocal system—when the glottis opens and flow through the model vocal tract is initiated, the model subglottal pressure accordingly decreases. However, because the bulk flow pattern

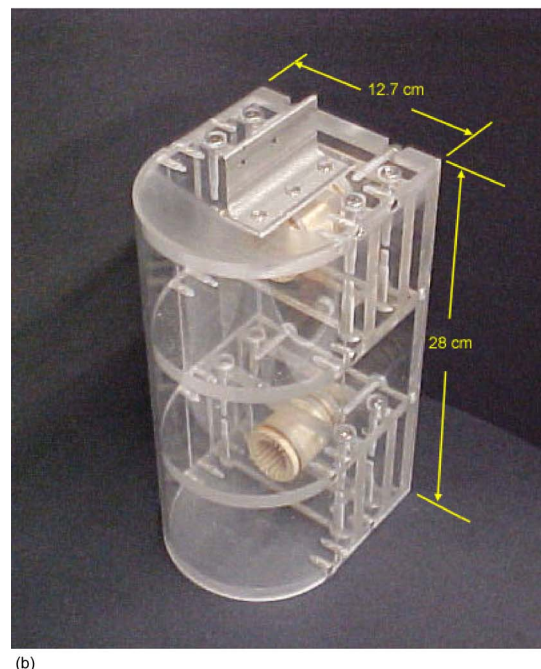
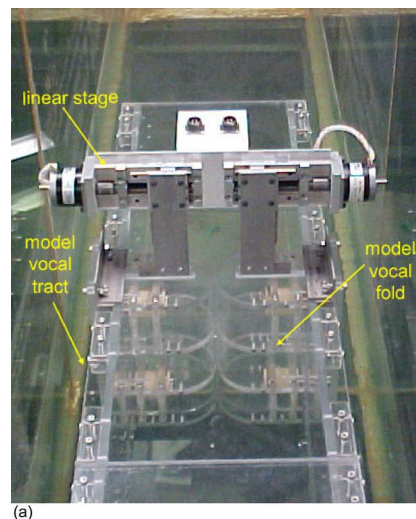


FIG. 2. (Color online) Photos of the test apparatus. (a) The model placed in the water channel test section, looking upstream. Visible are the model vocal folds, the drive mechanism (located above the model), and the glass walls and floor of the water channel test section. (b) One of the model vocal folds. The half-circular shape of the wall is readily visible. Also visible are the linear bearings, which allow smooth motion of the model, and (at top) the angle through which the linear stages transmit momentum to the model vocal fold.

around the model is maintained, the model subglottal pressure is always greater than the supraglottal pressure, maintaining the pressure gradient.

Figure 2(a) shows a photo of the test rig in the water channel, looking upstream. The model vocal folds can be seen, as well as the motor drives located just above the test rig. The test rig was fabricated from clear cast acrylic and is composed of the 28×28 cm square cross section duct 3.66 m in length, and the rigid model vocal folds. The vocal folds are placed 45.7 cm downstream of the test section inlet, so that the distance between the glottal exit plane and the model outlet is approximately 3.08 m.

The model folds were fabricated to a 6.35-cm-radius half-circle shape and measure $14 \times 12.7 \times 28 \text{ cm}^3$ (transverse, axial, and height dimensions, respectively). A photo of a model vocal fold is shown in Fig. 2(b). The target maximum glottal gap used in the experiments was 2.88 cm. Small gaps ($\sim 1.3 \text{ mm}$) were allowed between the top and bottom duct walls and the top and bottom of the vocal fold models to allow free travel and prevent wear due to friction between the walls and the moving model vocal folds. Similarly, small gaps of similar dimension were allowed between the side walls and the vocal fold model.

The experimental design also attempts to preclude artifacts from strongly affecting the flow in the glottal region. First, the water channel is set to run at an incident flow speed of roughly 14 cm/s , slow enough that only standing waves, of less than 1 mm in height, appear on the free surface. Second, it should be emphasized that the model vocal tract was completely immersed, so the water flowed around the top of the model, as well as the bottom and sides. Second, the model vocal tract is suitably long (see below) to ensure that any wake fluctuations at the model outlet have minimal effect on the flow in the model glottis region. Third, flow through the gaps between the vocal fold model and the side, top, and bottom walls was minimized. The importance of the latter two effects was determined using dye visualization and DPIV measurements downstream of the model vocal folds when closed, a condition when both the leakage flow and any upstream effect of the model wake should be maximum. Leakage flow was found to be minimal using visual inspection of dye in these regions using dye. In addition, DPIV measurements of the flow downstream of the vocal folds, when the folds were in the closed position, showed that any velocity fluctuations were no greater than the uncertainty of the DPIV measurement (0.16 cm/s) made with the same settings as used in the measurement of the flow during the period when the glottis is open. In other words, velocity fluctuations downstream of the closed model vocal folds due to leakage flow or the upstream effect of the model wake are less than 0.5% of the maximum velocity (28 cm/s) observed when the vocal folds are open.

The model vocal folds are moved using Servo Systems[®] stepper motors and linear stages such that a single motor step supplies a 0.003175 mm linear displacement of the vocal folds. A 1.6-mm-thick by 7.62-cm-long stainless steel strip attaches the linear stage to the vocal fold at the stainless steel angle visible in the photo of the model vocal fold. For motion stability, the vocal folds are mounted on two linear bearings, each of which slides over a 2.54-cm-diam stainless steel rod attached to the outer wall of the test section. Because of a slight clearance between the bearings and the mounting rods, the model vocal folds rocked slightly when moved by the motors. As discussed below, this rocking caused some frequency dependence of model vocal fold motion.

B. Data acquisition and reduction

DPIV measurements of the flow velocity field were performed during single realizations of one cycle of motion of

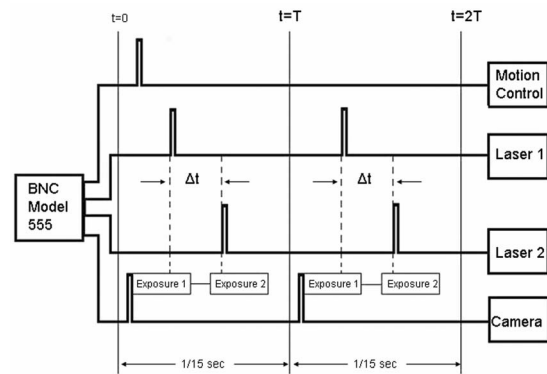


FIG. 3. Schematic of experiment control using waveform generator. A new acquisition interval begins every 15 Hz. The waveform generator sends pulses to the camera and lasers to provide proper timing of illumination of the two images in an image pair. During the first cycle, a pulse is sent to the motion controller to initiate the motion of the model vocal folds.

the model vocal folds. At each acquisition time step, a velocity field estimate was obtained from a pair of images acquired at a known time delay, typically $1\text{--}1.6 \text{ ms}$. The flow was seeded by silver-coated hollow glass spheres (Conducto-Fil, manufactured by Potter Industries, mean diameter $\sim 10 \mu\text{m}$). In each 1018×1008 pixel image the seed particles were illuminated by a single 150-ns illumination pulse from a NewWave Research Solo 120PIV dual Nd:Yag laser. Velocity field estimates were performed by applying a cross-correlation technique (see below) to two successive images. Image pairs were acquired at 15 Hz using a Redlake Mega-plus ES-1.0 digital video camera (8-bit resolution) operating in triggered double exposure mode.

Image acquisition, illumination, and model vocal fold motion control were coordinated using a Berkeley Nucleonics BH-555 waveform generator, as shown in Fig. 3. In the figure, T is the period of the image pair acquisition cycle ($1/15 \text{ s}$) and Δt is the time delay between acquisition of each image in the image pair. This time delay varied from case to case ($1\text{--}1.6 \text{ ms}$) but was chosen so that particle displacements were no greater than 8 pixels. In each acquisition cycle, two trigger pulses are sent to the laser, and a single pulse is sent to the camera. The camera needs only a single pulse when operating in triggered double exposure mode, in which the camera shutter opens twice in an acquisition cycle. The laser triggering is timed so that one each of the 150-ns laser pulses illuminates the field during the two shutter openings. Motion control of the model vocal folds was synchronized to the DPIV acquisition using the BH-555 waveform generator, which sent a trigger pulse to the motion controller during the first image pair acquisition cycle. To ensure that the whole flow evolution in a single glottal cycle was captured, the motion controller was set to wait for 200 ms after image acquisition began before initiating the motion.

Estimation of the velocity field was obtained from image pairs using a two-pass cross-correlation technique (Hsu, 2000; Hsu *et al.*, 2000). The first pass used a 128×128 pixel interrogation window to obtain the average displacement for the window, to the nearest pixel. The second pass used a smaller interrogation window of 64×64 pixels to fine tune the estimate to subpixel precision. Images were cropped

prior to processing so that they were centered vertically about the axis of symmetry. The processing program output provides, for each time step, a representation of the velocity vector and vorticity fields in the plane of the light sheet. After cropping the image, the measurement region is discretized into a 58×59 grid on which the velocity and vorticity are computed from the cross correlation of the illuminated particle patterns between the two images in each image pair.

III. RESULTS

A. Cases studied

Because this study is concerned with the frequency dependence of the flow, measurements were taken for four values of open time T_o . Dynamic similarity was obtained by matching Reynolds number of 8000, based upon a life-scale jet speed of 40 m/s, a glottal gap of approximately 3 mm, a kinematic viscosity of 1.5×10^{-5} m²/s for air, a Strouhal number $L/(U2T_o)$ based upon the same flow speed, a glottis length of 1.3 cm, and the vocal fold oscillation period. Note that $f=1/(2T_o)$, assuming that the vibration period $T=2T_o$. In other words, for purposes of constructing a life-scale equivalent frequency, the open quotient (fraction of cycle that glottis is open) is taken to be 0.5. The four frequencies studied, in life scale, are $f_{1s}=30, 58, 109,$ and 126 Hz, corresponding to $St=0.01, 0.018, 0.035,$ and 0.040 , respectively. This range of frequencies allowed acquisition of from 124 to 512 image pairs per open phase to be taken, depending on the oscillation frequency studied. For each cycle frequency studied, ten realizations of a single cycle of wall motion were acquired.

The target wall motion was a uniform speed of opening, then closing, the maximum opening occurring at halfway through the open phase of the cycle. The acceleration of the motors was set to the minimum allowed by the hardware, 1.27 cm/s² (4000 steps/s²). Figure 4 shows the wall motion for the four St studied: the closest distance between the centerline and each model vocal fold, normalized by the target maximum distance, versus the percent time the glottis is open. The lowest frequency case most closely approximates a straight-line motion, with some rounding at opening, as the walls reverse direction, and just before closing. This rounding becomes more pronounced as the cycle frequency increases. In addition, it can be observed that the opening is not entirely symmetrical, and that the asymmetry is greatest for the slowest motion. The dependence of the wall motion on cycle frequency is due to the slight clearance between the bearings and the mounting rod, as described above. When the vocal fold motion is initiated, or when it changes direction, a slight rocking occurs, and this rocking is manifested differently for different wall motion speeds.

As a reference case, the steady-state flow through a fully open glottis was measured. The steady state flow was allowed to develop for several minutes after the glottis was opened. A jet flow was observed to cling to one or the other wall, occasionally switching sides, due to the Coanda effect. The maximum speed in the narrowest part of the glottis was measured to be 28 cm/s.

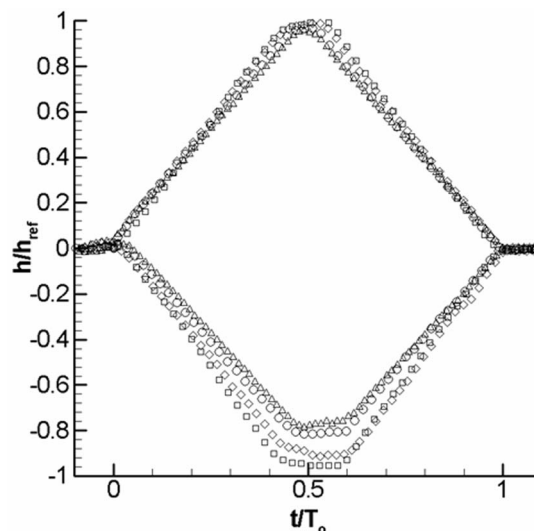


FIG. 4. Wall motion for the four cases of oscillation frequency studied. Glottal width, h , normalized by maximum glottal width, h_{\max} , is plotted versus time, t , from opening, normalized by time, T_o , glottis is open. Frequency is expressed nondimensionally as Strouhal number (St) and as life-scale frequency (f_{1s}). Δ : $St=0.010$, $f_{1s}=30$ Hz; \circ : $St=0.018$, $f_{1s}=58$ Hz; \diamond : $St=0.035$, $f_{1s}=109$ Hz; and \square : $St=0.040$, $f_{1s}=126$ Hz.

B. Evolution of spatial structure of flow

Figure 5 shows a typical set of images from the $St=0.035$ case, corresponding to a life-scale fundamental frequency of 109 Hz. Each of the four images shown is composed of the measured velocity field superimposed on the corresponding raw image so that the spatial relationship between the jet structure and model vocal folds may be observed at each time step. In each image the flow is from left to right. For clarity, not all vectors are shown. (Note that in some images, vectors appear as though the flow moves through the model walls. This is an artifact of the manner in which the velocity vectors are estimated: the vectors are placed at the center of the fine interrogation windows—see Sec. II B. When a correlation window contains mostly wall and only a small amount of flow, the correlation technique correlates on the particle motion in the flow, but places the vector at the center of the window, even if that location is inside the model. In this paper, we largely discuss the motion on the exit plane, where this artifact does not occur.)

Each vector field shown was chosen as an exemplar of an event common to all realizations acquired. The first image (a) corresponds to the opening phase ($t/T_o=0.15$), in which, soon after the glottis opens, a small puff of fluid is released in the form of a weak vortex pair, visible in the velocity disturbance just to the right of the minimum glottal opening. During this phase, a weak suction of fluid into the glottis is evident, along with a retardation of the convection speed of the starting vortex. A jet forms and follows behind the starting vortex, initially increasing in speed and size. The maximum jet width occurs halfway through the oscillation cycle [Fig. 5(b)]. From this point, until roughly three quarters of the way through the time the glottis is open [Fig. 5(c)], the jet speed increases rapidly, then begins to decrease until the flow shuts off *prior to closure* roughly 90% of the way through the open phase. After the jet flow shuts off in this

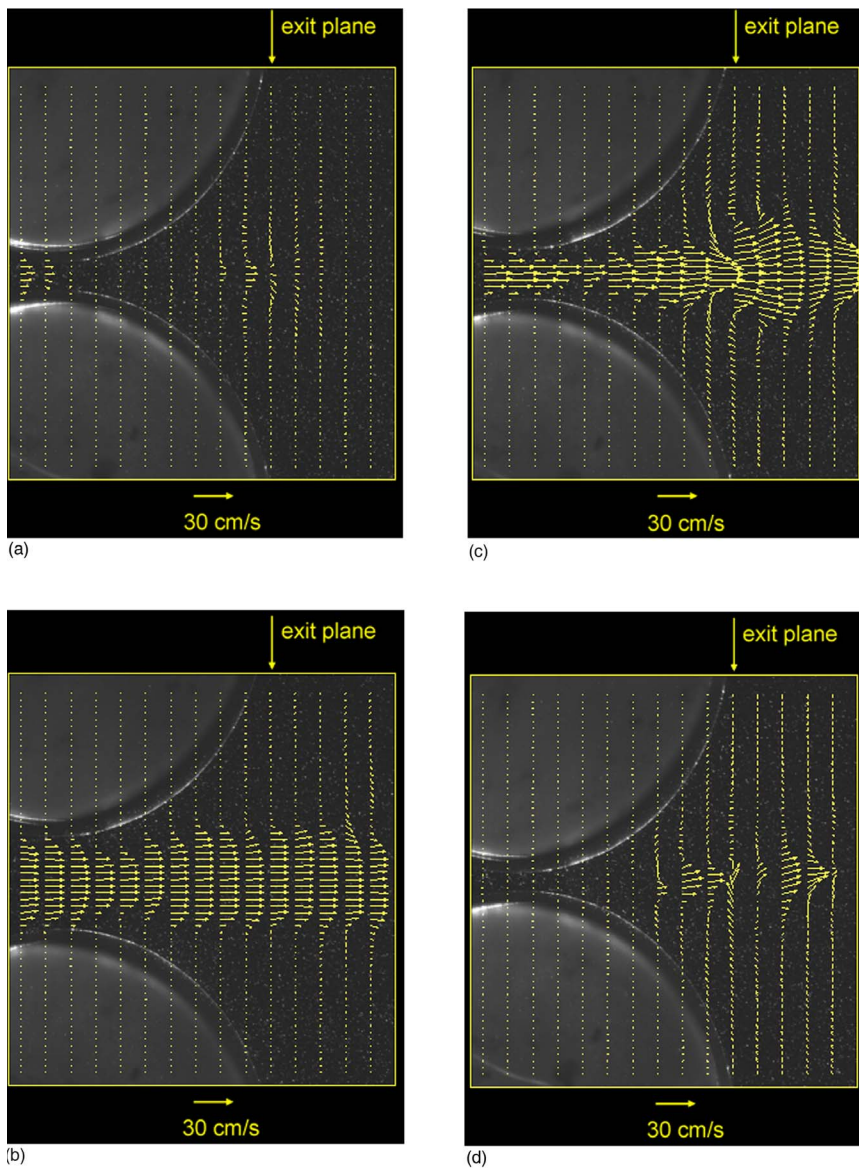


FIG. 5. (Color online) Images of jet velocity field from DPIV measurements of a single realization, superimposed on corresponding image of model vocal folds during motion, $St = 0.035$ ($f_{1s} = 109$ Hz). Arrow length is proportional to local flow speed. Time of each image is given as fraction of time, T_o , that the glottis is open: (a) $t/T_o = 0.15$, $h/h_{max} = 0.33$, starting vortex visible; (b) $t/T_o = 0.50$, $h/h_{max} = 0.99$, glottis fully open; (c) $t/T_o = 0.76$, $h/h_{max} = 0.58$, jet at maximum speed; (d) $t/T_o = 0.93$, $h/h_{max} = 0.18$, flow has shut off and jet is in decay phase.

manner, the remains of the jet convect downstream, mixing in an increasingly three-dimensional motion with the surrounding stagnant fluid, as shown in Fig. 5(d).

Several features of flow structure are immediately observable. First, glottal flow separates at roughly the same location, just downstream of the minimum constriction point. Second, starting at the point of minimum glottal width and moving toward the glottis exit plane (right to left in the images), a *vena contracta* is observed. Third, downstream of the minimum jet width, the jet shear layers begin to roll up into vortex pairs. For any sequence, the rate of vortex formation is initially slow, and the amplitude of the vortex pairs decreases as the flow reaches its plateau. As the walls begin to move together ($t/T_o > 0.5$) the vortex formation rate increases and the vortices that form become more intense. In addition, the number of vortices formed in a sequence increases with T_o . This behavior reflects the expectation that the rate of vortex formation is expected to scale more strongly on the local flow conditions: speed, acceleration, and shear layer thickness. Fourth, the shape of the velocity profile in the jet changes from a relatively flat shape in the

early stages of the cycle, then becoming more “fully developed” in the later stages. This suggests that the viscous resistance of the glottis is different during opening than it is during closure, and that friction plays a stronger role during closure than during opening. The final observation from Fig. 5 is that, strictly speaking, the flow is not turbulent, at least in the field of view shown here, though flow visualization indicates that the jet becomes turbulent further downstream.

Figures 6(a)–6(d) show snapshots of the velocity field at $t/T_o = 0.75$ for each of the four St studied. In these images, all vectors are shown in order to see more clearly the vortex pair structure. The spatial structure of the jet flow is roughly similar for all cases. Of particular interest is the comparison of Figure 6(a) with the rest—the jet flow is asymmetrical only in this case, but not nearly as asymmetrical as the steady flow case. It should be pointed out that out of ten realizations, only this one showed this behavior. In any case, comparison of steady and unsteady flow topology suggests that indeed the time for the jet flow to reach a state where it

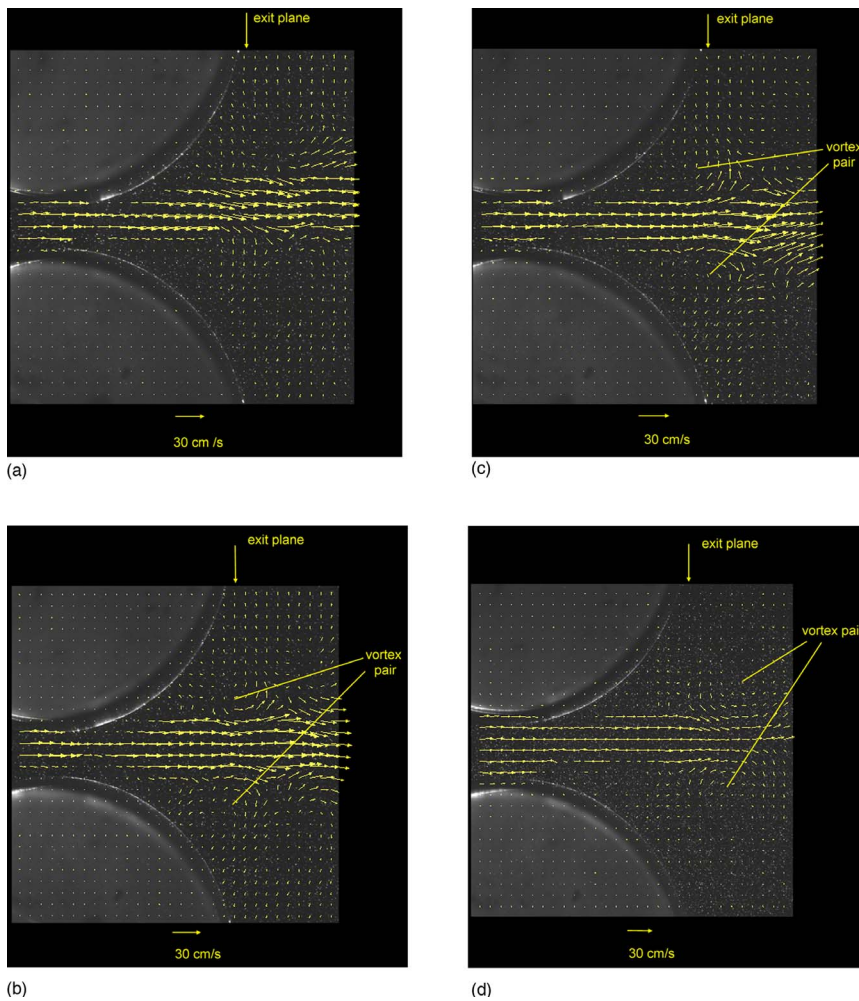


FIG. 6. (Color online) Images of jet velocity field from DPIV measurements superimposed on corresponding image of model vocal folds at $t/T_o = 0.75$ for all St. (a) $St=0.010$, (b) $St=0.018$, (c) $St=0.035$, and (d) $St=0.040$. Vortex pairs are visible in each image. In the $St=0.017$ case, the jet is seen to divert from the centerline.

fully “clings” to one or the other glottal wall is longer than a realistic glottal cycle period, as suggested by Hofmans *et al.* (2003).

C. Behavior of velocity maximum on glottal exit plane

Figure 7 shows waveforms of the maximum axial velocity at the exit plane for $St=0.035$, normalized by the velocity measured in the same location with the glottis fully open and the flow in a steady-state condition. The maximum velocity is considered because it generally denotes the dynamic behavior of the jet. If the centerline velocity was used in an asymmetric jet, it would reveal much less about the jet’s evolution. Of concern here are (1) the correlation of the velocity waveform shape with the behavior observed in the vector fields presented above, (2) the variability of the waveform shape, and (3) frequency variations in waveform shape.

Waveforms for a single realization, a low-pass filtered version of that realization, and the ensemble average of the ten realizations measured are shown in Fig. 7. Filtering was performed by convolving a nine-point triangle window with the original signal. The major difference between the three waveforms lies in the level of the high-frequency fluctuations. In the unfiltered single realization, each of the peaks is associated with the passage of a vortex pair through the exit plane. In the low-pass filtered waveform, these peaks are smoothed. Note that, in the ensemble averaged waveform,

vortex-associated fluctuations have largely disappeared, with the exception of that associated with the starting vortex and the vortex pairs that appear during the interval when the velocity is maximum ($0.6 < t/T_o < 0.8$). While only ten realizations are clearly not enough to converge an ensemble average, this result is suggestive that cycle-to-cycle fluctuations in vortex formation intervals are smallest for the starting vortex and during the interval when vortex formation is strongest.

The first feature of exit velocity behavior that should be noted is that the moment of flow onset does not coincide with the moment at which the glottis opens ($t/T_o=0$). The delay in all cases is on the order of several percent of T_o . Once the flow onset occurs, the flow rapidly accelerates over a brief interval until it reaches a peak, corresponding to the passage of the starting vortex through the exit plane. After this passage, the velocity rapidly decreases, then rises again as the next vortex passes through the exit plane. The absolute velocity during each subsequent vortex passage is higher than the previous until a plateau level is reached at roughly $t/T_o=0.45$. The fluctuations due to vortex pair passage are reduced during this plateau phase, which is more or less maintained until around $t/T_o=0.6$, after the maximum glottal width is reached ($t/T_o=0.5$). At this point in time the flow accelerates due to a combination of the narrowing glottis and the fact that the flow throughout the model vocal tract has

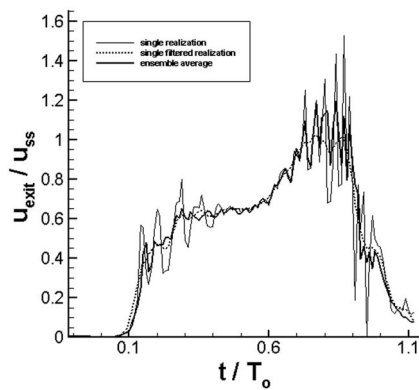


FIG. 7. Maximum axial velocity component at glottal exit plane, $St = 0.035$: comparison of (thin line) raw signal, (thick line) low-pass filtered signal, and (dotted line) ensemble average of ten realizations. Maximum jet speed on exit plane, u_{exit} , is normalized by maximum jet speed, u_{SS} , observed in steady flow through fully-open glottis. Time, t , from opening, is normalized by time, T_o , glottis is open.

finished its startup transient and has inertia. As the glottis narrows, the velocity increases until roughly $t/T_o = 0.9$, at which point the flow undergoes a rapid deceleration. As can be observed, the vortex pairs formed in this phase of motion are formed more rapidly and are more intense than previously, as shown in Figs. 5 and 6. Note that the exit plane velocity does not go to zero simultaneously with glottal closure because the residue of the jet is still moving away from the vocal folds, even after the glottis is closed, as seen in Fig. 5(d).

The above trends appear consistent with those observed in the hot-wire measurements of flow through (1) a canine larynx *in vivo* (Berke *et al.*, 1989), (2) an excised canine larynx (Alipour *et al.*, 1995), and (3) an *in vitro* model of the glottis (Mongeau *et al.*, 1997). Several discrepancies with these previous results are evident, however. First, a double peak is sometimes observed in the previous work, likely due to flapping of the jet, which may be a consequence of probe interference of the flow. Second, the temporal shape of the waveform is time reversed from that of Alipour *et al.* (1995). Third, the magnitude of the peak velocity, relative to that observed during the remainder of the cycle, is larger in the present experiment than previously observed. It is possible that compressibility plays a more important role during closure in the real glottis than previously understood. In other words, during closure, the incoming air, by virtue of its inertia, causes an accumulation of mass in the subglottal region. In this case, the subglottal pressure is split between driving the transglottal flow and compressing the subglottal air. In the experiment described here, this effect would be reduced because of the relative incompressibility of the working fluid (water), resulting in a higher jet velocity peak. However, it should be pointed out that these previous measurements were taken much farther downstream than the measurements described here, so that the jet, having mixed with the surrounding stagnant fluid, simply has a lower peak velocity. These assertions clearly require further investigation.

Figure 8 shows waveforms of single realizations of maximum jet speed, again on the exit plane, as a function of

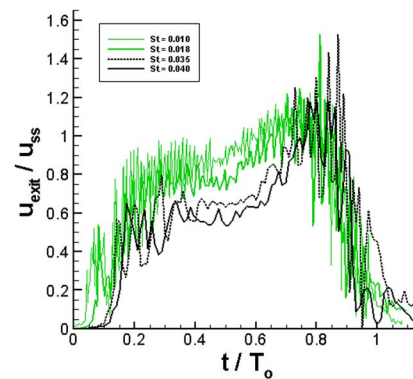


FIG. 8. (Color online) Single realizations of maximum axial velocity component at glottal exit plane versus t/T_o for four cycle frequencies studied. Maximum jet speed on exit plane, u_{exit} , is normalized by maximum jet speed, u_{SS} , observed in steady flow through fully-open glottis. Time, t , from opening, is normalized by time, T_o , glottis is open. Frequency is expressed nondimensionally as Strouhal number (St) and as life-scale frequency (f_{1s}): (thin gray line) $St=0.010$, $f_{1s}=30$ Hz; (gray solid line) $St=0.018$, $f_{1s}=58$ Hz; (black dotted line) $St=0.035$, $f_{1s}=109$ Hz; (black solid line) $St=0.040$, $f_{1s}=126$ Hz.

normalized time, for the four wall cycle frequencies studied. While the overall shapes of the waveforms are similar, the occurrences of the events described above are frequency dependent. The most obvious is the rate of vortex formation: the number of vortices shed per unit cycle decreases with frequency, suggesting that the vortex formation rate does not correlate as strongly to cycle frequency as to the local conditions at the jet formation point: jet speed, local acceleration, and shear layer thickness. Vortex strength also appears to increase with cycle frequency. In addition, several events described above occur earlier in the cycle for lower frequencies: flow onset, reaching the plateau level, acceleration after the glottis begins to close, and the time the maximum is reached. As a result, the normalized waveforms do not collapse, except during their final deceleration phase. However, the maximum value reached in each waveform shows little if any frequency dependence.

These trends do not seem to follow what would be expected if the flow were quasisteady, which, according to Mongeau *et al.* (1997) and Krane and Wei (2006), will only occur in the middle of the cycle, not in the startup or shutoff transient. In that case, the normalized waveforms would be expected to collapse in the middle intervals, but not during the initial and final intervals. Instead, only the shutoff interval shows a collapse, although the modulation of the waveform due to vortex shedding has a strong dependence on cycle frequency. Furthermore, the fact that the vortex behavior does not collapse upon scaling the velocity waveform in the manner presented here also demonstrates that the jet is inherently unsteady.

The mechanism responsible for frequency dependence in the middle interval is partially explained by the inertance of the flow in the vocal tract upstream and downstream of the glottis. This might explain why the jet speed plateau level decreases with frequency: as frequency of vibration increases, less time is available for subglottal fluid to accelerate, under the action of an identical pressure gradient, while the glottis is opening. However, if that were the whole ex-

planation, then the jet speed would not likely reach a plateau, but would increase until it reached the maximum. In addition, the maximum velocity should also vary with frequency. The measured behavior does not show this trend. Another mechanism is likely involved, perhaps relating to the unsteady development of the glottal jet. This is especially evident in the relation between the passage of the starting vortex through the exit plane and the occurrence of the plateau. This coincidence suggests that formation of the starting vortex modulates volume flow, and hence determines glottal resistance during the startup transient.

Finally, it should be noted that the fluctuations due to vortex formation show the most pronounced cycle-to-cycle variations in both amplitude and phase during the final deceleration near the end of the cycle. Air motion in this phase of the cycle contributes most to power and quality of the voice, suggesting that aperiodic vortex shedding may, in fact, modulate glottal volume flow during this perceptually relevant interval.

D. Volume flow waveform behavior

Volume flow per unit span, obtained by integrating the velocity distribution on the glottal exit plane, is shown in Fig. 9 for the four values of Strouhal number studied. As in the jet speed waveforms, the time coordinate is normalized by the total time the glottis is open, while the volume flow is normalized by the volume flow measured for the fully open, steady state case. As with the jet speed waveforms, the short-time waviness is attributed to vortices passing through the glottal exit plane.

The volume flow waveforms in Fig. 9 all have a similar shape. First, a weak negative volume flow is observed due to the suction of fluid while the vocal folds are opening and the through-flow has not yet been established. The onset of the through-flow is delayed from the opening of the glottis. Once established, the volume flow rises to a maximum that occurs after the glottis begins to close, and then a decline to zero before glottal closure, followed by a weak negative overshoot. Frequency dependence can be observed in the slope of the curve and the time the volume flow maximum level is reached. The slope is higher for low St , corresponding to the maximum occurring earlier in the cycle with lower St . These trends result in the familiar skewing of the volume flow waveform. The maximum value of volume flow shows a strong inverse frequency dependence, rising from 0.78 for $St=0.040$ to 0.87 for $St=0.010$. It should be pointed out that the variation of volume flow peak level across frequency is of the same order of magnitude as the maximum glottal opening variation, shown in Fig. 4. The two effects may thus be related. However, results from a study in which St was varied by changing the channel flow speed, but holding the wall motion speed constant (Barry, 2005), show frequency dependence trends similar to those reported here. This study will be the subject of a future paper.

The occurrence of nonzero flow after glottal closure needs some explanation. The flow in this interval is becoming three-dimensional, given that the cause of two-dimensionality (the through-flow) has disappeared once the

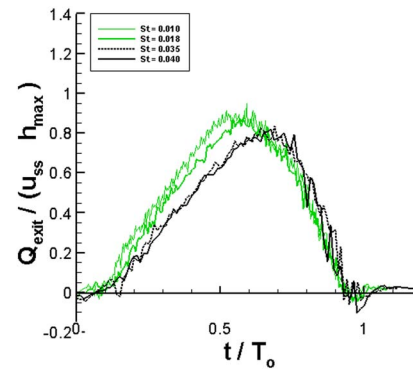


FIG. 9. (Color online) Ensemble-averaged glottal exit plane volume flow versus t/T_o for four cycle frequencies studied. Glottal volume flow, Q_{exit} , is normalized by the product of the maximum velocity u_{SS} observed in steady flow for a fully open glottis and the maximum glottal width during the cycle. Time, t , from opening, is normalized by time, T_o , glottis is open. Frequency is expressed nondimensionally as Strouhal number (St) and as life-scale frequency (f_{1s}): (thin gray line) $St=0.010$, $f_{1s}=30$ Hz; (gray solid line) $St=0.018$, $f_{1s}=58$ Hz; (black dotted line) $St=0.035$, $f_{1s}=109$ Hz; (black solid line) $St=0.040$, $f_{1s}=126$ Hz.

model glottis has closed. In such a case, the remnants of the jet undergo turbulent diffusion, an inherently three-dimensional process. It should be emphasized that the waveforms in Fig. 9 are not volume flow, but volume flow per unit span, computed from the data on a single plane of the flow. As a result, when the flow contains out-of-plane motion, volume flow per unit span can go negative. It is clearly not the result of a complete back-flow through the entire glottis, which would occur because the transglottal pressure difference switches sign. In the experimental setup used here, the flow around the vocal tract model would have to contrive a way to generate thrust, unlikely since bluff bodies in a uniform stream are unable to do so.

Perhaps the most surprising behavior shown in Fig. 9 is that the flow shuts off prior to glottal closure, and that this premature shutoff occurs at the same part of the cycle for each cycle frequency studied. The lack of frequency dependence of the volume flow is also surprising. It might be expected that the shutoff flow has a similar frequency dependence to the flow onset, as argued by Krane and Wei (2006). This behavior is, however, consistent with the collapse of the exit plane velocity waveforms (Fig. 8) during the final deceleration of the flow during glottal closure.

A final point concerning the volume flow waveforms echoes the final point made above about the particle velocity behavior: the fluctuations due to vortex formation show the most pronounced cycle-to-cycle variations in both amplitude and phase during the final deceleration near the end of the cycle. Again, this demonstrates that aperiodic vortex shedding does, in fact, modulate glottal volume flow during this perceptually relevant interval.

Whether the results presented here hold in all respects for phonatory flows is certainly an open question, given the differences in vocal fold wall behavior between the real and the model vocal folds. These differences were also present in the models used by Mongeau *et al.* (1997), Barney *et al.* (1999), Zhang *et al.* (2002), Devere *et al.* (2003), and Vilain

et al. (2004). Clearly similar measurements performed in a mode in which the walls move in a manner more consistent with the real vocal folds are warranted.

IV. DISCUSSION

The measurements presented above provide evidence that glottal flow is inherently unsteady. The glottal jet generally flows down the center of the passage, except for the lowest frequency studied, in contrast to the steady state, where the jet spent most of its time attached to one wall or the other. The flow does not start until well after the glottis is open. The flow shuts off before the glottis closes. The waveforms of particle velocity and volume flow show that the flow is strongly modulated by vortex shedding, and that this modulation varies with frequency. The low-frequency trends of these waveforms also show strong frequency dependence.

A. Frequency dependence of glottal flow

The unsteadiness of the startup flow was observed by *Mongeau et al.* (1997) and attributed to the action of the starting vortex. The starting vortex observed in measurements presented here was relatively weak, compared to those that followed. The starting vortex acts in combination with (1) the inertia not only of the flow in the model glottis, but that upstream of the vocal folds, and (2) the displacement of fluid by motion of the model vocal fold walls. The first of these would result in the occurrence of flow onset being proportional to the cycle frequency. In other words, the flow onset occurs at an earlier part of the cycle as the cycle frequency decreases. The latter effect is expected to be important when the glottis is narrow (*Deverge et al.*, 2003; *Vilain et al.*, 2004; *Krane and Wei*, 2006). The percentage of the cycle during which both effects are important is proportional to the Strouhal number (*Krane and Wei*, 2006), which is consistent with the measurements presented here.

The startup transient, as well as the middle interval between onset and the maximum flow, also appears to be strongly modulated by vortex shedding. In a sense, this is related to the inertance of the flow downstream of the glottis—the glottal jet. The amount of flow that can be pushed through the glottis for a given transglottal pressure gradient depends on the manner in which the fluid downstream of the glottis is accelerated. The glottal jet flow downstream of the glottis is divided into the high-momentum jet and the relatively quiescent ambient fluid. Because of transverse diffusion, the jet entrains ambient fluid by dragging it along. The rate at which this occurs depends on the flow topology. A purely laminar flow will diffuse slowly due to viscosity, i.e., molecular collisions, alone. If the jet shear layers roll up into vortices, or are turbulent, then the diffusion process is accelerated, and more ambient fluid is dragged along with the jet. Clearly, it takes time to accelerate not only the fluid in the jet, but the ambient fluid as well. Just as clearly, this process, depending as it does on the history of jet structure, is more complex than the acceleration of the fluid upstream of the glottis, which is well understood. This process is currently under investigation.

The part of the cycle after the startup transient and the maximum velocity is generally thought to be quasisteady (*Mongeau et al.*, 1997; *Vilain et al.*, 2004; *Krane and Wei*, 2006). The results here suggest that it may not be due to its strong frequency dependence. However, quasisteadiness implies that the instantaneous transglottal pressure gradient (not measured here) and the instantaneous velocity are related through the steady Bernoulli equation. Determining whether the frequency dependence of the flow observed in the current experiments is due to the time to accelerate the flow upstream and downstream (even with modifications to account for the unsteady behavior of the jet) will be the subject of future work.

B. Shutoff flow

Perhaps the most striking thing about the behavior of the flow between the maximum velocity and shutoff is that, in the ensemble mean, it appears to have at most a weak dependence on cycle frequency. This behavior is unexpected—the shutoff transient might have been expected to have the same frequency dependence as the startup transient, if it were dominated by the motion of the vocal fold walls, for example (*Deverge et al.*, 2003). In this case the flow would exhibit the same sort of frequency dependence as the startup transient (*Krane and Wei*, 2006). The fact that the flow does not display frequency dependence suggests that the flow in this interval is governed by processes that are in phase with the motion of the walls.

Deverge et al. (2003) showed how viscous blockage of glottal flow may restrict the amount of fluid a given subglottal pressure can pump through the glottis. If the boundary layers on the vocal fold walls are quasisteady (for discussion, see *Krane and Wei*, 2006), then the degree to which the boundary layers restrict glottal flow will be in phase with the wall motion. In this case, this mechanism may in fact govern the shutoff flow.

However, the hypothesis that viscous blockage governs the flow during shutoff may not completely explain why the flow seems to shut off prior to glottal closure. Premature flow shutoff may also be aided by the increased pressure in the glottis where the folds collide upon closure, as shown in the results of *Deverge et al.* (2003). However, this effect should display the same type of frequency dependence as the startup transient (*Krane and Wei*, 2006). Because the shutoff behavior appears frequency independent here, friction may, in fact, predominate.

Another possible mechanism for premature jet shutoff relates to the mismatch of momentum between the jet and the flow external to it. In a steady jet flow, a recirculation of the “stagnant” region surrounding the jet occurs. Recirculation is driven by turbulent and/or molecular diffusion of jet momentum, which causes the jet to drag essentially stagnant fluid along with it. This entrainment is strongest near the separation point. Fluid from downstream must be pulled upstream along the walls in order to replace that dragged along with the jet. As the jet injects more momentum into the supraglottal cavity, the recirculation gains more momentum. During closure, the jet momentum flux decreases rapidly. If

this decrease is not matched by a comparably rapid decrease in recirculation momentum, the recirculation flow may actually act to pinch off the jet before the glottis closes completely. Observations of the flow during glottal closure also indicate that the fluid external to the jet moves with the vocal fold walls as they close, which may enhance this effect. This additional mechanism for jet shutoff is consistent with and complementary to those proposed by [Deverge *et al.* \(2003\)](#), and is advanced here as a supplement to them.

While the ensemble mean volume flow shows little or no frequency dependence, the variability of vortex shedding in the final stages of the shutoff is clearly an inherently unsteady process. Part of the unsteadiness during this interval is likely due to the weakening of the jet, whose two-dimensional structure dominates the flow while it is strong. During the final moments of glottal flow during a cycle, the velocity fluctuation due to the vortex pairs is at its strongest level compared to the ensemble-averaged velocity, so their inherently unstable dynamics will dominate the flow.

Finally, the volume flow is clearly modulated by vortex shedding, and this inherently unsteady process exhibits perhaps the strongest frequency dependence of all the phenomena discussed here. As the cycle frequency increases, the number of vortices per cycle decreases, and the vortex strength increases. Vortex shedding also impacts voiced sound production through the interaction of the jet with the pharyngeal walls, the source of “breathy” sounds (see [Krane, 2005](#)). Clearly, the fact that neither of these contributions scales proportionally to cycle frequency suggests that the glottal jet makes an important contribution to voice perturbations and fluctuations.

V. SUMMARY

Experiments using a test rig that allows time- and space-resolved measurements of flow velocity field in a time-varying model glottis have been presented. The model is scaled up ten times and uses water as the working fluid in order to reduce the model flow time scales by a factor of 1500, allowing good time resolution of the flow at the 15-Hz sampling rate characteristic of the digital particle image velocimetry (DPIV) technique used to measure the velocity field.

DPIV measurements were made for four wall motion cycle frequencies. These frequencies range from (in life scale) 30 Hz to 126 Hz, thus bracketing the normal adult male speaking voice. Inspection of sequences of the velocity vector fields show that for up to the first roughly 10% of the open phase, the glottal flow is inherently unsteady. Once the vocal folds begin to close, the combination of closure and flow inertia combine to accelerate and narrow the jet until roughly 75% of the way through the open phase. During this interval, vortex shedding increases both in the intensity of the vortices shed and in the rate of shedding. After the maximum jet speed is reached, the jet rapidly decays in speed, and the timing of vortex pair formation becomes less synchronized to the wall motion. Finally, the jet shuts off at

approximately 90% of the way through the open phase, after which the decaying jet remnants are still present for some time.

The most important findings reported here are as follows:

- (1) Comparison of the spatial structure of a steady jet flow and the flow that occurs during wall motion shows that the steady state configuration is not reached during a phonation cycle, although some jet asymmetry was observed at the lowest frequency wall motion, supporting the notion that a steady-state glottal jet flow downstream of the glottis is not established during a single vibration cycle.
- (2) The onset and shutoff of transglottal flow do not coincide with glottal opening and closure.
- (3) The flow through most of the cycle is strongly frequency dependent, and the strongest frequency dependence is associated with vortex formation.
- (4) A lack of frequency dependence occurs during the flow shutoff phase: shutoff occurs, on average, at the same point in the cycle, independent of cycle frequency, but
- (5) the precise shutoff time of a particular realization is influenced strongly by cycle-to-cycle variability in vortex shedding during this interval.

These findings not only suggest that glottal flow is unsteady, they have important implications for understanding how the voice is produced and on fundamental limitations in controlling perturbations and fluctuations. As vocal fold vibration frequency increases, the volume flow waveform skewness also increases, so that the rate of change of volume flow waveform is more intense with increasing frequency. In the past, this skewing has been explained by the inertance of the flow upstream and downstream of the vocal folds, but the current results suggest that the inertance is modified by the unsteady behavior of the glottal jet. In addition, the occurrence of premature volume flow shut-off means that the steepness of that shutoff gradient is larger than would be estimated by observation of glottal area and the Bernoulli equation. All these unsteady flow behaviors directly affect voice power and quality through their influence on the glottal volume flow waveform. Finally, the variability in jet flow shutoff time may account for naturally occurring perturbations and fluctuations in the voice ([Hirschberg, 1992](#); [Lous *et al.*, 1998](#)).

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from Grant No. 1R01DC054642-01 and from the Institute for Deafness and Other Communication Disorders, National Institutes of Health. In addition we would like to thank J. Petrowski for invaluable assistance with the motion control and with model fabrication issues. We would also like to thank N. Krane for assistance in machining the model.

Alipour, F., Scherer, R., and Patel, V. (1995). “An experimental study of pulsatile flow in canine larynges.” *J. Fluids Eng.* **117**, 577–581.

Barney, A., Shadle, C. H., and Davies, P. O. A. L. (1999). “Fluid flow in a dynamic mechanical model of the vocal folds and tract. I. Measurements

- and theory," *J. Acoust. Soc. Am.* **105**(1), 444–455.
- Barry, M. (2005). "Unsteady glottal aerodynamics in a scaled up mechanical model of the vocal folds," M.S. thesis, Department of Mechanical and Aerospace Engineering, Rutgers University.
- Berke, G., Moore, D., Monkewitz, P., Hanson, D., and Gerratt, B. (1989). "A preliminary study of particle velocity during phonation in an in vivo canine model," *J. Voice* **3**(4), 306–313.
- Coker, C. H., Krane, M. H., Reis, B. Y., and Kubli, R. A. (1996). "Search for unexplored effects in speech production," in *Proceeding of ICSLP '96*, Philadelphia, PA, 1996.
- Cranen, B., and Boves, I. (1985). "Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models of speech production," *J. Acoust. Soc. Am.* **77**(4), 1543–1551.
- Deverge, M., Pelorson, X., Vilain, C., Lagrée, P.-Y., Chentouf, F., Willems, J., and Hirschberg, A. (2003). "Influence of collision on the flow through in-vitro rigid models of the vocal folds," *J. Acoust. Soc. Am.* **114**(6), 3354–3362.
- Hirschberg, A. (1992). "Some fluid dynamic aspects of speech," *Bull. Commun. Parl.* **2**, 1–30.
- Hofmans, G., Groot, G., Rancci, M., Graziani, G., and Hirschberg, A. (2003). "Unsteady flow through *in vitro* models of the glottis," *J. Acoust. Soc. Am.* **113**(3), 1658–1675.
- Hsu, T.-Y. (2000). "Turbulent secondary flow in the mixed boundary corner formed by a free surface and a vertical solid wall," M.S. thesis, Department of Mechanical and Aerospace Engineering, Rutgers University.
- Hsu, T.-Y., Grega, L. M., Leighton, R. I., and Wei, T. (2000). "Turbulent kinetic energy transport in a corner formed by a solid wall and a free surface," *J. Fluid Mech.* **410**, 343–366.
- Ishizaka, K., and Flanagan, J. L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.* **51**, 1233–1268.
- Krane, M. H. (2005). "Low-frequency aeroacoustic production of unvoiced speech sounds," *J. Acoust. Soc. Am.* **118**(1), 410–427.
- Krane, M. H., and Wei, T. (2006). "Theoretical assessment of unsteady effects in phonatory aerodynamics," *J. Acoust. Soc. Am.* **120**(3), 1578–1588.
- Krueger, P. S. (2005). "An over-pressure correction to the slug model for vortex ring circulation," *J. Fluid Mech.* **545**, 427–443.
- Krueger, P. S., and Gharib, M. (2003). "The significance of vortex ring formation to the impulse and thrust of a starting jet," *Phys. Fluids* **15**(5), 1271–1281.
- Krueger, P. S., and Gharib, M. (2005). "Thrust augmentation and vortex ring evolution in a fully pulsed jet," *AIAA J.* **43**(4), 792–801.
- Liljencrants, J. (1996). "Experiments with analysis by synthesis of glottal airflow," in *Proceedings of ICSLP '96*, Philadelphia, PA.
- Lous, N. J. C., Hofmans, G. C. J., Velduis, 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**(6), 1135–1150.
- Mongeau, L., Franche, N., Coker, C. H., and Kubli, R. A. (1997). "Characteristics of a pulsating jet through a small, modulated orifice, with application to voice production," *J. Acoust. Soc. Am.* **102**(2), 1121–1133.
- Shah, P., Atsavaprane, P., Hsu, T. Y., Wei, T., and McHugh, J. (1999). "Turbulent transport in the core of a trailing delta wing vortex," *J. Fluid Mech.* **387**, 151–175.
- Story, B., and Titze, I. (1995). "Voice simulation with a body-cover model of the vocal folds," *J. Acoust. Soc. Am.* **97**, 1249–1260.
- Teager, H. M. (1980). "Some observations on oral airflow during phonation," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-28**(5), 599–601.
- van den Berg, J., Zatema, J., and Doornenbal, P. (1957). "On the air resistance and Bernoulli effect in the human larynx," *J. Acoust. Soc. Am.* **29**, 626–631.
- Vilain, C. E., Pelorson, X., Fraysse, C., Deverge, M., Hirschberg, A., and Willems, J. (2004). "Experimental validation of a quasi-steady theory for the flow through the glottis," *J. Sound Vib.* **276**, 475–490.
- Zhang, Z., Mongeau, L., and Frankel, S. (2002). "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes," *J. Acoust. Soc. Am.* **112**(4), 1652–1663.

Effects of perturbation magnitude and voice F_0 level on the pitch-shift reflex

Hanjun Liu and Charles R. Larson^{a)}

Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208, USA

(Received 27 February 2007; revised 26 September 2007; accepted 28 September 2007)

The purpose of the present study was to investigate the responsiveness of the pitch-shift reflex to small magnitude stimuli and voice fundamental frequency (F_0) level. English speakers received pitch-shifted voice feedback ($\pm 10, 20, 30, 40,$ and 50 cents, 200 ms duration) during vowel phonations at a high and a low F_0 level. Mean pitch-shift response magnitude increased as a function of pitch-shift stimulus magnitude, but when expressed as a percent of stimulus magnitude, declined from 100% with ± 10 cents to 37% with ± 50 cents stimuli. Response magnitudes were larger and latencies were shorter with a high F_0 level (16 cents; 130 ms) compared to a low F_0 level (13 cents; 152 ms). Data from the present study demonstrate that vocal response magnitudes are equal to small perturbation magnitudes, and they are larger and faster with a high F_0 voice. These results suggest that the audio-vocal system is optimally suited for compensating for small pitch rather than larger perturbations. Data also suggest the sensitivity of the audio-vocal system to voice perturbation may vary with F_0 level. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2800254]

PACS number(s): 43.70.Gr, 43.70.Fq, 43.70.Jt, 43.71.Bp [AL]

Pages: 3671–3677

I. INTRODUCTION

Control of voice fundamental frequency (F_0) is an important issue in speech communication, and auditory feedback plays a significant role in this process. Numerous studies have demonstrated that subjects produce compensatory responses in voice F_0 following unanticipated perturbations in voice pitch feedback. The fact that the pitch-shift reflex functions to help stabilize voice F_0 have been demonstrated during sustained vowels (Bauer and Larson, 2003; Burnett *et al.*, 1998; Chen *et al.*, 2007; Hain *et al.*, 2000; Kawahara, 1995; Kiran and Larson, 2001; Larson *et al.*, 2001; Sivasankar *et al.*, 2005), glissandos (Burnett and Larson, 2002), speech (Bauer, 2004; Chen *et al.*, 2007; Xu *et al.*, 2004) and nonsense syllables (Donath *et al.*, 2002; Natke *et al.*, 2003; Natke and Kalveram, 2001).

Responses to pitch-shifted voice feedback are generally of two types. The far more prevalent type changes voice F_0 in the opposite direction to the pitch-shift stimulus and has been termed “opposing” response. It is thought that these responses correct for errors between intended F_0 and the actual F_0 produced. The second type, termed “following” response, is relatively rare and changes F_0 in the same direction as the stimulus. These responses are inherently destabilizing since if unchecked they would cause voice F_0 to drift further and further from the intended F_0 . It is presently unknown what causes “following” responses. Both response types have latencies of approximately 100 – 150 ms, and magnitudes are generally a fraction of the stimulus. For example, a 100 cents stimulus (100 cents = 1 semitone) generally yields about a 30 cents response. Table I lists the

stimulus magnitudes involved in pitch-shift reflex studies in recent years. Although stimulus magnitudes ranged from 25 to 600 cents, 100 cents being the most widely used (Bauer and Larson, 2003; Burnett and Larson, 2002; Donath *et al.*, 2002; Kiran and Larson, 2001; Natke *et al.*, 2003; Natke and Kalveram, 2001; Sivasankar *et al.*, 2005), response magnitudes rarely exceeded 60 cents. Larson *et al.* (2001) reported a response magnitude of 26 to a 25 cent stimulus, but it is the only report of a full compensation to the pitch perturbations. Similarly, Bauer *et al.* (2006) reported a response of 0.99 to a 1 dB loudness perturbation during production of vowels with a low voice amplitude. These two studies suggest that a full response only occurs with small magnitude stimuli. Therefore, one hypothesis tested in the present study is that the reflexive mechanism may respond to perturbations with magnitudes equal to the stimuli only for small variations in voice feedback.

Recent studies of the pitch-shift reflex have examined the task-dependent role of auditory feedback in control of voice F_0 during speech production (Bauer, 2004; Chen *et al.*, 2007; Natke *et al.*, 2003; Xu *et al.*, 2004). For example, significantly larger responses were found in singing compared to a speaking condition (Natke *et al.*, 2003), and in speech compared to a vowel condition (Chen *et al.*, 2007). To date, however, there is no evidence whether there is a task-dependent modulation of pitch-shift response magnitude in sustained vowels. A recent loudness-shift study has demonstrated larger responses when subjects sustained vowels using a soft voice compared to a normal voice amplitude (Bauer *et al.*, 2006). In an analogous way, voice F_0 level may have a similar effect on responses to pitch-shifted voice feedback. Compared to a low F_0 level, vocalizing at a high F_0 requires greater degrees of muscle contraction (Hirano *et al.*, 1970; Titze, 1994), which may affect response magnitudes or

^{a)}Author to whom correspondence should be addressed. Electronic mail: clarson@northwestern.edu

TABLE I. Summary of the studies that compare the stimulus magnitude in vowel phonations.

Study	Stimulus magnitude (cents)	Average response magnitude (cents)
Burnett <i>et al.</i> , 1998	25, 50,100,150,200,250,300	3–100
Larson <i>et al.</i> , 2001	25,100, 200	26–48
Chen <i>et al.</i> , 2007	50,100, 200	22–32
Larson <i>et al.</i> , 2007	50	10–18
Natke <i>et al.</i> , 2003	100	47–66
Kiran and Larson, 2001	100	30–72
Burnett and Larson, 2002	100	18–24
Bauer and Larson, 2003	100	12–15
Hain <i>et al.</i> , 2000	100	20–40
Donath <i>et al.</i> , 2002	100	50
Sivasankar <i>et al.</i> , 2005	100	9–23
Natke and Kalveram, 2001	100, 600	15–65

latencies to perturbations in voice pitch feedback. That is, a reflexive input to motor neurons that are discharging at a high rate may lead to a greater level of muscle contraction (and hence greater change in F_0) than an equal input to neurons discharging at a lower rate. Therefore, another hypothesis tested was whether response magnitudes to pitch-shifted voice feedback would be larger when subjects vocalized at a high F_0 level compared to a low F_0 level.

Additionally, in previous studies subjects were instructed to phonate at a “comfortable” F_0 or loudness level during sustained vowels or speech production. Brown *et al.* (1976) noted, however, that comfortable level for both frequency and intensity can vary markedly across the experimental sessions. Through a trial period of five successive days, researchers found that subjects varied voice F_0 as much as 30 Hz and amplitude by 25.3 dB during sustained vowels and phrases at a comfortable effort level. Such variation in vocalization could lead to greater variability in response measures. Therefore, in the present study, the subjects were instructed to vocalize the vowel /u/ to match a piano note with a constant F_0 level (high or low) at approximately 70 dB. It was anticipated that this control would lead to reduced variability in response measures and hence increased accuracy for the assessment of the effects of stimulus magnitude and F_0 level on responses to pitch-shift voice feedback.

The third independent variable manipulated in this study was stimulus direction: upwards or downwards shifts in voice pitch feedback. Stimulus direction was not hypothesized to have an effect on the responses during vowel phonations. However, in this experimental paradigm, by randomly altering stimulus direction, it further reduces the chances that the subjects would be accustomed to the stimulus direction. Such an expectation could conceivably affect the responses.

II. METHODS

A. Subjects

Twenty-two subjects (five males and 17 females; ages 19–28), most of whom were students at Northwestern University, participated in the experiment. All subjects passed a hearing screening at 25 dB hearing level bilaterally at 250,

500, 1000, 2000, and 4000 Hz, and none reported a history of neurological or communication disorders. All signed informed consent approved by the Northwestern University Institutional Review Board and were paid for their participation.

B. Apparatus

Subjects wore Sennheiser headphones with attached microphone (model HMD 280) in a sound-treated room throughout the testing. They were asked to vocalize the vowel /u/ at approximately 70 dB sound pressure level (SPL), self-monitoring their voice loudness from a Dorrugh Loudness Monitor (model 40-A) placed 0.5 m in front of them. The vocal signal from the microphone was amplified with a Mackie mixer (model 1202) and shifted in pitch with an Eventide Eclipse Harmonizer, mixed with 40 dB SPL pink masking noise (Goldline Audio Noise Source, model PN2; spectral frequencies 1–5000 Hz) with a Mackie mixer (model 1202-VZL), and then amplified with a Crown D75 amplifier and HP 350 dB attenuators at 80 dB SPL. MIDI software (Max/MSP v.4.1 by Cycling 74) was used to control the harmonizer. A Brüel and Kjær sound level meter (model 2250) and in-ear microphones (model 4100) were used to calibrate the microphone and headphones to make sure there was a gain of 10 dB SPL between the subject’s voice amplitude and the feedback loudness. The voice output, feedback and TTL control pulses were digitized at 10 kHz, low-pass filtered at 5 kHz, and recorded using Chart software (AD Instruments). Data were analyzed using event-related averaging techniques in Igor Pro (Wavemetrics, Inc.). A keyboard (Yamaha, PSR-310) was used to present musical notes through the headphones to the subject prior to each set of trials.

C. Procedures

Before the experiment, subjects were first instructed and tested on their ability to produce a /u/ vowel and match the pitch of two different piano notes from the keyboard. One note was close to average values of conversational F_0 and the other was a much higher level. These notes were based on the gender of the subject as well as on testing on the subjects to make sure they were in a comfortable range for the sub-

jects: for male subjects, C4 (261.83 Hz) was used for the high F_0 and C3 (130.81 Hz) for the low F_0 ; for female subjects, E4 (329.63 Hz) was used for the high F_0 and A3 (220 Hz) for the low F_0 level. Several practice trials were given to make sure the subjects were able to match the notes within 100 cents. Then the voice was recorded and the F_0 was measured with a fast Fourier transform algorithm (Chart software) to verify that the subjects matched the note. If their voice F_0 was higher than the piano note, they were asked to reduce the voice F_0 and tested again, and vice versa. Two subjects were excluded from the data analysis because one was unable to match the note with high F_0 , and the other was unable to keep the voice F_0 constant. Therefore, data were analyzed from 20 subjects.

After the training, the subjects were instructed to repeatedly sustain the vowel /u/ for approximately 5 s duration at either the high or low F_0 level. Before each vocalization, the MIDI program automatically presented the piano note (0.5 s duration) through the headphones. The intensity of the piano notes was not calibrated, but was judged to be at a comfortable loudness level by the experimenters. Subjects were requested to vocalize the vowel /u/ to match the note during the experiment. Production of 12 consecutive vocalizations constituted an experimental block, and all trials within one block consisted of the same piano note. For each vocalization within a block, the voice pitch feedback was increased, decreased or held constant (no stimulus) five times in succession (randomized sequence), resulting in 20 increasing, 20 decreasing, and 20 control perturbations in each block of 12 trials. During each vocalization, the inter-stimulus interval varied between 900 and 1500 ms. The duration of each stimulus was 200 ms and the magnitude was held constant at $\pm 10, 20, 30, 40,$ or 50 cents within each block. Since the initial pitch perturbation occurred between 500 and 850 ms after vocal onset and the sequence and timing of subsequent stimuli were randomized, subjects could not predict the timing or direction of stimuli that would occur on any given trial.

D. Data analysis

The voice wave form was processed in Praat (www.Praat.org) using an autocorrelation method to produce pulses for each glottal cycle in the vocalizations. This process was done separately for each subject and each experimental condition. These signals were then transformed in Igor Pro to produce a F_0 contour for each vocalization. Event-related averaging techniques were used to measure the voice F_0 response separately to each stimulus direction or for the control trials. The averaging process was done by triggering the averaging program on one of three TTL pulses generated by the MIDI program at the time of the recording. The averaging window included a 200 ms pre and 700 ms poststimulus period. The program then calculated an average wave form for each stimulus direction, or control condition, and for each experimental condition and subject. By subtracting the average wave of the control trials from the average wave for either increasing or decreasing stimulus test waves, a “difference wave” was calculated for upward or downward stimulus

TABLE II. Total number of following (FOL), opposing (OPP), and nonresponse (NR) across F_0 level.

	High	Low	Total
FOL	7	17	24
OPP	191	177	368
NR	2	6	8
Total	200	200	400

trials. A point-by-point series of t tests was run between all control and all test trials, in which response latency was defined as the point where the “ p ” values of significant differences dipped below 0.02 following the onset of the stimulus with a delay of at least 60 ms and remained significant for at least 50 ms. Those cases where the p wave failed to reach a value of at least 0.02 for at least 50 ms, commencing at least 60 ms after the stimulus, were labeled as nonresponses. Response magnitude was measured as the greatest value of the difference wave following the latency and before the time where the p wave re-crossed the 0.02 value indicating the end of the response. Also, percent response magnitude was calculated by dividing the response magnitude by the stimulus magnitude. Voice jitter measures were made in Praat (local jitter) on all the data files for each subject, which is expressed as the average absolute difference between consecutive periods, divided by the average period. Response magnitude, latency, percent response magnitude and voice jitter were submitted to significance testing using repeated-measures analyses of variance (ANOVAs) (SPSS, v. 11.0). Responses to upward and downward stimuli were not tested separately but were averaged together in the statistical analyses. Nonresponses were replaced by the mean value calculated from the measured data from other subjects for that condition. A log transformation was done on response magnitude, latency and percent magnitude measures to achieve a normal distribution and homogeneity of variance. Assumptions of compound symmetry and circularity for a repeated measures ANOVA were met. Counts of opposing and “following” responses were made under each condition, and a Chi-square analysis was done to assess significance of different counts across conditions. An alpha level of $p < 0.05$ was considered to be statistically significant.

III. RESULTS

From 20 subjects across five stimulus magnitudes and two F_0 levels, there were 400 possible responses ($20 \times 5 \times 2 \times 2$). Ninety-two percent of the responses “opposed” the stimulus direction. Only 2% of the responses did not meet our criteria of validity and were declared to be nonresponses. Chi-Square tests revealed a greater number of “following” responses in the low F_0 condition than the high F_0 condition ($\chi^2 = 4.167, df = 1, p = 0.041$) (see Table II), and no statistically significant differences in response types were found across stimulus direction and stimulus magnitudes.

Figures 1 and 2 show examples of the average responses to pitch-shifted feedback across five stimulus magnitudes and two stimulus directions for high and low F_0 levels. All traces have been de-meant in order to illustrate magnitudes

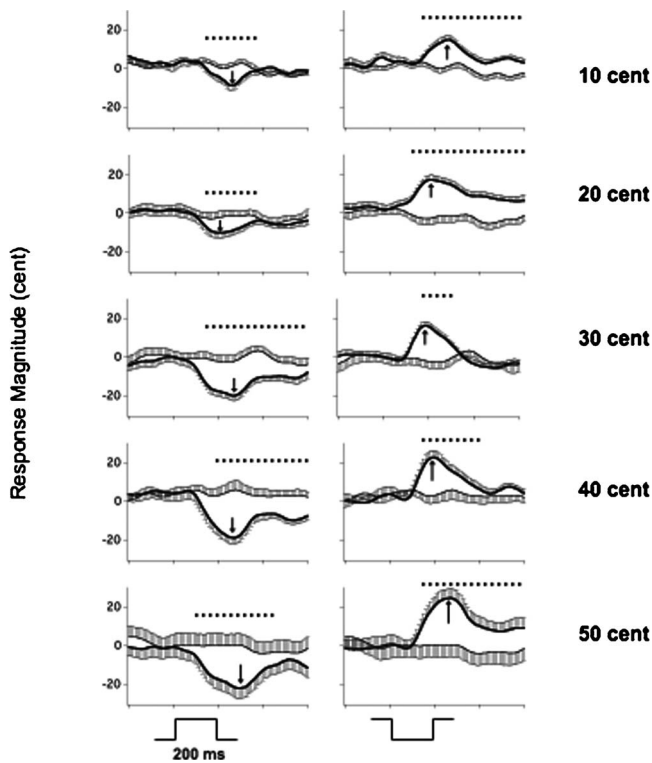


FIG. 1. Control (thin black line) and test average waves (thick black line) for stimulus magnitudes of 10, 20, 30, 40, and 50 cents at a high F_0 level for one subject. The vertical arrow indicates time where the response magnitude was measured. Error bars represent the standard error of the mean for a single direction. Dotted lines at top indicate time of significant difference between test and control waves. Vertical dashed lines indicate response onset. Square brackets at the bottom indicate the time and the direction of the stimulus. Voice F_0 values have been de-meant.

on the same scale. Responses to the upward stimuli are illustrated on the left and downward stimuli on the right, and the responses to $\pm 10, 20, 30, 40,$ and 50 cents are displayed from the top to the bottom. For each graph, the heavy line represents the average of the responses to the pitch perturbation and the light line the control responses. In all the plots of these two figures, the responses to the pitch perturbation are in the opposite direction to the stimulus. It is also clear, whether for the high or low F_0 condition, that F_0 contours of the control waves are relatively constant. It is difficult to see patterns in the responses in these figures; however, it appears that responses to downward stimuli may be somewhat larger than those to upward stimuli. It also appears that most responses began less than 200 ms after stimulus onset.

Figures 3–5 show boxplots of response magnitudes, response latencies, and percent response magnitude, respectively, as a function of stimulus magnitude for high and low F_0 levels. Values of response magnitude and latency are shown in Tables III and IV. Figure 3 illustrates the increase in response magnitude with stimulus magnitude, while an opposite effect is seen when the same data are plotted as percent response magnitude shown in Fig. 5. The percent response magnitudes were greatest with 10 cents stimuli ($\approx 100\%$) and decreased with greater stimulus magnitudes (Table III). A two-way repeated-measures ANOVA (stimulus magnitude and F_0 level) performed on the log transformed measures of response magnitude indicated significant main

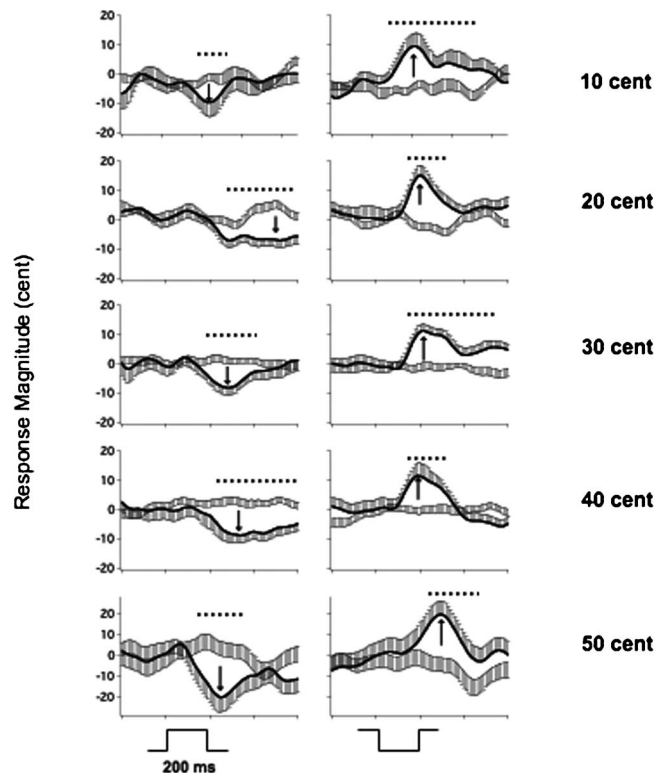


FIG. 2. Control (thin black line) and test average waves (thick black line) for stimulus magnitudes of 10, 20, 30, 40, and 50 cents at a low F_0 level for the same subject as in Fig. 1.

effects for stimulus magnitude ($F(4, 76)=43.989, p<0.001$) and F_0 level ($F(1, 19)=7.318, p<0.002$). There was no significant interaction between stimulus magnitude and F_0 level ($F(4, 76)=1.864, p>0.1$). Posthoc Bonferroni tests indicated that 40 (16.93 ± 7.72 cents) and 50 cents (18.48 ± 8.29 cents) conditions led to larger responses than 10 (9.06 ± 3.88 cents), 20 (12.27 ± 4.82 cents), and 30 cents stimuli (14.55 ± 6.65 cents) ($p<0.04$) conditions (Fig. 3). No significant differ-

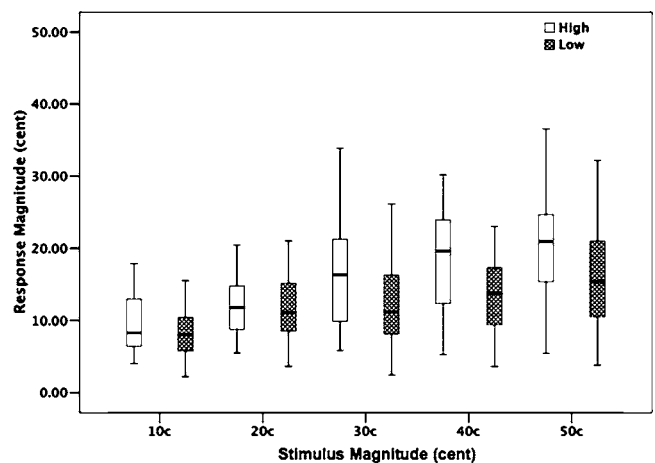


FIG. 3. Box plots illustrating the absolute response magnitude as a function of stimulus magnitude for the high F_0 and low F_0 levels. Shaded boxes are responses for low F_0 level and open boxes for high F_0 level. Box plot definitions: middle line is median, top and bottom of boxes are 75th and 25th percentiles, whiskers extend to limits of main body of data defined as high hinge $+1.5$ (high hinge $-$ low hinge), and low hinge -1.5 (high hinge $-$ low hinge).

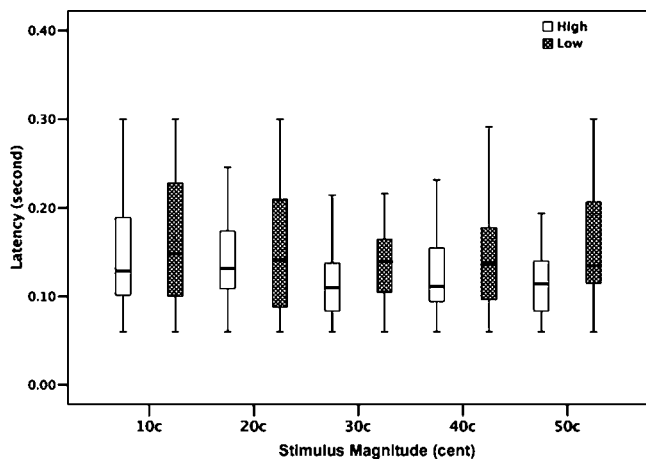


FIG. 4. Box plots illustrating the response latencies as a function of stimulus magnitude for the high F_0 and low F_0 levels.

ences were found between 20 and 30 cents stimuli ($p > 0.2$) and between 40 and 50 cents ($p > 0.9$).

For the latency, statistical analyses also revealed significant main effects for stimulus magnitude ($F(4,76) = 3.936, p < 0.01$) and F_0 level ($F(1,19) = 26.668, p < 0.001$). There was no significant interaction in the latency between stimulus magnitude and F_0 level ($F(4,76) = 1.247, p > 0.25$). Posthoc Bonferroni tests indicated that the 10 cents (157 ± 76 ms) condition produced a significantly longer latency than the 30 cents (128 ± 52 ms) condition ($p < 0.002$) (Fig. 4).

Statistical analysis of percent response magnitude also revealed significant main effects for stimulus magnitude ($F(4,76) = 78.508, p < 0.001$) and F_0 level ($F(1,19) = 7.471, p < 0.02$). Posthoc Bonferroni tests indicated significant differences between each stimulus magnitude except between 30 and 40 cents (Table V and Fig. 5). No significant interaction was found in the percent response magnitude between stimulus magnitude and F_0 level ($F(4,76) = 1.846, p > 0.1$).

A two-way repeated-measures ANOVA of voice jitter revealed significant main effects for F_0 level ($F(1,19) = 13.993, p < 0.002$) but not for stimulus magnitude

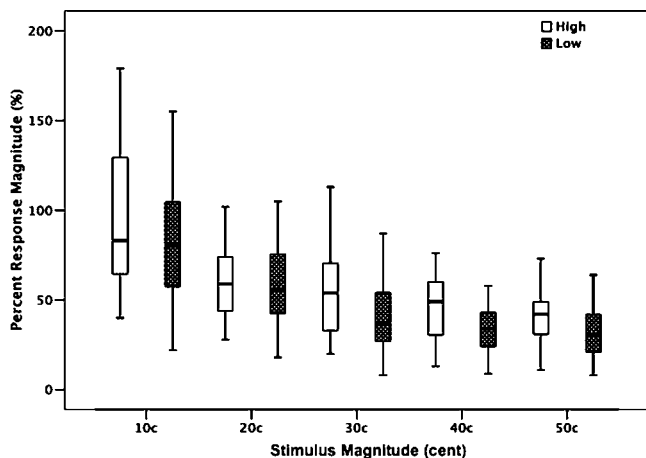


FIG. 5. Box plots illustrating the absolute percent response magnitude as a function of stimulus magnitude for the high F_0 and low F_0 levels.

TABLE III. Average response magnitudes in cents (SD) across stimulus magnitude and F_0 level.

	High F_0	Low F_0
10c	9.66 (4.18)	8.36 (3.44)
20c	12.46 (5.01)	12.08 (4.67)
30c	16.38 (6.94)	12.60 (5.79)
40c	19.24 (7.78)	14.49 (6.97)
50c	21.02 (8.38)	15.80 (7.40)

($F(4,76) = 1.232, p > 0.3$). Higher jitter values were observed for the low F_0 ($0.33\% \pm 0.12\%$) compared with the high F_0 level ($0.25\% \pm 0.09\%$). No significant interaction was found between stimulus magnitude and F_0 level ($F(4,76) = 0.151, p > 0.9$).

IV. DISCUSSION

The purpose of the present study was to investigate vocal responses to small pitch-shift stimuli as a function of vocal F_0 level and pitch-shift magnitude. It was found that response magnitudes were about equal to the 10 cents stimuli, but when expressed as a percent of stimulus magnitude, they decreased with larger stimulus magnitudes. We also found that larger responses were associated a high vocal F_0 level than with the low F_0 condition. In addition, response latencies were shorter for 30 cents compared with 10 cents stimulus magnitudes and with the higher vocal F_0 level compared to the low F_0 level. Thus, greater stimulus magnitudes or a high vocal F_0 level led generally to larger responses and shorter latencies.

The results also showed that, in the case of pitch perturbations as small as 10 cents, the audio-vocal system is capable of compensating for errors in the voice F_0 output. Due to the nonlinear relationship between cent and frequency, the pitch perturbations expressed in Hertz vary across the vocal F_0 level as follows:

$$\Delta F = \left(10^{\frac{\text{Acen}}{3986}} - 1 \right) \times F_0,$$

where ΔF is pitch shift in Hertz and F_0 is voice frequency in Hertz. Corresponding to the pitch changes from 10 to 50 cents, the frequency perturbations ranged from 0.76 to 3.85 Hz for males and from 1.28 to 6.47 Hz for females in the low F_0 condition, and from 1.52 to 7.70 Hz for males and 1.92 to 9.70 Hz for females in the high F_0 condition. Hence, one question that arises is whether the small pitch perturbations can be perceived by the auditory system. The minimum detectable change in frequency, or pitch discrimination level

TABLE IV. Average response latencies in ms (SD) across stimulus magnitude and F_0 level.

	High F_0	Low F_0
10c	145(68)	171(83)
20c	140(56)	145(67)
30c	115(41)	142(59)
40c	133(63)	154(80)
50c	121(52)	154(67)

TABLE V. Average percent response magnitude (SD) across stimulus magnitude and F_0 level.

	High F_0	Low F_0
10c	97(41)	84(34)
20c	62(25)	60(23)
30c	55(23)	42(19)
40c	48(19)	36(17)
50c	42(17)	32(15)

for pure tones, was reported as 1.0 Hz(8.6 cents) for both 200 Hz and for 400 Hz(4.3 cents) at a sensation level (SL) of 40 dB, and 1.2 Hz(5.2 cents) for 400 Hz at 80 dB SL (Wier *et al.*, 1977). Harris (1952) reported that at 40 dB SL the pitch discrimination threshold was 0.4 Hz(5.5 cents) for 125 and 0.75 Hz(5.2 cents) for 250 Hz, respectively. In general, pitch discrimination levels decreased with increases in sensation level regardless of frequency (Harris, 1952; Wier *et al.*, 1977). It is somewhat problematic to make a direct comparison between results of psychophysical testing on pure tones and the complex vocal signals in the present study, but assuming some correspondence between the two techniques, it appears that the 10 cents stimulus magnitudes in this study were somewhat greater than the minimal detectable change in frequency for a pure tone stimulus as reported by others. However, given that the 0.2 s pitch-shift stimuli were shorter in duration than those used by Wier *et al.* (1977) (0.5 s duration) or Harris (1952) (1.4 s duration), the pitch-shift stimuli used in this study may have been rather close to the threshold of detection.

In previous studies, it was shown that response magnitudes rarely exceeded 60 cents in the presence of 50, 100 or 200 cents stimuli (Bauer, 2004; Bauer and Larson, 2003; Burnett *et al.*, 1998; Chen *et al.*, 2007; Hain *et al.*, 2000; Kawahara, 1995; Kiran and Larson, 2001; Larson *et al.*, 2001; Sivasankar *et al.*, 2005). In such cases, percent response magnitudes were less than 50%, varying from approximately 10% to 40%. Larson *et al.* (2001) reported a mean response magnitude of 26 cents in response to 25 cents stimuli, where percent response magnitude was over 100%. In the present study, the mean response magnitudes were also about equal to the 10 cents stimuli (see Table III) but failed to equal the pitch perturbation for stimuli larger than 20 cents. Similar findings were also found in the loudness-shift study of sustained vowels (Bauer *et al.*, 2006), in which the response magnitude to 1 dB loudness perturbation was 0.99 dB at a soft amplitude level, but full compensation was not achieved for 3 or 6 dB amplitude stimuli.

Collectively, these observations indicate that the audio-vocal system can regulate voice F_0 or amplitude with a response magnitude equal to small perturbations, i.e., 10 cents or 1 dB. The failure of the system to correct for errors of larger magnitudes may be a self-protection mechanism to prevent environmental sounds from exerting a predominant influence over the voice output. If the audio-vocal system had a very high gain, such as the oculomotor in which a high gain is required to keep images stabilized on the retina (Glimcher, 1999), voice feedback or environmental sounds

could exert greater control over voice F_0 than voluntary mechanisms. Such mechanisms probably explain why even small stimuli may not always produce full compensation. For example, response magnitude of less than 14 cents was produced in response to 25 cents stimuli (Burnett *et al.*, 1998). Also, as Bauer *et al.* (2006) reported for loudness-shifted voice feedback, a gain of close to 1 was only reported when subjects attempted to maintain a relatively quiet voice amplitude, not when they vocalized at a louder amplitude. Hence, a full response to small stimuli may be related to mechanisms involved in the control of specific vocal conditions. To date, there is no evidence of a full response to pitch perturbations during speech production.

The finding that larger response magnitudes occurred with a high F_0 level compared to the low F_0 condition supports the hypothesis that task-dependent modulation of these responses occurs with vowels. This observation is analogous to the study in which it was shown that voice amplitude affected responses to loudness-shifted feedback (Bauer *et al.*, 2006). One explanation for this observation is that maintaining a relatively high F_0 that is clearly beyond the typical conversational level may require greater reliance on auditory feedback than a lower F_0 level. Another explanation is that response magnitude may be related to the phonation stability of sustained vowels (cycle-to-cycle variations in frequency). Previous studies of F_0 effects on the phonation stability have shown that pitch perturbation values (jitter) are greater in low F_0 conditions than in a high F_0 condition (Gelfer, 1995; Horii, 1979). In the present study it was also found that voice jitter was greater in the low F_0 condition compared to high F_0 , perhaps making the pitch-shift stimuli more salient in the high F_0 condition than in the low F_0 condition, which may have led to larger and more clearly defined responses.

It is also possible that the greater voice variability (jitter) in the low F_0 condition may be related to the greater number of “following” responses in this condition. Although the cause of “following” responses has not been identified, it was suggested (Hain *et al.*, 2000) that they are a result of the subject treating the feedback signal itself as the referent, as when a singer attempts to match a piano note. In such a case, the singer would adjust their F_0 towards the piano note (external referent). Alternatively, when a singer attempts to produce a remembered pitch, and if their voice pitch feedback does not agree with the memory (internal referent), they adjust their F_0 away from the feedback and towards the memory. In a more recent study (Larson *et al.*, 2007), greater numbers of “following” F_0 responses were observed with stimuli consisting of simultaneous changes in pitch and loudness feedback when the stimuli changed in opposite directions. It was suggested that the following responses were due to difficulty in correctly identifying the stimulus direction. Therefore, in the present study, the greater numbers of “following” responses in the low F_0 condition may have been caused by the greater difficulty in identifying the pitch-shift direction in the midst of a variable voice. It is noteworthy that some people are better than others at detecting the direction of pitch changes (Semal and Demany, 2006), which may explain why the numbers of “following” responses in this and previous studies do not follow a predictable pattern.

It is also possible that vocalizing in the high F_0 condition was a more difficult vocal task, which required greater attention to auditory feedback for regulation. Such task dependency has been reported in previous studies of the effects of perturbed feedback on voice control (Bauer *et al.*, 2006; Chen *et al.*, 2007; Natke *et al.*, 2003; Xu *et al.*, 2004).

Previous studies have demonstrated that response latency can be modulated as a function of stimulus magnitude (Larson *et al.*, 2001) and vocal task (Burnett and Larson, 2002; Chen *et al.*, 2007). The current findings showed longer latencies for the 10 cents stimuli, but with larger stimulus magnitudes, the latencies were shorter. Shorter latencies were also observed with the higher F_0 level. The shorter latencies observed with the greater stimulus magnitudes may reflect the fact that these stimuli were more easily perceived. Similar findings have been reported in the reaction time literature (Jaskowski *et al.*, 1994; Seitz and Rakerd, 1997). The finding that latencies decreased as vocal F_0 level increased is similar to the findings from pitch perception studies where it has been found that subjects are more sensitive to pitch deviations at a high frequency compared to a low frequency (Harris, 1952; Wier *et al.*, 1977). Thus the latency changes observed in the present study may reflect more general characteristics of the interaction between stimulus salience and the speed of response.

V. CONCLUSION

The results of the present study show that responses to perturbations in voice pitch feedback increase in magnitude as the pitch-shift stimuli increase from 10 to 50 cents. However, when response magnitude is considered as a percent of stimulus magnitude, percent response magnitudes decreased from over 90% for 10 cents stimuli to about 37% for 50 cents stimulus magnitudes. These findings suggest the audio-vocal system is optimally suited for stabilization of the voice around small perturbations. Larger response magnitudes were also recorded when subjects maintained a relatively high F_0 compared to a lower level, which may relate to either the lower voice F_0 variability found at high F_0 compared to low levels or a greater reliance on auditory feedback in the high F_0 condition. Altogether, the significant findings in this study may be due in part to the increased precision in the experimental procedures that required subjects to control their voice F_0 levels by matching a tone that was presented just before each vocalization.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. 1R01DC006243. The authors thank Chun Liang Chan for programming assistance and Kaavya Paruchuri for assistance in data recording. They authors also thank two anonymous reviewers who made important suggestions that improved the manuscript.

- Bauer, J. J. (2004). "Task dependent modulation of voice F_0 responses elicited by perturbations in pitch of auditory feedback during English speech and sustained vowels," Ph.D. Dissertation, Northwestern University, Evanston, IL.
- Bauer, J. J., and Larson, C. R. (2003). "Audio-vocal responses to repetitive

- pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique," J. Acoust. Soc. Am. **114**, 1048–1054.
- Bauer, J. J., Mittal, J., Larson, C. R., and Hain, T. C. (2006). "Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude," J. Acoust. Soc. Am. **119**, 2363–2371.
- Brown, W., Murry, T., and Hughes, D. (1976). "Comfortable effort level: An experimental variable," J. Acoust. Soc. Am. **60**, 696–699.
- Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (1998). "Voice F_0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.
- Burnett, T. A., and Larson, C. R. (2002). "Early pitch shift response is active in both steady and dynamic voice pitch control," J. Acoust. Soc. Am. **112**, 1058–1063.
- Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (2007). "Voice F_0 responses to pitch-shifted voice feedback during English speech," J. Acoust. Soc. Am. **121**, 1157–1163.
- Donath, T. M., Natke, U., and Kalveram, K. T. (2002). "Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables," J. Acoust. Soc. Am. **111**, 357–366.
- Gelfer, M. P. (1995). "Fundamental frequency, intensity, and vowel selection: Effects on measures of phonatory stability," J. Speech Hear. Res. **38**, 1189–1198.
- Glimcher, P. W. (1999). "Eye movements," in *Fundamental Neuroscience*, edited by M. J. Zigmond, F. E. Bloom, S. C. Landis, J. L. Roberts and L. R. Squire (Academic, San Diego), pp. 993–1010.
- Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (2000). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–141.
- Harris, J. D. (1952). "Pitch discrimination," J. Acoust. Soc. Am. **24**, 750–755.
- Hirano, M., Vennard, W., and Ohala, J. (1970). "Regulation of register, pitch and intensity of voice," *Folia Phoniatr.* **22**, 1–20.
- Horii, Y. (1979). "Fundamental frequency perturbation observed in sustained phonation," J. Speech Hear. Res. **22**, 5–19.
- Jaskowski, P., Rybarczyk, K., and Jaroszyk, F. (1994). "The relationship between latency of auditory evoked potentials, simple reaction time, and stimulus intensity," *Psychol. Res.* **56**, 59–65.
- Kawahara, H. (1995). "Hearing Voice: Transformed auditory feedback effects on voice pitch control," *Computational Auditory Scene Analysis' and International Joint Conference on Artificial Intelligence*, Montreal.
- Kiran, S., and Larson, C. R. (2001). "Effect of duration of pitch-shifted feedback on vocal responses in Parkinson's Disease patients and normal controls," J. Speech Lang. Hear. Res. **44**, 975–987.
- Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (2001). "Comparisons of voice F_0 responses to pitch-shift onset and offset conditions," J. Acoust. Soc. Am. **110**, 2845–2848.
- Larson, C. R., Sun, J., and Hain, T. C. (2007). "Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F_0 and amplitude control," J. Acoust. Soc. Am. **121**, 2862–2872.
- Natke, U., Donath, T. M., and Kalveram, K. T. (2003). "Control of voice fundamental frequency in speaking versus singing," J. Acoust. Soc. Am. **113**, 1587–1593.
- Natke, U., and Kalveram, K. T. (2001). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," J. Speech Lang. Hear. Res. **44**, 577–584.
- Seitz, P. F., and Rakerd, B. (1997). "Auditory stimulus intensity and reaction time in listeners with longstanding sensorineural hearing loss," *Ear Hear.* **18**, 502–512.
- Semal, C., and Demany, L. (2006). "Individual differences in the sensitivity to pitch direction," J. Acoust. Soc. Am. **120**, 3907–3915.
- Sivasankar, M., Bauer, J. J., Babu, T., and Larson, C. R. (2005). "Voice responses to changes in pitch of voice or tone auditory feedback," J. Acoust. Soc. Am. **117**, 850–857.
- Titze, I. R. (1994). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).
- Wier, C. C., Jesteadt, W., and Green, D. M. (1977). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.
- Xu, Y., Larson, C., Bauer, J., and Hain, T. (2004). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.

Acoustic cues to lexical segmentation: A study of resynthesized speech

Stephanie M. Spitzer^{a)} and Julie M. Liss

Motor Speech Disorders Laboratory, Department of Speech and Hearing Science, Arizona State University, Box 870102, Tempe, Arizona 85281-0102

Sven L. Mattys

Department of Experimental Psychology, University of Bristol, 12A Priory Road, Bristol BS8 1TU, United Kingdom

(Received 15 February 2007; revised 27 July 2007; accepted 28 September 2007)

It has been posited that the role of prosody in lexical segmentation is elevated when the speech signal is degraded or unreliable. Using predictions from Cutler and Norris' [J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121 (1988)] metrical segmentation strategy hypothesis as a framework, this investigation examined how individual suprasegmental and segmental cues to syllabic stress contribute differentially to the recognition of strong and weak syllables for the purpose of lexical segmentation. Syllabic contrastivity was reduced in resynthesized phrases by systematically (i) flattening the fundamental frequency (F0) contours, (ii) equalizing vowel durations, (iii) weakening strong vowels, (iv) combining the two suprasegmental cues, i.e., F0 and duration, and (v) combining the manipulation of all cues. Results indicated that, despite similar decrements in overall intelligibility, F0 flattening and the weakening of strong vowels had a greater impact on lexical segmentation than did equalizing vowel duration. Both combined-cue conditions resulted in greater decrements in intelligibility, but with no additional negative impact on lexical segmentation. The results support the notion of F0 variation and vowel quality as primary conduits for stress-based segmentation and suggest that the effectiveness of stress-based segmentation with degraded speech must be investigated relative to the suprasegmental and segmental impoverishments occasioned by each particular degradation. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2801545]

PACS number(s): 43.71.An, 43.71.Gv [MSS]

Pages: 3678–3687

I. INTRODUCTION

There exists no known set of acoustic cues that reliably and consistently mark word boundaries in connected speech (Lehiste, 1972; Nakatani and Shaffer, 1978). Nonetheless, the task of lexical segmentation is executed with relative ease and nearly flawlessly in the vast majority of communicative environments. It has been hypothesized that at least a portion of our success in lexical segmentation can be attributed to our knowledge about the prosodic structure of our language. Cutler and her colleagues proposed the *metrical segmentation strategy* (MSS), in which listeners capitalize on the rhythmic structures and statistical probabilities of a language to identify the location of word boundaries (Cutler and Norris, 1988). For example, in English, listeners should be successful in their lexical segmentation if they attend to strong syllables as potential word onsets because of the high occurrence of strong syllables word-initially in the English language (Cutler and Carter, 1987). This hypothesis is supported by data on naturally occurring and experimentally induced slips of the ear (Cutler and Butterfield, 1992; Smith *et al.*, 1989), in which listeners are more likely to mistakenly insert a lexical boundary before a strong syllable than a weak one (e.g., the misperception “I’ll play” for the target “ap-

ply”), and mistakenly delete a lexical boundary more often before a weak syllable than before a strong one (e.g., the misperception “caller” for the target “call a”).

Although the MSS hypothesis and subsequent implementations of the principle (e.g., McQueen *et al.*, 1994; Norris *et al.*, 1995) do not suggest that syllabic strength is the only source of information for lexical segmentation, there is converging evidence that listeners do indeed attend to prosodic variation in their segmentation of the connected speech stream when other sources of information are degraded or unreliable (Cutler and Butterfield, 1992; Smith *et al.*, 1989; Mattys, 2004; Mattys *et al.*, 2005; Norris *et al.*, 1997). Further, the work of Liss and colleagues (dysarthric speech) and Mattys and colleagues (manipulated speech) suggests that some forms of speech degradation have a more deleterious effect on the application of the MSS than others (Liss *et al.*, 1998, 2000, 2002; Mattys *et al.*, 2005).

This begs the question, then, whether certain aspects of prosodic information that contribute to syllabic strength contrasts have a greater effect on lexical segmentation than others. The need to tease apart the respective weights of segmentation cues has already been highlighted across the prosodic, segmental, and acoustic domains. Indeed, McQueen and Cutler (2001) wrote, “There is [...] now a relatively long list of cues which listeners appear to use for segmentation and which vary among languages, including phonotactics, allophonics, and other acoustic-phonetic cues,

^{a)}Author to whom correspondence should be addressed. Electronic mail: spitzer@asu.edu

silence, and metrical cues based on the language's rhythmic structure. An important issue which remains to be addressed is the relative ranking of these cues: do some cues carry more weight in segmentation than others?" (p. 481). Empirical efforts in that direction were made by [Mattys *et al.* \(2005\)](#), who found that listeners do not treat all segmentation cues equally. Pitting cues against each other, they showed that lexical and contextual sources of information override sublexical cues in case of conflict and that, among sublexical cues, segmental and acoustic regularities such as phonotactics and coarticulation override stress cues. Stress, however, had a substantial contribution to segmentation when the signal was presented in white noise or deprived of lexical and segmental cues (see also [Mattys, 2004](#)). These results and those by [Liss *et al.* \(1998, 2000\)](#) converge in showing the importance of stress-based segmentation when the signal is acoustically/phonetically degraded, be this degradation artificially induced or naturally occurring.

Because the perception of stress in English has been shown to be determined mainly by fundamental frequency and syllable duration contrasts ([Fry, 1955](#); [Lehiste, 1970](#); [Lieberman, 1960](#); [Mattys, 2000](#)), a question is whether these dimensions are equally important for segmentation. To date, this question has not been addressed directly, and we are only offered a few conflicting leads from studies that have manipulated acoustic parameters to assess their impact on the more general issues of speech intelligibility and perception. Some of these variables include fundamental frequency contour ([Laures and Weismer, 1999](#); [Mens and Povel, 1986](#)), segment duration ([Hammen *et al.*, 1994](#); [Hertrich and Ackermann, 1998](#); [Mens and Povel, 1986](#)), and vowel quality ([Bond, 1981](#); [Fear *et al.*, 1995](#); [Quené and Koster, 1998](#)).

When it comes to speech segmentation, the available information is sparse. In general, the presence of multiple stress parameters is shown to be more effective for segmentation than that of isolated ones ([Streeter, 1978](#)), at least for adult listeners ([Thiessen and Saffran, 2003, 2004](#)). However, the segmentation of Finnish and Dutch, both languages with word-initial stress predominance, can be facilitated by solely raising F0 in word-initial positions ([Vroomen *et al.*, 1998](#)). On the other hand, in English, lengthening the final, but not the initial syllable of synthesized trisyllabic words assists listeners in segmenting new words from an artificial language ([Saffran *et al.*, 1996](#)). While the absence of segmentation benefits from initial lengthening might indicate that duration is not a strong cue for word onsets, the effect of final lengthening could possibly reflect a general phenomenon across languages rather than a language-specific feature ([Hoequist, 1983a, b](#)). Finally, it should be noted that [Cutler \(1986\)](#) emphasized the role of segmental in addition to suprasegmental information for stress-based segmentation. In Cutler's account, stress contrastivity is inferable not only by suprasegmental cues such as F0 and duration but also—and perhaps mostly—by the fact that strong syllables typically bear a full vowel whereas weak syllables are often reduced. Accordingly, stress-based segmentation would be more a matter of vowel quality than one of suprasegmental informa-

tion, with lexical access initiated on syllables containing a subset of vowels (e.g., /a, u/) and refrained on syllables containing another (e.g., /l/).

The aim of the present investigation was to examine a set of cues that may contribute differentially to the perception of strong and weak syllables for the purpose of lexical segmentation in impoverished speech. In this study, we reduced syllabic contrastivity by systematically altering suprasegmental cues (i.e., flattening the fundamental frequency and/or equalizing vowel durations) and the segmental cue of vowel identity (i.e., changing all vowels into near-schwas) using speech resynthesis. Using the phrase-transcription task of [Liss *et al.* \(1998, 2000\)](#), our primary question was whether the manipulations produced independent effects on lexical boundary error patterns. Predictions from the MSS hypothesis were used as a framework for interpretation of the results. Those manipulations that elicited error patterns consistent with MSS would be regarded as less detrimental to the application of the stress-based segmentation strategy, whereas those that elicited mis-segmentation patterns closer to chance would be regarded as more detrimental to identifying syllabic strength and therefore to the application of stress-based segmentation.

II. METHODS

A. Phrases

Development of the phrases used in this investigation has been described in detail elsewhere ([Liss *et al.*, 1998](#)). Briefly, the stimuli consisted of 40 six-syllable phrases, which alternated in syllabic strength. Half of the phrases contained a strong-weak (SWSWSW) pattern and the other half a weak-strong (WSWSWS) pattern. Strong syllables received prosodic emphasis and contained full vowels (which were longer in duration and higher in F0 variation, relative to weak syllables). Weak syllables did not receive prosodic emphasis and contained a reduced vowel or a schwa. The phrases consisted of a series of one- and two-syllable words, with low semantic transitional probability. A list of the phrases is provided in Appendix A.

B. Initial speech samples

A 31-year-old male with a standard American dialect produced the speech samples for this study. The speech samples were collected in a sound-treated room, recorded directly onto a PC using an Audio Technica ATM31 microphone and Kay Elemetrics CSL (ASL, KayPentax, 1999, 2004) for the highest fidelity recording possible (22050 kHz; 16 bit). The speaker read the phrases in a normal, conversational manner approximately three times each. A single rendition of each phrase was chosen which contained no pauses or misarticulations.

C. Resynthesis

Phrase resynthesis was conducted with Analysis Synthesis Lab (ASL, KayPentax, 1999). Six resynthesis conditions were created for the present investigation. In addition to a baseline resynthesis condition in which no additional ma-

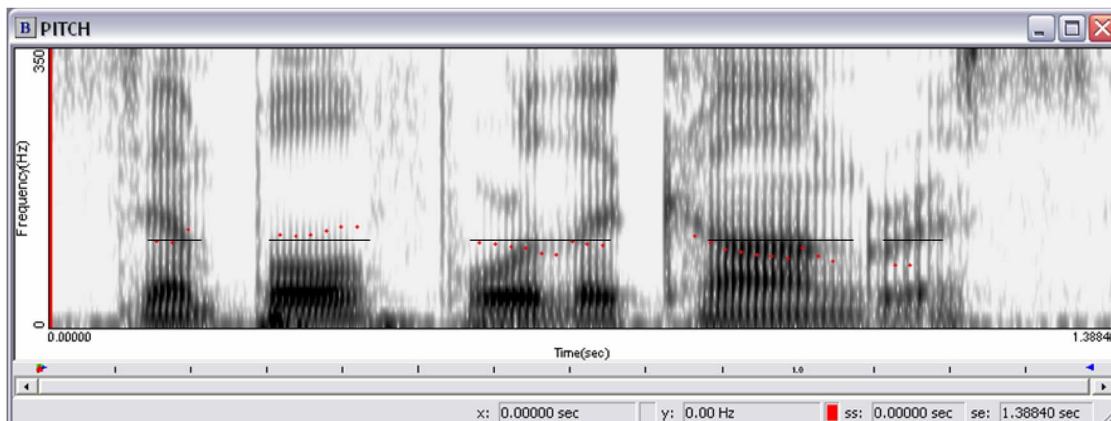


FIG. 1. (Color online) Spectrogram of the phrase “simple for the congress.” The dotted line is the original F0 track, as produced by the speaker. The solid line is the manipulated F0 track after F0 flattening; the value was set to the mean F0 across the phrase.

nipulations were made (Control), five conditions manipulated either suprasegmental or segmental cues, or a combination thereof. These included: (i) A flattened fundamental frequency condition (F0); (ii) a condition in which the vowels were lengthened to the duration of the longest strong vowel in the phrase (Duration); (iii) a combination of the F0 and Duration conditions (Suprasegmental Combined); (iv) a weakened vowel condition in which the first and second formants of all vowels in the phrase were replaced with schwa vowel values (Vowel); and (v) a condition including all single cue manipulations (Total Combined). Phrases in all conditions were subjected to a custom rms intensity equalization procedure in MATLAB prior to their use in the listening experiment (Louizou, 1999, 2006). Each condition is detailed in the following.

1. Resynthesized condition (Control)

The ASL (1999, 2004) software performs speech resynthesis using LPC-extracted parameters. These analysis/synthesis routines were implemented using pitch-synchronous, residual-driven LPC for the highest quality synthesis possible. Because the quality of the resynthesis is increased greatly by downsampling the speech files prior to manipulation, all phrases were downsampled to 10 kHz. We used resynthesized phrases rather than the original phrases as a control for the quality of the resynthesis procedure. Acoustically, these phrases were highly similar to the originals; untrained listeners were unable to distinguish them from the original phrases. All manipulations for the single and combined cue conditions were conducted on this set of resynthesized phrases.

2. Fundamental frequency flattened condition (F0)

The mean F0 (Hz) of each phrase was measured in CSL by standard procedures, in which the F0 was calculated by the *pitch contour* function in CSL. The F0 contour was then flattened to this mean value using the numerical editor function in ASL. It was determined that this method yielded a more natural-sounding result than setting the F0 contours to a single value for all phrases. A similar procedure was employed in the Laures and Weismer (1999) study to flatten the

F0 contours of their phrases. The perceptual effect of this manipulation was to lend a monotone, robotic quality to an otherwise intelligible phrase. This manipulation is illustrated in Fig. 1, where both the original and the flattened F0 contours are overlaid on the spectrogram of the phrase “simple for the congress.”

3. Equalized vowel duration condition (Duration)

Although there are a variety of procedures for achieving equalized vowel durations across syllables, we elected to lengthen all vowels within a phrase to match the longest strong vowel duration in that phrase. Vowel durations were lengthened using an ASL algorithm that duplicates the signal within a 10-ms frame of the vowel. A single frame was chosen at each vowel temporal midpoint. This frame was then duplicated to achieve the desired duration for that vowel. The longest vowel duration in each phrase was chosen as the standard, and the remaining vowels were lengthened to equal this target duration. The phrases were necessarily lengthened as a result of this manipulation (see Appendix B). This procedure was preferred over one that would have preserved total phrase duration because segment shortening would have likely resulted in a greater loss of segmental identity than segment lengthening. Perceptually, the phrases had an equal-and-even rhythmic quality. The effect of this manipulation is illustrated in Fig. 2.

4. Suprasegmental combined condition (Suprasegmental Combined)

This condition was created to assess the combined effect of the suprasegmental manipulations (F0 and Duration) on lexical segmentation. The phrases were manipulated in the manners described earlier: The F0 contour was set to the mean value for each phrase and the vowels were set to the longest vowel duration in each phrase.

5. Weakened vowel condition (Vowel)

The purpose of this condition was to reduce the possible contribution of vowel quality to perceived syllabic strength (Cutler, 1986; Fear *et al.*, 1995). To do so, we replaced the first two formant values of all vowels with those of a schwa.

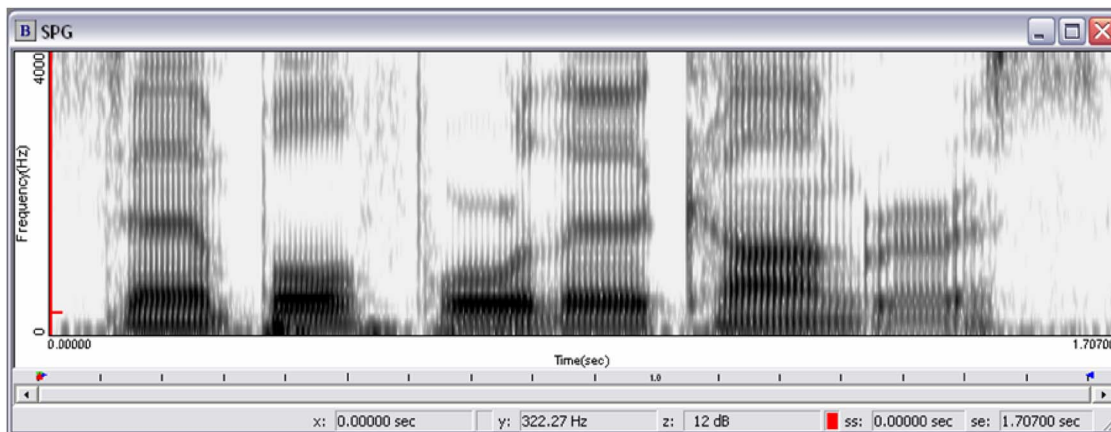


FIG. 2. (Color online) In the Duration condition, the phrase “simple for the congress” was manipulated to equalize duration of the six vowels. Five vowels were lengthened to equal the longest vowel in the phrase (the last strong vowel in this example).

Only the first two formants were manipulated since that portion of the spectral information has been reported to contain the majority of the vowel-differentiating information (Hillenbrand *et al.*, 1995; Kewley-Port and Zheng, 1999; Neary, 1989). Furthermore, the resynthesis algorithms are more reliable for the lower frequencies due to the interaction of filter order and sampling frequency with the LPC analysis.

Using ASL, the formant values of each vowel were changed to the following: $F1=623$ Hz and $F2=-1200$ Hz (Hillenbrand *et al.*, 1995). These values were included in the range of the study speaker’s own first and second formant values for schwa vowel productions. Vowel duration was not affected by this manipulation. When the $F2$ of the original strong vowel was particularly high (as it was for 10 /i/ vowels, out of 240 total vowels), the slope of the vowel transitions was modified using an automatic smoothing algorithm to avoid abrupt changes in formant values from the original to the resynthesized samples. Thus, the entire extent of the first two formants for each vowel was manipulated to approximate a centralized vowel. Because the higher formant frequencies remained unchanged, not all information related to vowel identity was eliminated. Perhaps for this reason, vowel identity was not completely indiscernible, and the per-

ceptual experience was one of accented or unusual speech. The spectrogram for a phrase in this manipulation condition is shown in Fig. 3.

6. Total combined condition (Total Combined)

All single cue manipulations ($F0$, Duration, and Vowel) were conducted on the resynthesized phrases to assess the reduction of suprasegmental and segmental cues on lexical segmentation. A subset of the phrases in the above-described manipulation conditions can be heard on our website (see Spitzer, 2007).

D. Preliminary verification of resynthesized phrases

The success of the phrase-transcription paradigm requires the elicitation of an interpretable number of lexical boundary errors (LBEs), as has been demonstrated previously by studies of faint speech, speech in noise, and impaired speech (Cutler and Butterfield, 1990, 1992; Liss *et al.*, 1998, 2000, 2002; Smith *et al.*, 1989). To achieve this goal across all our conditions, it was necessary to reduce the overall intelligibility of the phrases. This was accomplished by merging each phrase with broadband noise (1–5000 Hz,

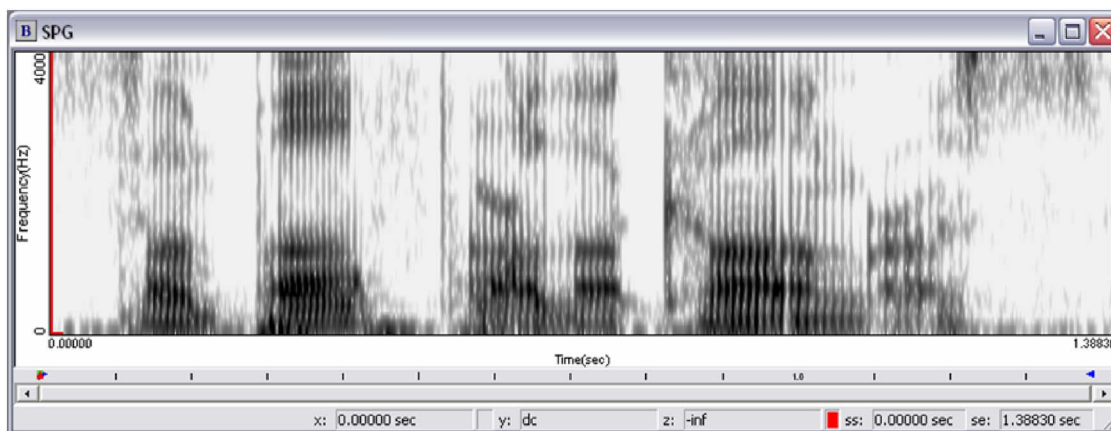


FIG. 3. (Color online) In the Vowel condition, the phrase “simple for the congress” was manipulated to shift the $F1$ and $F2$ values of all vowels to 623 and 1200 Hz, respectively, for the total duration of each of the 6 vowels. The vowel durations were untouched in this manipulation.

5-ms transition) for a signal-to-noise ratio (SNR) of +18 dB. This SNR was selected because, in preliminary testing, it resulted in intelligibility decrements of the order necessary to elicit an interpretable number of LBEs for the cue-manipulated conditions.

E. Listeners, materials, and listening task

Participants were 90 undergraduate and graduate students enrolled in classes at Arizona State University. All were native English speakers between the ages of 18 and 46 (mean age=23.88, s.d.=6.34) who self-reported normal hearing and no disease or conditions known to affect speech or language processing. The group contained 77 females and 13 males. Listeners were compensated for their participation.

The materials consisted of 40 phrases \times 6 conditions (240 phrases in total). Excluding the phrases in the Suprasegmental Combined condition, the phrases in the 5 other conditions were organized in 5 lists. Each list included 8 different, randomly selected phrases from each of the 5 conditions. Within each list, conditions were arranged in blocks (i.e., the 8 phrases per condition were presented in sequence). The order of conditions was counterbalanced across the 5 lists. Phrases were preceded by a nonresynthesized female voice saying the item phrase number, and followed by a 15-s ISI for phrase transcription. Because new software became available during the course of this study, transcription data for the Suprasegmental Combined condition were collected using ALVIN (Hillenbrand, 2005), a stimulus-presentation software package designed for speech perception experiments. Phrases were randomly presented in an open transcription task with no feedback or replay capabilities.

The first 75 listeners were randomly assigned to 1 of the 5 lists. The next 15 listeners transcribed phrases in the Suprasegmental Combined condition. Thus, the full data set consisted of 15 complete transcriptions of the 40 phrases from each of the 6 conditions. All listeners performed the task wearing Sennheiser HD 25 SP headphones in a quiet room free of visual and auditory distractions. At the beginning of the experiment, the signal volume was set to a comfortable listening level by each listener, and remained at that level for the duration of the task. After instructions to write down exactly what they heard, listeners were presented with 3 practice phrases, not in noise. These practice phrases were not resynthesized and did not contain any of the same words found in the 40 experimental phrases. Listeners who were not able to transcribe the practice phrases with 92% accuracy were excluded from the study (i.e., listeners were allowed a single word error for all three phrases). No listeners in this study were removed for violating this or any other criterion.

Following the practice phrases, listeners heard each experimental phrase only once, and immediately transcribed what they had heard. The listeners were told they would hear a male speaker produce a series of phrases played in a background of noise, and that all the words in the phrases were real English words. They were encouraged to guess on a word if they were not entirely sure about what they heard. They were not given any feedback about their performance during the course of the experiment.

TABLE I. Examples of words-correct intelligibility scores (%WC), lexical boundary errors (LBE), and word substitutions (WSubs) from listeners in all six conditions in response to the same target phrase. Word substitutions are shown in *italic*, insertion LBEs in **bold**, and deletion LBEs underlined.

Target phrase: amend the slower page			
Listener transcripts	%WC	LBE	WSubs
Control: amend the <i>solar</i> page	75	0	1
F0: come in the slower <i>tape</i>	50	1 IS	1
Duration: <i>among</i> the <u>novocane</u>	25	1 DS	1
Vowel: <i>among</i> the slow <u>inflate</u>	25	1 IW; 1 DS	1
SupraSeg: <u>Amanda's</u> <i>lovers' pace</i>	0	1 DW	2
TotalComb: <i>robo</i>	0	0	1

III. TRANSCRIPT ANALYSIS

The transcripts were coded independently by two trained judges for the number of words correctly transcribed as well as the presence, type (insertion or deletion), and location (before strong or before weak syllables) of lexical boundary errors (LBEs). In addition, transcription errors that did not violate lexical boundaries were tallied (Word Substitutions) as well as instances in which no attempt was made at transcription of a phrase (No Response). It should be noted that word substitutions and LBEs are mutually exclusive. Word substitution errors occur when the transcription is correctly parsed (thus no lexical boundary error) but the response is not the target word. For example, a listener's response of "advice" for the target "convince" is a word substitution, whereas the response "the fence" is an insertion of boundary before a strong syllable. Inter-rater reliability was 0.9875 prior to consensus, and 1.000 after consensus since any items not in agreement were excluded from the analysis. Criteria for scoring words correct were identical to those in Liss *et al.* (1998), and included tolerance of word-final morphemic alterations which did not affect the number of syllables (i.e., "boats" for "boat," but not "judges" for "judge"), as well as substitutions of function words "a" and "the." Examples of listener transcripts and coding of LBEs are shown in Table I; these errors were made in response to the same phrase across the 6 different conditions.

The main dependent variables, tallied for each listener and phrase across conditions, were: (1) The number of words correctly transcribed out of the total number of words possible (intelligibility score) and (2) the number, type (insertion versus deletion), and location (before strong versus before weak syllables) of LBEs. Accordingly, LBEs fell in four categories: Insertion of a word boundary before a strong syllable (IS), insertion of a word boundary before a weak syllable (IW), deletion of a word boundary before a strong syllable (DS), and deletion of a word boundary before a weak syllable (DW). Finally, we calculated (3) a MSS ratio, defined as the number of MSS-consistent LBEs, namely, insertions of word boundaries before strong syllables and deletions of word boundaries before weak syllables, divided by the total number of LBEs (IS+DW/total LBEs) for each set of 15 transcriptions of the 40 phrases per condition. A MSS ratio greater than 0.50 was taken as evidence of stress-based segmentation.

TABLE II. Summary measures.

Condition	Control	F0	Duration	Vowel	SupraSeg.	TotalComb.
Intelligibility (%)	92.93	59.02	53.46	51.67	28.21	15.22
s.d.	2.24	5.22	5.28	6.07	5.14	2.62
Exact target	445	144	99	78	32	1
No response	0	22	15	19	23	150
Word subs.	89	341	396	461	242	354
Total LBE	43	239	325	242	540	432

IV. RESULTS

A. Intelligibility

Because phrases from the Control, F0, Duration, Vowel, and Total Combined conditions were distributed across 5 test lists, it was first necessary to verify that differences in intelligibility scores were not simply the effect of listener group (group effect) or nonrandom distribution of phrases among the sets (phrase effect). The mean intelligibility for the 8 phrases within each condition (across all 5 sets) was computed for each listener and entered in a two-way analysis of variance (group \times phrase). This analysis revealed no significant effect of either group $F(4,20) < 1$ or phrase $F(4,20) < 1$. These results suggest that there was no significant difference in the intelligibility of the phrases within each subset of 8 phrases, thus allowing a comparison of the intelligibility scores across the 5 experimental conditions without further consideration for subset assignment. In the remainder of the analysis, the data sets consisted of the 15 transcriptions for each of the 6 conditions (15 transcriptions \times 40 phrases = 600 phrase transcriptions per condition).

Table II shows the mean intelligibility values and standard deviations for each of the six test conditions. A one-way analysis of variance, $F(5,84)=493.65$, $p < 0.001$, revealed significant intelligibility differences in all pairwise comparisons ($p \leq 0.005$), except for Duration vs. Vowel, where $p = 0.33$ (Fisher's least significant difference, LSD, correction was used in all pairwise comparisons). In descriptive terms, the Control condition, in which noise was added to otherwise unmanipulated resynthesized phrases, suffered only a small reduction in intelligibility (93% correct word identification), due most likely to the unusual meaning of the phrases rather than to any signal degradation. The three single-cue conditions, Duration, F0, and Vowel, caused a substantial intelligibility decrement compared to the Control condition. As expected, the two combined-cue conditions, Suprasegmental Combined and Total Combined, reduced intelligibility even further, with Total Combined less intelligible than Suprasegmental Combined (15% and 28%, respectively). Thus, transcription accuracy was markedly impaired by both suprasegmental and segmental alterations and further worsened when alterations were combined.

B. Analysis of lexical boundary errors (LBEs)

The overall performance pattern, together with a breakdown by LBE type and location, are reported in Tables II and III. Presenting the phrases in noise was expected to elicit an interpretable number of LBEs and this was achieved, with a

total of 1821 LBEs across 3600 phrase transcriptions (40 phrases \times 15 listeners \times 6 conditions). The Control condition was omitted from all subsequent analyses because, on account of its relatively preserved intelligibility, it elicited too few perceptual errors and even fewer LBEs for mis-segmentation patterns to be interpretable—less than 3 LBEs, on average, were committed per set of 40 phrases in the Control condition. An analysis of variance performed on the 5 remaining conditions showed marked disparities in the average number of LBEs per condition, $F(4,70)=23.57$, $p < 0.001$. Except for the F0 versus Vowel comparison, which did not reach significance, $p = 0.94$, all LSD-corrected pairwise comparisons showed reliable differences ($p < 0.01$ in all but Vowel versus Duration, $F[1,28]=5.53$, $p < 0.05$). In sum, the occurrence of LBEs varied from one condition to another but, in general, the pattern followed that of intelligibility, with lower intelligibility accompanied by a higher incidence of LBEs. The link between the two variables is expected, as it reflects the greater opportunity for mis-segmentations caused by a perceptually less intelligible input. Note, too, that there was a preponderance of boundary insertions over deletions for all conditions (all p 's < 0.005), which indicates that listeners tended to extract short words over long ones, perhaps because of the generally higher frequency of the former.

In stark contrast to the LBE frequency data, the patterns of LBEs, that is, the distribution of lexical boundary insertion and deletion errors, either before strong or weak syllables, were not directly predictable from intelligibility. A comparison of MSS ratios across conditions allowed us to evaluate the extent to which the test manipulations affected adherence to stress-based segmentation. As mentioned earlier, the MSS ratio breaks down LBEs into those that are predicted by MSS, i.e., inserting a word boundary before a strong syllable and deleting a word boundary before a weak

TABLE III. Distribution of lexical boundary errors. Percentage and type of LBEs are shown across the cue-manipulation conditions. The MSS ratio = (IS+DW)/total LBE.

LBE type/location	F0	Duration	Vowel	SupraSeg.	TotalComb.
IS (%)	50.21	53.54	56.61	49.26	42.82
IW (%)	11.72	10.46	16.12	23.70	16.44
DS (%)	14.23	8.62	16.12	9.44	12.50
DW (%)	23.85	27.38	11.16	17.59	28.24
MSS ratio	0.74	0.81	0.68	0.69	0.71

syllable, and those that are not, i.e., deleting a word boundary before a strong syllable and inserting a word boundary before a weak syllable. Since the MSS ratio is the proportion of LBEs predicted by MSS divided by the total number of LBEs, high values reflect stronger adherence to MSS whereas lower values reflect either weaker adherence to MSS or failure to do so because of a lack of perceived syllabic strength contrast. A first observation is that the MSS ratio was significantly greater than 0.50 in all conditions ($t[14]$, with all p 's < 0.001), suggesting that listeners relied on stress-based segmentation to cope with signal distortion across the board.¹ However, an analysis of variance showed that the strength of adherence to this strategy varied across conditions, $F(4, 70) = 4.36$, $p < 0.005$. LSD-corrected pairwise comparisons highlighted a pattern of results quite different from that found for intelligibility and total LBE data. Compared to the Duration condition, all conditions showed a significant attenuation in adherence to MSS: F0, $F(1, 28) = 5.04$, $p = 0.03$, Suprasegmental Combined, $F(1, 28) = 17.06$, $p < 0.001$; Vowel, $F(1, 28) = 11.76$, $p < 0.005$; Total Combined, $F(1, 28) = 11.19$, $p < 0.005$. None of the other comparisons reached $p < 0.20$. Implications for stress-based segmentation are presented in the following.

V. DISCUSSION

The findings of this investigation provide additional evidence that listeners rely on syllabic stress contrasts to assist in word-boundary identification in degraded speech (Liss *et al.*, 1998, 2000, 2002; Mattys, 2004; Mattys *et al.*, 2005; Smith *et al.*, 1989). All experimental conditions elicited large numbers of LBEs that generally conformed to predictions generated by the MSS. For all conditions, insertion errors before strong syllables outnumbered by two to five times those before weak syllables. With one exception, lexical boundary deletion errors occurred more often before weak than before strong syllables. Thus, even when attempting to reduce or eliminate the major cues to syllabic stress, 68%–81% of LBEs occurred in the predicted locations, which highlights the robustness of the strategy and its likely reliance on both the present cues and cues not tested in this study.

Despite the overall adherence to the MSS predictions, there were significant differences in LBE patterns across the various cue conditions. Among the single-cue manipulations, those involving the reduction of intonational information (F0) and vowel identity (Vowel) were most detrimental to metrical segmentation. Equalizing vowel duration (Duration) was significantly less so. Thus, listeners' reliance on syllable strength for the purpose of lexical segmentation appears to be contingent on the availability of both suprasegmental (F0) and segmental (Vowel) sources of information. It is worth noting that we found no evidence for compensatory mechanisms between the cues. For example, there was no indication that the effect of a lack of intonational cues was compensated for by an adequate durational contrast or full/reduced vowel contrast. The same goes for a lack of full/reduced vowel contrast. MSS's resilience to durational alteration suggests that vowel duration contrasts, though

critical for disentangling lexically ambiguous junctures in nondegraded speech (Shatzman and McQueen, 2006), are not a necessary condition for the application of stress-based segmentation, at least in the form of degradation used in this investigation. Finally, the absence of any cumulative effects on the MSS ratio when more than one cue is removed suggests a floor effect when either F0 or vowel-identity cues are reduced.

Although the statistical analyses of the MSS ratios reveal a similar magnitude of effect for the F0 and Vowel manipulations, additional evaluation of the performance patterns in these two conditions suggests the Vowel condition was the more perceptually challenging of the two. Both conditions elicited a nearly identical number of lexical boundary errors (239 and 242 for F0 and Vowel, respectively); however, the Vowel condition elicited both significantly lower intelligibility scores and a lower MSS ratio. The latter finding is due, in part, to the fact the only *nonpredicted* LBE pattern in this study was found in the deletion errors in the Vowel condition. Here deletions before strong syllables occurred 1.4 times more often than those before weak syllables. This, along with the relatively lower adherence for the insertion errors compared with the F0 condition, suggests a more substantial challenge in syllable stress assignment in the Vowel than the F0 condition. The additional reduction in intelligibility over the F0 condition can be traced, in part, to the high number of word substitution errors in the Vowel condition, in which lexical boundaries were preserved, but the transcribed words were incorrect. It can be surmised that the loss of vowel identity should have a much greater computational impact than F0 alteration on word identification given the widening of the lexical neighborhood and resulting lexical confusions. In sum, although both vowel identity (segmental) and F0 contrasts (suprasegmental) seem critical for an effective use of stress-based segmentation, assessment of overall performance patterns suggests that the MSS might be slightly more segmentally than suprasegmentally guided (Cutler, 1986), with vowel identity (full versus reduced) outweighing the main correlates of stress perception (F0 and duration).

As for the greater disruption caused by the F0 than the duration manipulation, the literature on the acoustic correlates of perceived stress in English (e.g., Fry, 1955; Lehiste, 1970; Lieberman, 1960) has indeed documented the leading role of F0 in discriminating stressed from unstressed syllables. However, duration, though second to F0, is often described as a substantial correlate of perceived stress as well (e.g., Mattys, 2000), which seems at odds with the limited impact its disruption had on the use of MSS. An explanation might be sought in the literature on auditory perceptual grouping. Tones alternating in F0 are shown to be perceived as strong-initial groups (Handel, 1989; Hay and Diehl, 1999). Similarly, Vroomen *et al.* (1998) reported enhanced word segmentation when the initial syllables of the words of an artificial language were pitch accented. In contrast, duration alternations tend to promote iambic grouping (Hay and

Diehl, 1999, 2007; Rice, 1992; Woodrow, 1909; see also Saffran *et al.*, 1996). Thus, a reduction in durational contrasts can be seen as beneficial for stress-based segmentation, as it allows F0 to operate unimpeded. However, when pitch is neutralized, the remaining duration cues work against the MSS. Therefore, the relative contributions of acoustic cues to the MSS in English seem to be a function of not only the weight of these cues for stress perception *per se*, but also the grouping patterns promoted by these cues when listeners are confronted with continuous streams.

In this context, the different weights of vocalic, intonational, and durational cues for stress-based segmentation are likely to be language-dependent. Patterns of disruption are expected to depend on at least two factors: (1) The location and statistical predominance of stress in the words of the language and (2) the perceptual correlates of stress in that language, including both segmental and suprasegmental cues. For instance, it is likely that F0 disruptions in stress-initial Dutch, Finnish, and Hungarian will cause a reduced usage of the MSS similar to that observed in the present study, since all four languages exhibit a strong association between F0 and stress perception (Iivonen *et al.*, 1998; Siptar and Torkency, 2000; Vroomen *et al.*, 1998). However, the magnitude of the F0 disruption should be greater in Finnish and Hungarian than in English and Dutch, as the former are fixed-initial-stress languages whereas the latter only show a stress-initial predominance. Furthermore, we predict that the effect of equalizing duration should be even less consequential in Hungarian than in English because duration, in Hungarian, serves to indicate phonemic contrasts rather than stress (Siptar and Torkency, 2000). In contrast, the word-final lengthening typically observed in French words (Banel and Bacri, 1994; Delattre, 1966) should cause the equalization of duration to affect the segmentation of French more than that of English, whereas the flattening of F0 should have a relatively minor impact (see Welby, 2007, for a discussion on the optional word-initial F0 rise in French). Finally, this same line of logic can be extended to naturally degraded speech, such as that occurring from disease or disorder or as the result of foreign accent or dialect. It would hold that the patterns of degradation or disruption to the segmental and suprasegmental cues to syllable stress would have language-specific consequences to stress-based segmentation.

VI. CONCLUSION

This study provides additional evidence for listeners' reliance on stress for segmenting words from degraded speech. However, the acoustic cues generally held to support stress perception showed differential effects on the application of stress-based segmentation. Equalizing vowel durations showed less detrimental effects than flattening F0 or reducing vowel identity, which demonstrates the leading role of the latter two for segmentation. Further, the results suggest that the MSS might be slightly more segmentally than supra-segmentally guided (Cutler, 1986), with vowel identity outweighing F0 and duration in the effective use of stress-based lexical segmentation.

ACKNOWLEDGMENTS

This research was supported by Research Grant Nos. 5R29 DC02672 and 5R01 DC6859 from the National Institute on Deafness and Other Communicative Disorders, National Institutes of Health (J.M.L.), as well as a grant from the Biotechnology and Biological Sciences Research Council (BBSRC, 7/S18783) to S.L.M. Gratitude is extended to Stephen Goldinger, Pamela Mathy, and Michael Dorman, for their significant contributions to the development of this project.

APPENDIX A: TEST PHRASES (STRONG VOWELS ARE SHOWN IN BOLD AND UNDERLINED)

amend the slower page
 his display collects it
 she describes a nuisance
call a random voter
nothing more of humor
release the legal batch
behind the broken arch
forgot the dinner place
assume a sanction change
ballot formal circles
 to rest and not appear
never shout forgive them
 the trial can destroy
soon the men were asking
apply the bill within
elect against your time
jumping for the lesson
union was beneath her
test a law for methods
alive and eager smile
 the giants are secure
convince the council here
we debate the cover
resign tonight at last
remove the boat across
 a term arranged inside
attack a drifting heart
simple for the congress
impeach the new results
sticks are best for pencils
 the tariff on the edge
judge upon the discount
watch him join the caucus
until the nation fell
rings amused a teacher
compose respect and leave
choose a course of justice
growing voice for taxes
 will travel after court
include a short refrain

APPENDIX B: INCREASES IN PHRASE DURATION AFTER MANIPULATIONS

Phrase name	Original (s)	Manipulated (s)	Difference (s)
Amend	1.2863	1.9787	0.6924
Arch	1.3145	1.7999	0.4854
Ballot	1.3257	1.9535	0.6278
Bill	1.2486	1.7543	0.5057
Boat	1.4143	1.8767	0.4624
Caucus	1.5364	2.1374	0.6010
Compose	1.8791	2.7767	0.8976
Congress	1.3884	1.8526	0.4642
Council	1.5640	2.2931	0.7291
Court	1.7451	2.3135	0.5684
Debate	1.3088	1.8526	0.5438
Dinner	1.5542	2.0549	0.5007
Display	1.5835	2.1749	0.5914
Elect	1.5265	1.9473	0.4208
Forgive	1.4857	2.0249	0.5392
Giant	1.4708	2.0175	0.5467
Heart	1.3398	1.8053	0.4655
Humor	1.5047	2.0492	0.5445
Impeach	1.7713	2.1827	0.4114
Judge	1.3195	1.8227	0.5032
Justice	1.7013	2.2949	0.5936
Law	1.5186	2.1533	0.6347
Legal	1.5600	2.1299	0.5699
Lesson	1.2451	1.6589	0.4138
Men	1.5360	2.0399	0.5039
Nation	1.4934	1.9672	0.4738
Nuisance	1.8397	2.2899	0.4502
Pencils	1.8001	2.1725	0.3724
Refrain	1.5973	2.4899	0.8926
Resign	1.6806	2.1599	0.4793
Rest	1.5679	1.9199	0.3520
Ring	1.4365	1.8913	0.4548
Sanction	1.8430	2.2255	0.3825
Smile	1.4800	1.9699	0.4899
Tariff	1.4686	2.0515	0.5829
Taxes	1.5103	2.0927	0.5824
Term	1.6217	2.5199	0.8982
Trial	1.5923	2.2199	0.6276
Union	1.5533	2.3759	0.8226
Voter	1.5698	2.3594	0.7896
Mean	1.5296	2.0913	0.5617
(s. d.)	(0.1652)	(0.2306)	(0.1422)

¹Although Cutler and Butterfield (1992) did not calculate a MSS ratio on their data, examination of their Table 6 (p. 228) indicates a value of approximately 0.75. Their LBEs were derived from transcripts of normal speech presented at very low sound level.

Banel, M-H., and Bacri, N. (1994). "On metrical patterns and lexical parsing in French," *Speech Commun.* **15**, 115–126.

Bond, Z. S. (1981). "Listening to elliptic speech: Pay attention to stressed vowels," *J. Phonetics* **9**, 89–96.

Cutler, A. (1986). "Forbear is a homophone: Lexical prosody does not constrain lexical access," *Lang Speech* **29**, 201–220.

Cutler, A., and Butterfield, S. (1990). "Durational cues to word boundaries in clear speech," *Speech Commun.* **9**, 485–495.

Cutler, A., and Butterfield, S. (1992). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," *J. Mem. Lang.* **31**, 218–236.

Cutler, A., and Carter, D. M. (1987). "The predominance of strong syllables in the English vocabulary," *Comput. Speech Lang.* **2**, 133–142.

Cutler, A., and Norris, D. G. (1988). "The role of strong syllables in segmentation for lexical access," *J. Exp. Psychol. Hum. Percept. Perform.* **14**, 113–121.

Delattre, P. (1966). *Studies in French and Comparative Phonetics* (Mouton, La Haye).

Fear, B., Cutler, A., and Butterfield, S. (1995). "The strong/weak syllable distinction in English," *J. Acoust. Soc. Am.* **97**, 1893–1904.

Fry, D. B. (1955). "Duration and intensity as physical correlates of linguistic stress," *J. Acoust. Soc. Am.* **27**, 765–768.

Hammen, V. L., Yorkston, K. M., and Minifie, F. D. (1994). "Effects of temporal alterations on speech intelligibility in Parkinsonian dysarthria," *J. Speech Hear. Res.* **37**, 244–253.

Handel, S. (1989). *Listening* (MIT, Cambridge, MA).

Hay, J. F. and Diehl, R. L. (1999). "Effect of duration, intensity, and F0 alternation on rhythmic grouping," *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, CA.

Hay, J. F., and Diehl, R. L. (2007). "Perception of rhythmic grouping: Testing the Iambic/Trochaic Law," *Percept. Psychophys.* **69**, 113–122.

Hertrich, I., and Ackermann, H. (1998). "Auditory perceptual evaluation of rhythm-manipulated and resynthesized sentence utterances obtained from cerebellar patients and normal speakers: A preliminary report," *Clin. Linguist. Phonetics* **12**, 427–437.

Hillenbrand, J. M. and Gayvert, R. T. (2005). "Open source software for experimental design and control," *J. Speech Lang. Hear. Res.* **48**, 45–60.

Hillenbrand, J., Getty, L., Wheeler, K., and Clark, M. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.

Hoequist, C. (1983a). "Durational correlates of linguistic rhythm categories," *Phonetica* **40**, 19–31.

Hoequist, C. (1983b). "Syllable duration in stress-, syllable- and mora-timed language," *Phonetica* **40**, 203–237.

Iivonen, A., Niemi, T., and Paananen, M. (1998). "Do F0 peaks coincide with lexical stresses? A comparison of English, Finnish, and German," edited by Stefan Werner, *Nordic Prosody. Proceedings of the VII Conference, Joensuu 1996* (Peter Lang, Frankfurt), pp. 142–158.

Kewley-Port, D., and Zheng, Y. (1999). "Vowel formant discrimination: Towards more ordinary listening conditions," *J. Acoust. Soc. Am.* **106**, 2945–2958.

Laures, J., and Weismer, G. (1999). "The effects of flattened fundamental frequency on intelligibility at the sentence level," *J. Speech Lang. Hear. Res.* **42**, 1148–1156.

Lehiste, I. (1970). *Suprasegmentals* (MIT, Cambridge, MA).

Lehiste, I. (1972). "The timing of utterances and linguistic boundaries," *J. Acoust. Soc. Am.* **51**, 2018–2024.

Lieberman, P. (1960). "Some acoustic correlates of word stress in American English," *J. Acoust. Soc. Am.* **32**, 451–454.

Liss, J. M., Spitzer, S. M., Caviness, J. N., and Adler, C. (2000). "LBE analysis in hypokinetic and ataxic dysarthria," *J. Acoust. Soc. Am.* **107**, 3415–3424.

Liss, J. M., Spitzer, S. M., Caviness, J. N., and Adler, C. (2002). "The effects of familiarization on intelligibility and lexical segmentation of hypokinetic and ataxic dysarthria," *J. Acoust. Soc. Am.* **112**, 3022–3031.

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., and Edwards, B. W. (1998). "Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech," *J. Acoust. Soc. Am.* **104**, 2457–2466.

Mattys, S. L. (2000). "The perception of primary and secondary stress in English," *Percept. Psychophys.* **62**, 253–265.

Mattys, S. L. (2004). "Stress versus coarticulation: Towards an integrated approach to explicit speech segmentation," *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 397–408.

Mattys, S. L., White, L., and Melhorn, J. F. (2005). "Integration of multiple segmentation cues: A hierarchical framework," *J. Exp. Psychol. Gen.* **134**, 477–500.

McQueen, J. M., and Cutler, A. (2001). "Spoken word access processes: An

- introduction," *Lang. Cognit. Processes* **16**, 469–490.
- McQueen, J. M., Norris, D., and Cutler, A. (1994). "Competition in spoken word recognition: Spotting words in other words," *J. Exp. Psychol. Learn. Mem. Cogn.* **20**, 621–638.
- Mens, L. H., and Povel, D. J. (1986). "Evidence against a predictive role for rhythm in speech perception," *Q. J. Exp. Psychol. A* **38a**, 177–192.
- Nakatani, L. H. and Shaffer, J. A., (1978) "Hearing words without words: Prosodic cues for word perception," *J. Acoust. Soc. Am.* **63**, 234–245.
- Neary, T. M. (1989). "Static, dynamic, and relational factors in vowel perception," *J. Acoust. Soc. Am.* **85**, 2088–2113.
- Norris, D., McQueen, J. M., and Cutler, A. (1995). "Competition and segmentation in spoken-word recognition," *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 1209–1228.
- Norris, D., McQueen, J., Cutler, A., and Butterfield, S. (1997). "The possible-word constraint in the segmentation of continuous speech," *Cogn. Psychol.* **34**, 193–243.
- Quené, H., and Koster, M. (1998). "Metrical segmentation in Dutch: Vowel quality or stress?" *Lang Speech* **41**, 185–202.
- Rice, C. C. (1992). "Binary and ternarity in metrical theory: Parametric extensions," Doctoral dissertation, University of Texas, Austin, TX.
- Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). "Word segmentation: The role of distributional cues," *J. Mem. Lang.* **35**, 606–621.
- Shatzman, K. B., and McQueen, J. M. (2006). "Prosodic knowledge affects the recognition of newly acquired words," *Psychol. Sci.* **17**, 372–377.
- Siptar, P., and Torkency, M. (2000). *The Phonology of Hungarian* (Oxford University Press, New York).
- Smith, M., Cutler, A., Butterfield, S., and Nimmo-Smith, I. (1989). "The perception of rhythm and word boundaries in noise-masked speech," *J. Speech Hear. Res.* **32**, 912–920.
- Spitzer, S. (2007). "Synthetic speech," <http://www.asu.edu/clas/shs/liss/> (accessed 27 July 2007).
- Streeter, L. A. (1978). "Acoustic determinants of phrase boundary perception," *J. Acoust. Soc. Am.* **64**, 1582–1592.
- Thiessen, E. D., and Saffran, J. R. (2003). "When cues collide: Statistical and stress cues in infant word segmentation," *Dev. Psychol.* **39**, 706–716.
- Thiessen, E. D., and Saffran, J. R. (2004). "Spectral tilt as a cue to word segmentation in infancy and adulthood," *Percept. Psychophys.* **66**, 779–791.
- Vroomen, J., Tuomainen, J., and de Gelder, B. (1998). "The roles of word stress and vowel harmony in speech segmentation," *J. Mem. Lang.* **38**, 133–149.
- Welby, P. (2007). "The role of early fundamental frequency and elbows in French word segmentation," *Speech Commun.* **49**, 28–48.
- Woodrow, H. (1909). "A quantitative study of rhythm: The effect of variations in intensity, rate and duration," *Arch. Psychol. (Frankf)* **14**, 1–66.

On the perception of similarity among talkers

Robert E. Remez^{a)}

Department of Psychology, Barnard College, 3009 Broadway, New York, New York 10027

Jennifer M. Fellowes

Department of Psychiatry, New York Presbyterian Hospital, 180 Ft. Washington Avenue, New York, New York 10032

Dalia S. Nagel

Department of Ophthalmology, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, New York 10029

(Received 14 November 2006; revised 24 September 2007; accepted 25 September 2007)

A listener who recognizes a talker notices characteristic attributes of the talker's speech despite the novelty of each utterance. Accounts of talker perception have often presumed that consistent aspects of an individual's speech, termed *indexical* properties, are ascribable to a talker's unique anatomy or consistent vocal posture distinct from acoustic correlates of phonetic contrasts. Accordingly, the perception of a talker is acknowledged to occur independently of the perception of a linguistic message. Alternatively, some studies suggest that attention to attributes of a talker includes indexical linguistic attributes conveyed in the articulation of consonants and vowels. This investigation sought direct evidence of attention to phonetic attributes of speech in perceiving talkers. Natural samples and sinewave replicas derived from them were used in three experiments assessing the perceptual properties of natural and sine-wave sentences; of temporally veridical and reversed natural and sine-wave sentences; and of an acoustic correlate of vocal tract scale to judgments of sine-wave talker similarity. The results revealed that the subjective similarity of individual talkers is preserved in the absence of natural vocal quality; and that local phonetic segmental attributes as well as global characteristics of speech can be exploited when listeners notice characteristics of talkers.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799903]

PACS number(s): 43.71.Bp, 43.71.Sy, 43.71.An [MSS]

Pages: 3688–3696

I. ON THE PERCEPTION OF SIMILARITY AMONG TALKERS

The perception of characteristic attributes of an individual talker is customarily considered to occur independently of the resolution of the linguistic properties of utterances (Abercrombie, 1967; Bricker and Pruzansky, 1976; Halle, 1985). Conceptualizations of linguistic perception have typically appealed to a function by which a perceiver discovers linguistic form despite the variation in correspondence between a linguistic property and its acoustic correlates, whether the variation is attributed to coarticulation of phonetic segments or to affective or anatomical properties of a talker. Concurrently, the perception of a talker from a speech sample is said to be a kind of second message carried by an utterance, and perception of talker-specific attributes to depend on the resolution of *indexical* acoustic properties common to all of a talker's utterances, special to none. Classic approaches to indexical perception have offered evidence of perceptual sensitivity to acoustic correlates of size variation in vocal anatomy across talkers (Fant, 1966; Ladefoged and Broadbent, 1957; reviewed by Kreiman, 1997; and Pisoni, 1997). Others have appealed to persistent characteristics of use rather than a direct effect of anatomy as a reliable cause of long-term qualitative consistency in a speaker's

speech, such as a tendency to whispery voice, or heavy nasality (Laver, 1980; Nolan, 1983). Research on qualitative variation of voices similarly has implicated basic auditory attributes, such as the pitch and pitch range of phonatory frequency, the timbre of the voice, as well as more abstract attributes, such as vocal strength or melodiousness, in the identification of talkers by ear (for example, Krauss *et al.*, 2002).

Recent refinements of this approach have included specification of the acoustic attributes unique to female voices (Klatt and Klatt, 1990) independent of linguistic contrasts, and estimation of the effects of variation in vowel spectra on phoneme quality (Frieda *et al.*, 1999) independent of specific individual talkers. A dissociation of linguistic and indexical perception in the effects of brain activity and injury (Belin *et al.*, 2004; Neuner and Schweinberger, 2000; Stevens, 2004; Van Lancker *et al.*, 1988) has also granted a biological license to the speculation that linguistic and indexical perception are distinct functions fed by different sensory attributes: Short-term elements pertain to symbolic contrasts, and long-term characteristics pertain to distinctions among talkers.

A. Identifying a sine-wave talker

Despite evidence of this dissociation of linguistic and indexical perceptual functions, each devoted to its kind of acoustic correlate and each promoting the perception of dif-

^{a)}Electronic mail: remez@columbia.edu

ferent attributes of speech, some findings oppose the hypothetical independence of linguistic perceptual analysis and the perception of individual characteristics. Common among these is the premise that some grammatically regulated subphonemic properties of phonetic expression are nested within dialect and idiolect, and if perceivers notice, track and remember these linguistic attributes they can serve as indexical properties even though they do not stem from anatomical or physiological differences among talkers. Such findings have shown that perceivers require a phonetically diverse sample in order to resolve some distinctive aspects of individual talkers (Pollack *et al.*, 1954), and that familiarity with a specific talker promotes linguistic resolution of novel utterances (Lieberman, 1963; Nygaard *et al.*, 1994; Smith, 2004; also, see Hawkins, 2003; and a review by Pardo and Remez, 2006). In some tests of this approach to individual and linguistic identification, sine-wave replicas of natural speech (Fellowes *et al.*, 1997; Remez *et al.*, 1997; Sheffert *et al.*, 2002; see, also, Brungart *et al.*, 2006) were used to assess the conditions in which linguistically regulated attributes of speech promote the perception of talkers as well as words. These projects revealed that both strangers and long-time acquaintances could identify talkers under listening conditions in which the acoustic correlates of natural voice quality were eliminated from the speech samples. The use of sine-wave replicas of natural samples permitted such tests, in which a tone was set equal in frequency and amplitude to each of the three lowest oral resonances. A three-tone replica of a natural utterance fails to evoke an impression of natural vocal timbre, though it preserves the spectrotemporal variation sufficient to elicit impressions of detailed phonetic characteristics of consonants and vowels. Because sine-wave replicas of speech spectra are produced without setting a tone analog to match the fundamental frequency of the natural model, listeners in these tests identified talkers without relying on veridical impressions of vocal pitch (Remez and Rubin, 1984, 1993), although this attribute of speech has been favored as a perceptually and forensically useful acoustic correlate of individual differences (Bricker and Pruzansky, 1976; Hollien, 2002; von Dommelen, 1987). Indeed, Fellowes *et al.* (1997) eliminated the acoustic correlates of vocal tract scale variation as well, by transposing the tone analogs of the formants in frequency. Although this condition imposed an acoustic constraint impairing the perceptual identification of a talker's sex, listeners were able to identify most individuals nonetheless (Fellowes *et al.*, 1997). These findings show that many of the acoustic correlates of voice quality that had been candidates for use as indexical information are unnecessary for individual identification by listeners.

Overall, this pattern of results encouraged the hypothesis that listeners can exploit phonetic attributes in remembering and identifying individual talkers. Because habits of phonetic expression are presumably consistent in the speech of an individual and are distinct among individuals whether by dialect or idiolect, an intelligible utterance, even one that is distorted or anomalous in timbre, presents a potential for distinguishing talkers perceptually (Remez, *in press*). Of

course, linguistically regulated subphonemic phonetic attributes are also useful in recognizing spoken words (Luce *et al.*, 2000).

This claim that talkers are identifiable by idiolect without access to qualitative attributes of the voice was based on indirect and opportunistic evidence. Performance levels in tests of individual identification from sine-wave samples were high, especially when listeners were personally acquainted with the talkers to be identified (Remez *et al.*, 1997; cf. Sheffert *et al.*, 2002). The distribution of errors of identification was used to estimate the perceptual similarity among the ten individuals whose utterances composed the set of test materials. The basic finding can be formulated succinctly: No single acoustic feature or dimension of variation of the sine-wave samples matched the perceptual similarity estimated from errors of identification. Neither similarity in duration, nor in central spectral tendency, nor in the spectral distribution of stressed vowels, nor in the rate of syllable production explained the similarity analyses based on misidentification of sine-wave talkers.

Although this pattern of results encouraged it, the hypothesis of an idiolectal cause for the patterns of similarity was an especially risky speculation because it was proposed on the evidence of very few data. Errors were rare given such good performance. The present project sought to improve the quality of this evidence about the causes of subjective similarity of talkers without natural vocal timbre. The three experiments of this report use direct assessments of subjective similarity and a variety of utterance types in tests that aimed to calibrate the perceived similarity of talkers without impressions of natural vocal quality.

B. The present tests

Three experiments investigated the apparent similarity among a set of talkers, in an attempt to explain prior findings about the identification of talkers from sine-wave samples. In Experiment 1, simple and direct reports of similarity were used to estimate the apparent similarity among the set of ten talkers used in the empirical precedents (Fellowes *et al.*, 1997; Remez *et al.*, 1997; Sheffert *et al.*, 2002) with two kinds of acoustic signals, natural samples and sine-wave replicas, and two different sentences. The finding of this study is that the pattern of perceptual similarity among a group of talkers is largely preserved over an acoustic transformation from natural sample to sine-wave signal. In Experiment 2, the contribution of vocal quality and phonetic inventory was estimated in tests that used temporally reversed natural and sine-wave samples. The results of this study showed that the loss of veridical phonetic properties affected impressions of similarity of natural and sine-wave samples alike, though a core of subjective similarity remained. Experiment 3 aimed to ascertain the role of the tone analog of the first formant, an index of the anatomical scale of a talker (Goldstein, 1980), in the perception of similarity among sine-wave talkers. These results showed that this single component of the tone complex of a sine-wave sentence does contribute to impressions of similarity of sine-wave talkers, as if it is used perceptually to index consistent qualitative differences among them.

TABLE I. Mean frequency and dispersion measures of the three lowest vocalic formants of the sentences used as natural samples or as models for sine-wave synthesis. Ten talkers each contributed two sentences. Male talkers are designated by M, female by F. The two sentences are: Yell: The drowning man let out a yell. Scarves: The scarves were made of shiny silk.

Talker	Sentence	First formant				Second formant				Third formant			
		Mean	s.d.	High	Low	Mean	s.d.	High	Low	Mean	s.d.	High	Low
M ₁	Yell	480.1	133.4	693	280	1496.3	336.0	2330	914	2438.9	233.0	2817	1651
	Scarves	486.8	98.5	607	277	1428.8	289.6	2013	937	2186.4	233.2	2482	1666
M ₂	Yell	488.1	175.1	784	192	1690.1	362.0	2352	961	2623.2	204.3	3165	2085
	Scarves	475.4	123.9	635	244	1570.9	415.2	2377	732	2394.5	405.6	3061	1726
M ₃	Yell	488.9	170.5	783	160	1619.1	323.6	2269	1004	2410.8	206.9	2831	2068
	Scarves	434.4	119.7	608	197	1454.3	312.5	2023	871	2188.2	185.3	2746	1907
M ₄	Yell	493.8	153.8	802	203	1583.6	350.9	2376	910	2580.5	294.9	3101	1620
	Scarves	500.1	128.7	706	305	1453.0	349.3	2220	822	2297.9	377.7	2830	1360
M ₅	Yell	450.5	141.4	707	168	1390.1	278.0	2087	858	2277.5	256.9	2760	1430
	Scarves	439.6	112.3	616	264	1372.7	297.9	1848	510	2162.3	272.7	2517	1549
F ₁	Yell	511.0	218.0	945	259	1926.0	414.7	2720	1243	3017.1	337.8	3549	1878
	Scarves	485.7	121.1	678	271	1750.3	432.6	2655	1162	2734.7	415.8	3197	1802
F ₂	Yell	488.9	166.5	864	263	1939.9	399.2	2650	1165	2906.8	230.8	3458	2199
	Scarves	648.3	111.6	932	424	1764.5	407.6	2760	1216	2841.7	287.8	3281	2203
F ₃	Yell	634.0	220.1	1024	245	1734.6	392.1	2650	969	2663.6	342.4	3442	1584
	Scarves	633.5	101.9	798	361	1622.6	484.9	2566	855	2431.8	361.1	2984	1787
F ₄	Yell	602.3	203.6	996	187	1995.4	479.9	2875	1240	2976.1	241.3	3458	2443
	Scarves	662.2	101.7	803	381	1814.1	361.1	2590	1084	2773.5	266.9	3132	2188
F ₅	Yell	504.7	159.1	794	281	1820.8	294.5	2450	1225	2982.7	309.9	3774	2218
	Scarves	634.9	146.2	812	375	1722.1	412.9	2520	1104	2766.3	127.2	3062	2458

Nonetheless, a test of partial correlation showed that the similarity of sine-wave talkers cannot be attributed solely to impressions of vocal pitch or scale evoked by the tone analog of the first formant. Taken together, the results indicate a role of global acoustic properties of speech as well as local phonetic segmental attributes in the perceptual similarity observed among these talkers.

II. EXPERIMENT 1

A. Simple reports of similarity

1. Method

a. Test materials. This test used natural samples of spoken sentences and sine-wave replicas modeled on them. The natural samples of two sentences, “The drowning man let out a yell,” and “The scarves were made of shiny silk,” were spoken within a list of sentences by each of ten talkers, five males and five females. The talker ensemble was relatively heterogeneous, representing different American English and British English dialects. Each talker practiced speaking the sentences in the list, and the test used samples of utterances that were produced comfortably and without dysfluency. The sentences were recorded in a sound-attenuating chamber, low-pass filtered at 4.5 kHz, digitally sampled at 10 kHz, equated for amplitude, and stored as sampled data with 12-bit amplitude resolution.

Sine-wave replicas of the 20 natural speech tokens were created by estimating the frequencies and amplitudes at 5-ms intervals of the three oral formants and the intermittent nasal and fricative formants, relying interactively on two representations of the spectrum: (1) Linear predictive coding and (2) the discrete Fourier transform. The derivation of these estimates was performed by hand, with multiple passes for correction of errors. Three time-varying sinusoids were

then synthesized from these estimates to replicate the oral and nasal formant pattern, and a fourth sinusoid to replicate an intermittent fricative pole, based on the center frequencies and amplitudes obtained in the acoustic analysis (Rubin, 1980; two sine-wave synthesizers are available at <http://www.haskins.yale.edu/featured/sws/information.html>).

(Table I presents a summary of the measures of each talker and sentence.) The sine-wave synthesis procedure preserved patterns of spectrotemporal change of the vocal resonances, while eliminating the fundamental frequency, harmonic relations, and fine-grained spectral details of natural speech. Subjectively, sine-wave sentences are intelligible though unnatural in vocal quality (Remez *et al.*, 1981, 2002). The 20 natural samples and 20 sine-wave sentences were stored in sampled data format; they were sequenced for use in testing and converted from digital records to analog signals delivered to each listener at a nominal level of 65 dB SPL over a calibrated Telephonics TDH-39 headset.

b. Procedure. Each volunteer listener participated in a single test appraising the likeness of the ten talkers in the set. There were four tests overall, two natural (Natural Yell and Natural Scarves) and two sine-wave (Sine-wave Yell and Sine-wave Scarves). Participation was blocked by test, and each listener was assigned randomly to a single acoustic type and sentence. Within each test, the trial structure was identical. A pair of sentences each produced by a different talker of the ten was presented separated by 1 s of silence. The listener then indicated the subjective likeness of the two talkers, using a five-point scale. Along a Similarity scale, the listener reported that the two talkers on the trial were *very similar* (a value of 1) or *not very similar* (a value of 5). Along a Dissimilarity scale, the listener reported that the two talkers on the trial were *very dissimilar* (a value of 1) or *not very dissimilar* (a value of 5). A participant used a single

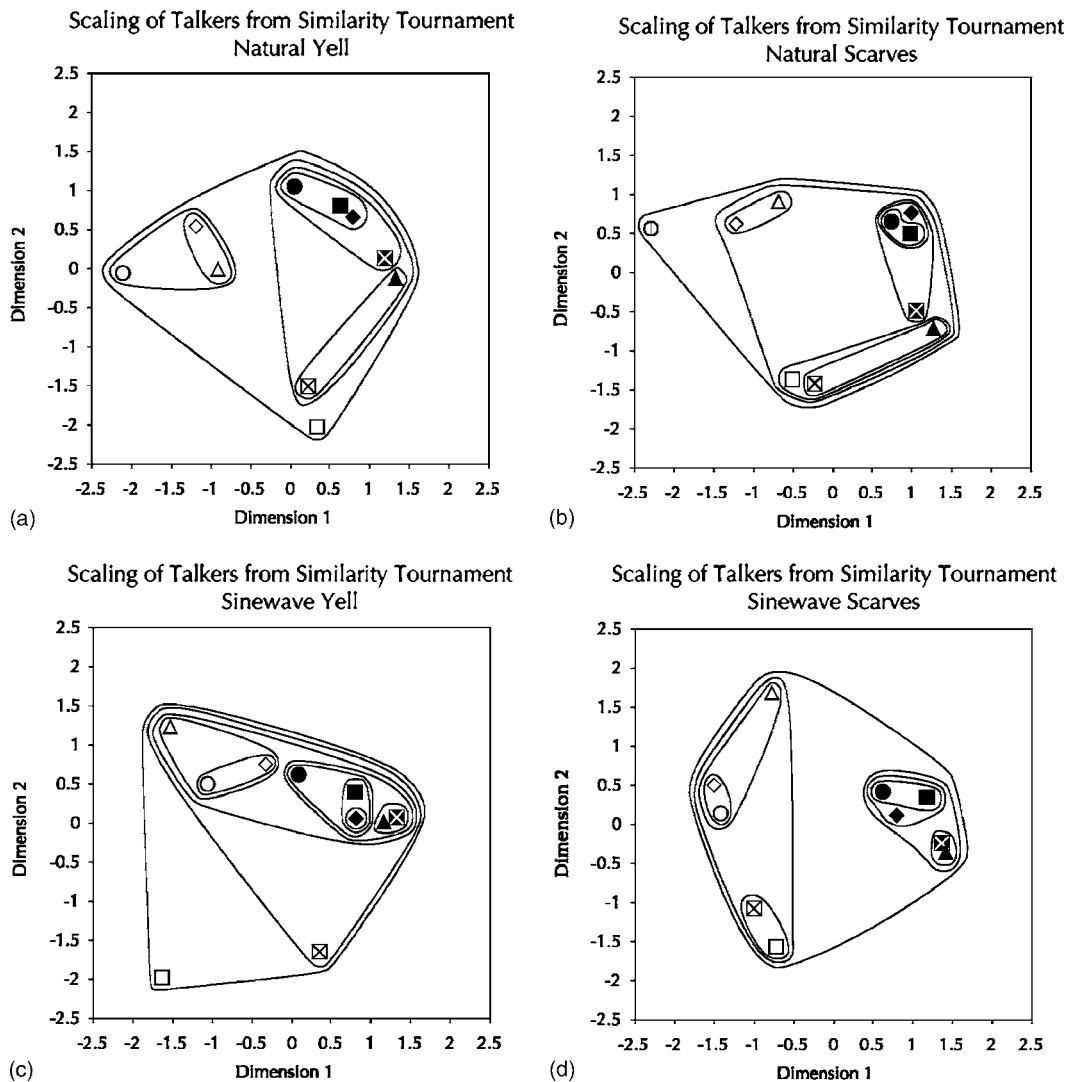


FIG. 1. Multidimensional scaling and hierarchical clustering analysis of the group data for four tests of similarity in Experiment 1: (a) With natural samples of the sentence, “The drowning man let out a yell.” (b) With natural samples of the sentence, “The scarves were made of shiny silk.” (c) With sine-wave replicas of “The drowning man...” (d) With sine-wave replicas of “The scarves were made...,” Tests of hypotheses use ranked similarities derived from the hierarchical clustering analyses. In each panel, the placement of the bullets represents the solution of the scaling analysis; the placement of the enclosing curves represents the solution of the clustering analysis. Black bullets are male talkers; white bullets are female talkers.

scale, Similarity or Dissimilarity, for the entire procedure. There were 3 s of silence between successive trials, and 6 s of silence after every tenth trial.

To compose a sequence in which each of the ten talkers was presented with every other required 90 trials, 45 in each order. Each of these 90 trials occurred twice in a test session, composing a procedure of 180 trials, overall.

c. Subjects. One hundred and eleven students registered in the Barnard College Subject Pool participated in the listening tests of Experiment 1. Each was a right-handed native speaker of English reporting no history of language disorder. None was familiar with sine-wave replicas of speech. Course credit was granted for participation.

B. Results and discussion

With so few reports per subject, the measures were pooled within condition, and statistical analyses were performed solely on the group performance. The reports of three participants who failed to follow instructions were excluded from the data set.

Across the four conditions, the data treatment was the same. The results of the tests using a Dissimilarity scale (1=*very dissimilar*, 5=*not very dissimilar*) did not differ from those of the tests using a Similarity scale, and were reflected to coincide with the Similarity scale, in which the more dissimilar a pair was judged to be, the greater the value of the index. A square matrix was created for each of the four test conditions, Natural Yell, Natural Scarves, Sine-wave Yell, and Sine-wave Scarves, representing the accumulated reports of all 27 subjects in a group for each of the 90 ordered pairs of talkers. Each group matrix was analyzed using the technique of multidimensional scaling (Kruskal, 1964) to produce a two-dimensional representation of each of the four similarity tournaments. The solutions of the scaling analyses are shown in the four panels of Fig. 1.

Each group matrix was also used to derive a hierarchical clustering (Johnson, 1967) of the ten talkers in each condition, producing a likeness classification implicit in the similarity reports of the subjects. The dendrogram of the cluster

analysis was used to produce a similarity ranking in each of the four tests, for testing the principal hypotheses of this study.

Overall, the subjects produced consistent and differentiated reports of similarity, as shown by the indices of variance accounted for in the scaling solutions: For Natural Yell, $r^2=0.76$; for Natural Silk, $r^2=0.78$; for Sine-wave Yell, $r^2=0.87$; for Sine-wave Silk, $r^2=0.96$. To determine the role of natural vocal timbre in the assessment of similarity, we took the performance of the natural version of each sentence as a standard, and compared it to the sine-wave version. This was achieved by testing for correlated similarity ranks obtained in the cluster analyses for natural and sine-wave variants of Yell (Spearman's $r=0.85$, $p<0.01$) and Silk (Spearman's $r=0.92$, $p<0.01$). This outcome shows that listeners largely judged the ten talkers to be similar in the same way whether their judgments were concurrent with impressions of natural vocal quality or not.

Evidence of an influence of vocal timbre on the judgment of similarity among the talkers would be seen in a differential contribution of the phonetic sample to natural and sine-wave signals. To estimate this, we compared the performance of the two natural sentences to each other (Spearman's $r=0.92$, $p<0.01$) and the two sine-wave sentences to each other (Spearman's $r=0.76$, $p<0.01$). The stronger correlation of the natural sentences is arguably an indication that some talker attributes driving perceptual similarity in natural samples are preserved despite a difference in phonetic inventory distinguishing the sentences Yell and Scarves. Sine-wave sentences do not evoke impressions of natural vocal quality, in contrast to natural utterances. Accordingly, we expected to find a smaller correlation between the two sine-wave sentences, an effect of their different segmental inventory, than between the two natural sentences, which share qualitative attributes of vocal sound production, however the segmental inventory varies.

Two attributes of each talker that are plausibly preserved over variation in phonetic inventory are the vocal quality and the pitch height and range of phonation. Neither is present in sine-wave sentences, and the listener is obliged to resolve a talker's characteristics by attending to other attributes. The phonetic samples differ in each sentence, and because variation in phonetic samples is known to influence assessments of a talker's identity (Lieberman, 1963; Pollack *et al.*, 1954) it is reasonable to expect the similarity classification to follow the variation in the phonetic sample in the absence of other sustaining attributes. Nonetheless, this relative difference between natural and sine-wave similarity does not negate the clear indication that listeners were consistent in their overall performance in natural and sine-wave conditions. Whether this can be attributed to consistency in individual expression of segmental phonetic properties or to other more global aspects of vocal sound production (Laver, 1980) can only be resolved by additional tests.

In order to determine the contribution of qualitative attributes and of phonetic form to the natural and sine-wave performance in Experiment 1, Experiment 2 used a test with temporally reversed natural samples of Yell of each acoustic type. Although temporal reversal abolishes lexical access and

harshly distorts the apprehension of many consonants, it preserves vowels at the syllable nuclei, and retains the glottal spectrum associated with impressions of vocal timbre along with the range of frequency variation of phonation. To the extent that assessments of similarity depend on these attributes alone, the performance for temporally veridical and reversed natural samples should not differ (cf. Van Lancker *et al.*, 1985). In the case of sine-wave replicas of speech, though, performance would be especially affected by temporal reversal if the appraisal of similarity rests on the phonetic details disrupted by reversal.

III. EXPERIMENT 2

A. Similarity of temporally reversed signals

1. Method

a. Test materials. Temporally reversed versions of the ten natural and ten sine-wave sentences in the Natural Yell and Sine-wave Yell sets were created by inverting the order of the digital records composing the files used in the temporally veridical conditions of Experiment 1. The result of this method applied to natural samples produced test items that preserved the glottal period and the spectrum variation of the natural and sine-wave sentences, arguably preserving quality while abolishing lexical access and perceptual resolution of many of the segmental phonetic attributes. The reversed sine-wave sentences were no less natural in vocal quality than the temporally veridical versions. However, temporal inversion affected the time-critical acoustic correlates of many consonants and gliding vowels, while sparing the tonal replicas of quasi-steady-state vowels at syllable nuclei and consonants that evolve more slowly, such as those of nasal and fricative manner.

b. Procedure. Two similarity tournaments were prepared from the reversed sentences, Reversed Natural Yell and Reversed Sine-wave Yell. Participation was blocked by test, and each listener was assigned randomly to a single acoustic type. Within each test, the structure of the trials and of the tests paralleled the procedure of Experiment 1.

c. Subjects. Thirty-five volunteers registered in the Barnard College Subject Pool participated in the listening tests of Experiment 2. Each was a right-handed native speaker of English reporting no clinical history of language disorder. No volunteer had participated in Experiment 1, nor was any familiar with sine-wave replicas of speech. Course credit was granted for participation.

B. Results and discussion

A single participant was excluded from the group data for declining to complete the test session. This left 17 subjects in each group. The individual reports were pooled within condition, and statistical analyses were performed solely on the group performance, repeating the practice of the first experiment of this report.

Consistent with the observations in Experiment 1, the judgments of similarity in these two tests were consistent and differentiated, and the multidimensional scaling solutions were again effective in two dimensions (for Reversed Natural Yell, $r^2=0.93$; for Reversed Sine-wave Yell, $r^2=0.80$). To test the hypothesis that a portion of the similarity relations among the natural talkers was based on phonetic properties,

we compared the ranked similarities of the Natural Yell performance of the first experiment (derived from a hierarchical clustering analysis) with the ranked similarities of the Reversed Natural Yell items assessed in this procedure. The performance based on reversed samples was highly correlated with performance in the first experiment based on veridical samples (Spearman's $r=0.78$, $p<0.01$), indicating an effect of qualitative attributes evoked by natural speech samples in judgments of perceptual similarity. However, this picture of perceptual similarity is complicated by the correlated performance that was observed for temporally veridical Sine-wave Yell signals in Experiment 1 and the temporally reversed signals in this procedure (Spearman's $r=0.72$, $p<0.01$). A principle stating that the perceived similarity of sine-wave talkers derives from the accumulation of phonetic segmental details cannot explain the parallel effect of temporal reversal on apparent similarity with and without natural vocal quality.

If temporal reversal preserves timbre at the expense of time-critical phonetic segments, then the effects that were observed here on talker perception with sine-wave sentences warrant a similar explanation, at least initially. Perhaps listeners assessed talker similarity by attending to qualitative and phonetic attributes of the samples, natural and sine-wave alike. Although sine-wave replicas of speech do not conserve natural vocal timbre, the time-varying tones differ in average frequency and in frequency excursion over the test set. A listener accustomed to relying on multiple attributes—some long term and others short term—to appraise the similarity of talkers might do the same with the sine-wave cases as the natural cases, in contrast to our claim that phonetic attributes govern the perception of talkers when veridical qualitative attributes are unavailable (Fellowes *et al.* 1997; Remez *et al.*, 1997; Sheffert *et al.*, 2002). Although sine-wave replicas lack natural vocal quality, the frequency differences in tone components might evoke consistent differences in impressions of auditory form, affecting perceived similarity.

Empirical precedents found that variation in qualitative attributes was not required for perceptual identification of talkers from sine-wave replicas. For one, Fellowes *et al.* (1997) had demonstrated the identification of sine-wave talkers when the tone components had all been transposed in frequency to exhibit the same average values across the set, and with such items there can be little qualitative difference among the talkers, leaving phonetic properties as the likely grain for differentiation and identification of individuals. For another, Brungart *et al.* (2006) found that sine-wave sentences exhibit sufficient talker-specific characteristics to resist some effects of masking by a concurrent speech signal. In the present test, the sine-wave items replicated the natural frequency values estimated for vocal resonance, and across the test set each sentence evoked a potentially unique spectrum pitch despite the absence of a fundamental frequency, harmonic excitation, and broadband formants. Indeed, in a circumstance similar to this experiment, listeners reported an apparent vocal pitch of sine-wave sentences despite the absence of the familiar auditory excitation correlated with glottal pulsing. Tests revealed that the tone analog of the first formant supplied the acoustic correlate of sine-wave sen-

tence intonation (Remez and Rubin, 1984, 1993). This attribute of a sine-wave replica survives temporal reversal, and could be responsible in part for the finding of similar perceptual scaling of temporally veridical and reversed sine-wave sentences in the absence of natural vocal quality.

The third experiment of this series aimed to account for the consistency of sine-wave talker similarity by testing the role of the tone analog of the first formant, all other things being equal. In this study, a similarity tournament was conducted using this single constituent of the sine-wave replica of each of the ten talkers. These single time-varying sinusoids were not intelligible, and did not evoke a perceptual impression of vocalization (cf. Remez *et al.*, 1981). Nonetheless, in intelligible sine-wave utterances, this tone component plays a complex role, as the surrogate for a vocal resonance associated with consonant manner and voicing and with the height of vowels. This tone analog of the first formant also is responsible for impressions of the weird intonation that accompanies sine-wave utterances (Remez and Rubin, 1984, 1993). As a replica of the first formant, it is arguably an acoustic marker of variation in the pharyngeal cavity, which differs in scale among talkers (Goldstein, 1980; Stevens, 2004). One way to determine whether the auditory form of this tone contributed a qualitative impression of a sine-wave talker in our prior tests is to compare similarity judgments of this isolated tone to judgments of intact sentences. However, because isolated tones do not evoke phonetic impressions, an auditory form similarity judgment was used as the task. The outcome of this test held the potential to explain the ability of listeners to treat sine-wave replicas of utterances in two ways, as phonetic effects of idiolect of specific talkers, and as qualitative effects devoid of linguistic significance, much as the conventional view dichotomizing talker identification and phonetic perception warrants.

IV. EXPERIMENT 3

A. A test of the tone analog of the first formant

1. Method

a. Test materials. A set of new test items was developed from the ten sentences of the Sine-wave Yell set. This new set consisted of the tone analog of the first formant (T_1) of each of the ten sentences, synthesized as a single tone without concurrent sinusoids replicating the higher frequency formants. This was accomplished by taking the synthesis parameters for the tone analog of the first formant of each sentence replica and converting it to an analog signal composed of a single time-varying sinusoid. Each tone analog of a natural first formant realized the frequency and amplitude estimates of the original. Other aspects of preparation and delivery of acoustic test materials to listeners followed the practice of the first and second experiments.

b. Procedure. A single similarity tournament was used in this experiment, Sine-wave Yell T_1 . Within each test, the structure of the trials and of the tests paralleled the procedure of Experiment 1, although the instructions described the test items as electronic melodies. We asked the listener to compare two electronic melodies presented in each trial and then to appraise their likeness. Continuing the plan of the prior studies, some subjects reported similarity and others reported dissimilarity.

c. Subjects. Nineteen students registered in the Barnard College Subject Pool participated in the listening tests of Experiment 3. No volunteer had been tested in any other condition of this set of experiments, and none had previously encountered sine-wave replicas of speech. Each was a right-handed native speaker of English reporting no history of language disorder. Course credit was granted for participation.

B. Results and discussion

The reports of two participants who did not complete the procedure were excluded from the compiled data, leaving 17 subjects in the group. The individual reports were pooled and statistical analyses were performed solely on the group performance.

Multidimensional scaling solutions were effective in two dimensions (Sine-wave Yell T_1 , $r^2=0.89$). To test the hypothesis that the subjective similarity relations among sine-wave talkers reduces to pitch impressions or vocal tract scaling deriving from attention to T_1 , the tone analog of the first formant, we compared the ranked similarities of the Sine-wave Yell T_1 performance (determined by a hierarchical clustering analysis) in Experiment 3 with the ranked similarities of the Sine-wave Yell items assessed in Experiment 1. The classifications were highly correlated (Spearman's $r=0.84$, $p<0.01$), encouraging an account of the similarity among sine-wave talkers, to a first approximation, as due to similarity of the lowest frequency tone components. This outcome was surprising, because an analysis of the physical properties of the samples (Remez *et al.*, 1997) had previously shown that the talkers were not well distinguished acoustically by the average frequency, or frequency range, or dispersion of frequency variation of the first formant. Prior ventures in similarity scaling of talkers based on error data (Fellowes *et al.*, 1997; Remez *et al.*, 1997) had also disconfirmed a reliance on criterial auditory properties of the samples to determine the perceived similarities. To explain the discrepancy of the precedents and present findings, it is useful to consider the possibility that the apparent similarity of two sine-wave talkers is influenced by the frequency variation in T_1 , with additional attributes of the sine-wave samples contributing concurrently to the perception of the properties of talkers.

To estimate the extent to which the first formant and its tone analog determined the similarity scaling, we applied the findings of Experiments 1 and 3 in a test of partial correlation. This permitted us to assess the degree of agreement of similarity judgments evoked by natural and by sine-wave samples of the talkers while excluding the contribution of the tone analog of the lowest resonance, at least insofar as the perceptual tests in Experiment 3 calibrated it. A test of partial rank correlation based on the hierarchical clustering analyses found that the performance of the Natural Yell and the Sine-wave Yell sets remained correlated (Kendall's $\tau=0.31$) when the contribution of the tone analog of the first formant, as assessed in the Sine-wave Yell T_1 data, was neutralized.

This analytical strategy revealed that the congruent classification of sine-wave and natural talkers does not stem solely from the sensory appraisal of the first formant as a global property by which to scale a talker's vocal size. In-

stead, the finding of a substantial degree of correspondence without this factor exposes the room for other properties to influence the impressions of indexical attributes of individuals. Overall, Experiment 3 provides evidence of the relative contributions of global, qualitative factors and local, perhaps idiolectal factors in the perception of a talker from a speech sample. It seems fair to conclude that a variety of impressions determined the resolution of indexical attributes when a listener encountered a talker on a trial in our tests. Among these is an impression of scale evoked by global variation in the lowest resonance, as this test shows, as well as an impression of dialect or idiolect that is not reducible to the perception of vocal tract scale.

V. GENERAL DISCUSSION

How do listeners perceive similarity among talkers? This question was motivated by an unusual outcome in a set of tests of talker identification from sine-wave signals. Although it has become customary to refer to global acoustic properties of vocal sound production in explaining the attributes of speech on which talker identification is based, such an appeal was frustrated by the anomalous quality of sine-wave utterances. Lacking fundamental frequency, harmonic spectra, and broadband resonances, sine-wave signals are described as unnatural by listeners (Remez *et al.*, 1981) and are considered to be unnatural in quality even when they are intelligible (Remez *et al.*, 2002). The present findings from direct assessment of subjective similarity corroborate the estimates based on mistaken identity in studies by Fellowes *et al.* (1997) and by Remez *et al.* (1997), namely, that characteristic properties of a talker survive an acoustic transformation from natural sample to time-varying sinusoids. In the present project, Experiment 1 revealed that the perceived similarity among this set of talkers was much the same whether the judgments were based on natural samples or on sine-wave replicas. Experiments 2 and 3 together revealed that a combination of global and local properties is likely to contribute to the assessment of talker attributes when a listener encounters a sine-wave utterance.

The converging assessment of natural and sine-wave similarity performed in this study shows that a stable pattern of perceived similarity among sine-wave talkers is attributable to the robust properties of the talkers conveyed acoustically, and is not an artifact of the method employed by Remez *et al.* (1997) and Fellowes *et al.* (1997) to assess identification. These precedents directed a listener to choose which member of a pair of unidentified sine-wave samples had been spoken by a specific individual indicated by a printed name or exemplified in a brief natural sample. The prior findings encouraged the hypothesis that talkers are perceived from sine-wave patterns as if these synthetic utterances were speech samples with anomalous timbre—perhaps no more unusual than a sample of a familiar talker speaking with laryngitis—although the evidence to secure this conjecture was indirect. To test the hypothesis required perceptual measures with two critical controls.

The first control necessary to assess the claim that listeners perceived sine-wave talkers as if they were hearing the

natural models was a test of apparent similarity using both kinds of sample. In comparing the tests in these two acoustic conditions here, it was possible to determine that the pattern of performance manifest in each acoustic type resembled the pattern observed with the other. It appears from these tests that listeners do indeed express detailed sensitivity to properties of the speech of these individuals, and the pattern is highly correlated across the acoustic types.

The second control required to decide the claim was an experimental manipulation to test the perceptual role of a prominent global property, the frequency variation of the first formant in natural and in sine-wave samples. Our prior investigations had employed a statistical technique (cf. Walden *et al.*, 1978) in assessing the proximate causes of perceived similarity. This approach had proved fruitful in generating hypotheses about the phonetic properties available for indexical perception, chiefly because idiolectal attributes seemed far more distinctive and characteristic than did the global acoustic properties featured in the talker identification literature. The perceptual test was performed in Experiment 3 here. Although the result indicated a clear contribution of the tone analog of the first formant to a listener's perceptual appraisal of sine-wave talkers, the finding overall is consistent with the notion that indexical properties of talkers include linguistically governed properties like dialect and idiolect and rather more simple scalable properties like vocal tract size.

Limits to the generality of the findings. Although the tests reported here arguably expose the relative salience of phonetic details in perceptual appraisals of talker similarity, the findings show the availability of these attributes in the perception of individuals without defining the role that such properties play in ordinary listening. A test of direct magnitude estimation of the likeness of two speech samples in a controlled listening environment differs greatly from a challenge to identify a talker under ordinary circumstances. Indeed, the stability of recognition even in controlled tests can depend on the number and variety of other talkers in the test set (Bricker and Pruzansky, 1976; Hecker, 1971; Papcun *et al.*, 1989). New tests are required to determine whether the attributes of an individual's speech that are promoted to prominence by judgments of likeness have a legitimate function in perceptual identification. In this regard, it will also be necessary to calibrate the perceptual effect of aspects of speech that distinguish talkers by dialect and by idiolect (for instance, Remez *et al.*, 2004).

A second limit to consider derives from the kind of speech samples that were used in the present tests and in the immediate empirical precedents (Fellowes *et al.*, 1997; Remez *et al.*, 1997). All of the utterances were taken from sentence lists read aloud by the talkers. As such, the listeners in these tests heard speech produced in a normative register, which might have been less characteristic of the individual talkers than casually produced speech despite the instruction to speak comfortably. Additional tests will be required to determine the relation between the assortment of phonetic attributes in fluent reading and in spontaneous registers, indeed, among the differing formal and informal registers of spontaneous speech (for example, Labov, 1986).

This study took an experimental approach to the question of the perceptual effect of the tone analog of the first formant, and in comparison we must acknowledge that the opportunity to take an experimental approach to dialect and idiolect will be difficult. Certainly, the phonetic attributes at play among these ten talkers are not obscure: Some individuals spirantized the coronal stop releases while others released them with a clean, brief burst; some geminated an intervocalic stop hold (VCCV), while others more nearly approximated a VCV; some raised or diphthongized low vowels, while others more nearly produced a clear singleton. Our hypothesis is that, in aggregate, these phonetic aspects of an individual's speech contribute to individual identification when vocal quality is ordinary as well as anomalous. But, the problem of treating such properties as discriminanda in a factorial design requires identifying individual talkers whose natural dialect and idiolect vary in the manner required by a research method of parametric variation. New studies will determine the resolution with which empirical tactics can address this intriguing question about speech perception.

ACKNOWLEDGMENTS

The authors thank Philip E. Rubin, Michael Studdert-Kennedy, Jennifer Van Dyk, and Cynthia Y. Yang for their guidance and criticism of this project and report; and, we thank Floye Sumida, Robert Boruchowitz, and Harriet Greisser for advice and encouragement about the prospects of understanding this difficult and subtle problem. The research was sponsored by an award to Barnard College (DC00308) from the National Institute on Deafness and Other Communication Disorders.

- Abercrombie, D. (1967). *Elements of General Phonetics* (Aldine, Chicago).
- Belin, P., Fecteau, S., and Bédard, C. (2004). "Thinking the voice: Neural correlates of voice perception," *Trends Cogn. Sci.* **8**, 129–135.
- Bricker, P. D., and Pruzansky, S. (1976). "Speaker recognition," in *Contemporary Issues in Experimental Phonetics*, edited by N. J. Lass, (Academic, New York), pp. 295–326.
- Brungart, D. S., Iyer, N., and Simpson, B. D. (2006). "Monaural speech segregation using synthetic speech signals," *J. Acoust. Soc. Am.* **119**, 2327–2333.
- Fant, C. G. M. (1966). "A note on vocal tract size factors and nonuniform *F*-pattern scalings," *Speech Transmission Laboratory Quarterly Progress and Status Report 4*, Royal Institute of Technology, Stockholm, Sweden, pp. 22–30.
- Fellowes, J. M., Remez, R. E., and Rubin, P. E. (1997). "Perceiving the sex and identity of a talker without natural vocal timbre," *Percept. Psychophys.* **59**, 839–849.
- Frieda, E., Walley, A., Flege, J., and Sloane, M. (1999). "Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect," *Percept. Psychophys.* **61**, 561–577.
- Goldstein, U. G. (1980). "An articulatory model for the vocal tracts of growing children," Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Halle, M. (1985). "Speculations about the representation of words in memory," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. A. Fromkin (Academic, New York), pp. 101–114.
- Hawkins, S. (2003). "Roles and representations of systematic fine phonetic detail in speech understanding," *J. Phonetics* **31**, 373–405.
- Hecker, M. H. L. (1971). "Speaker recognition: An interpretive survey of the literature," *ASHA Monogr.* **16**, 1–103.
- Hollien, H. (2002). *Forensic Voice Identification* (Academic, San Diego).
- Johnson, S. C. (1967). "Hierarchical clustering schemes," *Psychometrika* **32**, 241–254.
- Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis and perception of

- voice quality variations among female and male talkers," *J. Acoust. Soc. Am.* **87**, 820–857.
- Krauss, R. M., Freyberg, R., and Morsella, E. (2002). "Inferring speakers' physical attributes from their voice," *J. Exp. Soc. Psychol.* **38**, 618–625.
- Kreiman, J. (1997). "Listening to voices: Theory and practice in voice perception research," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego), pp. 85–108.
- Kruskal, J. B. (1964). "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika* **29**, 1–27.
- Labov, W. (1986). "Sources of inherent variation in the speech process," in *Invariance and Variability in Speech Processes* edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 402–425.
- Ladefoged, P., and Broadbent, D. E. (1957). "Information conveyed by vowels," *J. Acoust. Soc. Am.* **29**, 98–104.
- Laver, J. (1980). *The Phonetic Description of Voice Quality* (Cambridge University Press, Cambridge, UK).
- Lieberman, P. (1963). "Some effects of semantic and grammatical context on the production and perception of speech," *Lang Speech* **6**, 172–187.
- Luce, P. A., Goldinger, S. D., Auer, E. T., Jr., and Vitevitch, M. S. (2000). "Phonetic priming, neighborhood activation, and PARSYN," *Percept. Psychophys.* **62**, 615–625.
- Neuner, F., and Schweinberger, S. R. (2000). "Neuropsychological impairments in the recognition of faces, voices and personal names," *Brain Cogn* **44**, 342–366.
- Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition* (Cambridge University Press, Cambridge, UK).
- Nygaard, L., Sommers, M., and Pisoni, D. (1994). "Speech perception as a talker-contingent process," *Psychol. Sci.* **5**, 42–46.
- Pardo, J. S., and Remez, R. E. (2006). "The perception of speech," in *Handbook of Psycholinguistics*, 2nd Edition, edited by M. Traxler and M. A. Gernsbacher (Academic Press, San Diego), pp. 201–248.
- Papcun, G., Kreiman, J., and Davis, A. (1989). "Long-term memory for unfamiliar voices," *J. Acoust. Soc. Am.* **85**, 913–925.
- Pisoni, D. B. (1997). "Some thoughts on 'normalization' in speech perception," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego), pp. 9–32.
- Pollack, I., Pickett, J. M., and Sumby, W. H. (1954). "On the identification of speakers by voice," *J. Acoust. Soc. Am.* **26**, 403–406.
- Remez, R. E., "Spoken expression of individual identity and the listener," in *Expressing Oneself/Expressing One's Self: A Festschrift in Honor of Robert M. Krauss*, edited by E. Morsella (Taylor & Francis, London), in press.
- Remez, R. E., Fellowes, J. M., and Rubin, P. E. (1997). "Talker identification based on phonetic information," *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 651–666.
- Remez, R. E., and Rubin, P. E. (1984). "On the perception of intonation from sinusoidal sentences," *Percept. Psychophys.* **35**, 429–440.
- Remez, R. E., and Rubin, P. E. (1993). "On the intonation of sinusoidal sentences: Contour and pitch height," *J. Acoust. Soc. Am.* **94**, 1983–1988.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). "Speech perception without traditional speech cues," *Science* **212**, 947–950.
- Remez, R. E., Wissig, S. C., Ferro, D. F., Liberman, K., and Landau, C. (2004). "A search for listener differences in the perception of talker identity," *J. Acoust. Soc. Am.* **116**, 2544.
- Remez, R. E., Yang, C. Y., Piorkowski, R. L., Wissig, S., Batchelder, A., and Nam, H. (2002). "The effect of variation in naturalness on phonetic perceptual identification," *J. Acoust. Soc. Am.* **111**, 2432.
- Rubin, P. E. (1980). "Sinewave synthesis," Technical report, Haskins Laboratories, New Haven, CT.
- Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., and Remez, R. E. (2002). "Learning to recognize talkers from natural, sinewave and reversed speech samples," *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 1447–1469.
- Smith, R. (2004). "The role of fine phonetic detail in word segmentation," Doctoral dissertation, University of Cambridge, Cambridge, UK.
- Stevens, A. A. (2004). "Dissociating the cortical basis of memory for voices, words and tones," *Cognit. Brain Res.* **18**, 162–171.
- Van Lancker, D., Cummings, J. L., Kreiman, J., and Dobkin, B. H. (1988). "Phonagnosia: A dissociation between familiar and unfamiliar voices," *Cortex* **24**, 195–209.
- Van Lancker, D., Kreiman, J., and Emmorey, K. (1985). "Familiar voice recognition: Patterns and parameters. I. Recognition of backward voices," *J. Phonetics* **13**, 19–38.
- von Dommelen, W. A. (1987). "The contribution of speech rhythm and pitch to speaker recognition," *Lang Speech* **30**, 325–338.
- Walden, B. E., Montgomery, A. A., Gibeily, G. J., Prosek, R. A., and Schwartz, D. M. (1978). "Correlates of psychological dimensions in talker similarity," *J. Speech Hear. Res.* **21**, 265–275.

The effect of word learning on the perception of non-native consonant sequences

Lisa Davidson,^{a)} Jason Shaw, and Tuuli Adams
New York University, New York, New York 10003

(Received 4 December 2006; revised 28 August 2007; accepted 1 October 2007)

Previous research in cross-language perception has shown that non-native listeners often assimilate both single phonemes and phonotactic sequences to native language categories. This study examined whether associating meaning with words containing non-native phonotactics assists listeners in distinguishing the non-native sequences from native ones. In the first experiment, American English listeners learned word-picture pairings including words that contained a phonological contrast between CC and CVC sequences, but which were not minimal pairs (e.g., [ftake], [fətalʌ]). In the second experiment, the word-picture pairings specifically consisted of minimal pairs (e.g., [ftake], [fətake]). Results showed that the ability to learn non-native CC was significantly improved when listeners learned minimal pairs as opposed to phonological contrast alone. Subsequent investigation of individual listeners revealed that there are both high and low performing participants, where the high performers were much more capable of learning the contrast between native and non-native words. Implications of these findings for second language lexical representations and loanword adaptation are discussed. © 2007 Acoustical Society of America.
[DOI: 10.1121/1.2801548]

PACS number(s): 43.71.Hw, 43.70.Mn, 43.71.Es [ARB]

Pages: 3697–3709

I. INTRODUCTION

Research in cross-language speech perception has shown that listeners faced with perceiving non-native sounds often have trouble distinguishing between a native phoneme and a non-native phoneme that is similar to it, or between two similar unfamiliar sounds that are both close to a native phoneme that is acoustically similar (e.g., Best, 1995; Cebrían, 2000; Flege *et al.*, 2003; Kuhl and Iverson (1995)). For example, a classic finding by Werker and Tees (1984) demonstrated that English-speaking adults have considerable difficulty in distinguishing between a dental /t/ and a retroflex /ʈ/ or the velar ejective /k'/ and the uvular ejective /q'/. Models of cross-language speech perception such as the Perceptual Assimilation Model (Best, 1995) and the Speech Learning Model (Flege, 1995) claim that the phonetic similarities between non-native segments and native ones, defined in terms of acoustic and articulatory characteristics, are predictive of whether or not listeners will be able to develop new perceptual categories for these sounds. Other studies demonstrate that the perception of a non-native segment is affected by prosodic position or surrounding segmental context (e.g., Strange *et al.*, 2001; 2004). For example, it is more difficult for Japanese speakers to distinguish between English /t/ and /l/ in prevocalic than in postvocalic contexts (Lively *et al.*, 1993).

In general, the focus of the cross-language speech perception literature has been on the perception of individual phonemes. A smaller but important set of studies have examined the perception of non-native phonotactics; that is, cases where the phonemes in question may be allowed by the na-

tive phonology of the participant, but not in the same sequential combinations allowed in the language being learned (or tested). As pointed out by Peperkamp (2007), there are a number of ways in which listeners may perceive non-native consonant sequences. Some studies have shown that listeners whose language contains at least one type of obstruent + approximant sequence are more likely to perceive an unattested obstruent+approximant sequence as an instance of a legal phonotactic pattern (Hallé *et al.*, 1998; Massaro and Cohen (1983); Moreton, 2002; Pitt, 1998). For example, Hallé *et al.* (1998) found that French listeners perceived natural tokens of /t/ and /d/ sequences as /k/ and /g/, respectively. Massaro and Cohen (1983) demonstrated that English speakers listening to a sound intermediate between /l/ and /r/ more often labeled the sound as /r/ when in the environment of /t_i/ and as /l/ in /s_i/.

Studies that have examined obstruent-obstruent or obstruent-nasal sequences have indicated a different type of perceptual response. A series of studies by Dupoux and colleagues (Dehaene-Lambertz, Dupoux, and Gout, 2000; Dupoux *et al.*, 1999, 2001) and follow-up studies by Kabak and Idsardi (2007) and Berent *et al.* (2007) using both identification and discrimination tasks indicate that another way in which listeners may process non-native CC sequences is to assimilate them to phonotactically permissible CVC sequences. For example, Japanese listeners frequently categorize items like [ebzo] as [ebuzo]. However, the likelihood of occurrence for “perceptual epenthesis” may be conditioned by the native language of the listener and the particular composition of the sequence. For example, whereas perceptual epenthesis was high for Japanese speakers in most obstruent-obstruent contexts, Davidson (2007) showed that English speakers transcribing non-native fricative-initial consonant

^{a)}Electronic mail: lisa.davidson@nyu.edu

sequences wrote a vowel between the consonants less than 15% of the time, which is much lower than the rate of perceptual epenthesis for Japanese listeners.

The fact that non-native listeners have difficulty distinguishing certain contrasts not found in their own language also has important ramifications for the nature of bilingual or second language lexical entries. Previous research on the phonetic characteristics of lexical representations of experienced second language learners has shown that the phonetic specifications of these representations are not as detailed as native language lexical entries (MacKay *et al.*, 2001; Sebastián-Gallés and Soto-Faraco, 1999). For example, Bradlow and Pisoni (1999) demonstrated that non-native listeners performing word recognition tasks are negatively affected when factors such as phonological neighborhood density are manipulated; non-native listeners have difficulty recognizing even familiar words that have many phonetically similar neighbors (see also Imai *et al.*, 2005).

In the present experiments, English-speaking adults were taught sets of word-meaning pairs that differ only with respect to whether or not they contained a phonotactically possible initial CVC (e.g., [zəɡamo]) or an impossible CC sequence (e.g., [zɡamo]). Unlike previous studies with adults that provided little incentive to distinguish between these minimally different words (e.g., Davidson, 2007; Dupoux *et al.*, 1999), it was hypothesized that associating pictures to unfamiliar and non-native words would actually assist the adults in attending to the critical acoustic detail differentiating these words. Two previous adult picture-word learning tasks using both native and non-native stimuli demonstrate that it is possible for adults to learn new words and even new non-native contrasts (Curtin *et al.*, 1998; Storkel *et al.*, 2006), but these studies did not explicitly test whether picture-word learning causes discrimination to improve over auditory-only tasks. A more relevant study by Hayes-Harb (2007) indicated that when adult learners are trained on a novel phonemic contrast (e.g., voiced [g] versus unaspirated [k] for English listeners), those participants who were taught minimal pairs in conjunction with pictures were more accurate in a following AX discrimination task than those who were trained using only statistical information.

Following the phonemic category results from Hayes-Harb (2007), it is expected that adults will still show phonological interference from native phonotactics, but that by matching words with meaning, they will have incentive to tease apart the fine phonetic differences between attested CVC and phonotactically illegal CC sequences. The stimuli used in this study and the comparison of word-learning with transcription and discrimination tasks are based on previous studies that also investigated the production and perception of words with fricative-initial CC and CVC sequences (Davidson, 2006, 2007). In the following section, these studies will be briefly reviewed to introduce the stimuli that are employed in the present experiments and to situate them within a larger picture of language acquisition, loanword adaptation, and cross-language speech perception.

Production and perception of fricative-initial sequences. The purpose of Davidson (2006) was to examine the nature of native English speakers' repairs of phonotactically unat-

tested initial CC sequences. Based on previous research demonstrating that speakers often insert vocalic material into non-native consonant clusters (Broselow and Finer, 1991; Davidson *et al.*, 2004), the study was designed to determine whether the vocalic material produced between the two consonants is (a) phonological vowel epenthesis, or (b) a result of insufficient articulatory overlap of the two consonant articulations, which is realized as a transitional schwa due to a period of open vocal tract between the consonants. English speakers were asked to repeat fricative-initial CC and matching CəC sequences produced by a Slovak speaker (e.g., [zɡamo] and [zəɡamo]). Accurate production of the initial cluster depended on whether the sequence began with [f], [z], or [v], but participants produced 45% of the tokens without vocalic material between the consonants and approximately 30% of the tokens with insertion of vocalic material (these tokens will be referred to in the text as C^əC). Acoustic measurements of the inserted schwas (C^əC) and lexical schwas (CəC) showed that inserted C^əC schwas were shorter and had a lower F1 midpoint, which is consistent with a brief transition between two constrictions but not with movement toward an open vowel target.

Given that English speakers produced C^əC tokens as a considerable proportion of their attempts to produce CC sequences, the output of Davidson (2006) is an interesting test case for the role of perception in loanword adaptation (e.g., Broselow, 2004; Kang, 2003; Peperkamp and Dupoux, 2003). That is, when presented with CC sequences, English speakers produce a mixture of CC and C^əC utterances. If these English speakers are considered as a group of initial borrowers ("disseminators"), then we can ask how a larger group of monolingual speakers ("recipients") learning these loanwords from the disseminators would interpret their productions of CC and C^əC. Davidson (2007) used the output of the speakers of the production study to investigate how recipients perceive phonotactically illegal CC sequences produced by native English speakers, and how they classify C^əC sequences. They may treat any vocalic material as a vowel, or they may recognize that the acoustic properties of [ə] are different from those of [ə̃], leading them to recover the disseminators' attempt to produce a target CC sequence.

Davidson (2007) used both a transcription task and an AX discrimination task to examine these questions. In the transcription task, participants were presented with CCVCV, CəCVCV, and C^əCVCV utterances produced by one of the English speakers from Davidson (2006). Participants wrote CC tokens as a cluster and CəC with a vowel each about 70% of the time, but C^əC tokens were written with a vowel only 54% of the time. These results suggest that recipients recognize that C^əC tokens tend to be different from tokens with a lexical vowel, but nevertheless may have a slight tendency to categorize the vocalic material as a vowel since the task encourages them to assume only two categories. In the AX discrimination task, listeners had trouble distinguishing between all of the possible pairings (i.e., CC/C^əC, CəC/C^əC, and even CC/CəC), but were significantly worse in the discrimination of CəC/C^əC as opposed to CC/C^əC.

While both results suggest that recipients treat C^oC as a variant of both C^əC and CC, the findings of Davidson (2007) indicate that speakers may even extend this confusion to contrasting C^əC and CC pairs if they are not presented in a task that forces them to provide the stimuli with a linguistic representation. In the present set of experiments using an experimental paradigm other than a discrimination task, this hypothesis is tested with a picture-word learning task. In the first experiment, participants are taught C^əC and CC-initial words that contain the same consonants but that are not minimal pairs (e.g., [zdati] and [zədanu]) in conjunction with pictures. This experiment tests whether the presence of phonological contrast alone, in conjunction with meaning, is enough to improve discrimination between C^əC and CC. In the second experiment, the C^əC and CC stimuli are taught as minimal pairs. In both experiments, only C^əC and CC are explicitly taught in the training phases, but C^oC tokens are added as a condition in the test phase in order to determine how listeners classify these sequences once they have established representations of the C^əC and CC words. The results of the current experiments will be discussed with respect to the AX discrimination task from Davidson (2007) to evaluate whether learning sound in conjunction with meaning improves discrimination. Note that with respect to the C^oC items, the purpose of these experiments is not to try to teach a three-way distinction between phonotactic categories, but rather to further study how recipients may treat disseminators' C^oC tokens in loanword adaptation.

II. EXPERIMENT 1

A. Method

1. Participants

The participants were 40 native American English speakers who were recruited from Craigslist and from flyers posted around New York University. No participant had any experience with Slavic, Hebrew, or any other languages that have the initial obstruent clusters used in the experiment. No participants were phonetically trained. None reported any history of speech or hearing impairments. Eight further participants began the experiment but did not complete the training phase (see Sec. II A 3).

2. Materials

Two sets of stimuli were used in Experiment 1; one for the training phase and one for the test phase. Both sets of stimuli were the same subset of the utterances produced by native English speakers who were participants in Davidson (2006), but there was a different speaker for the training and test phases in Experiment 1. In Davidson (2006), participants were asked to produce both words with fricative-initial sequences (i.e., [zgamo], [fnada]) and their counterparts with a schwa (i.e., [zəgamo], [fənada]) that were originally presented to them as spoken by a native Slovak speaker. Although Slovak does not allow word-initial C^əC sequences, the speaker was a phonetician and proficient English speaker, and was able to produce C^əC with a reduced schwa.

In Davidson (2006), the Slovak pseudowords were presented to the English speakers over computer speakers, ac-





	CC	C ^ə C				
Block 1	[vbano]	[vəbaki]		[vbano]		
	[vgane]	[vəgapo]				
	[fnagu]	[fənapi]				[fnagu]
	[zbagi]	[zəbasi]				
Block 2	[zmapi]	[zəmagu]		[fətalu]		
	[zdati]	[zədanu]				
	[vzaku]	[vəzaba]				[zəmagu]
	[ftake]	[fətalu]				

FIG. 1. (Color online) CC and C^əC stimuli (left) and example picture-word pairings (right) used in the training and test phase.

companied by a written representation of the word in English orthography on a computer screen. The participants then repeated each word aloud once. The participants' responses that were used for both the training and test stimuli were recorded at 22 kHz in a soundproof booth with a Marantz PMD-680 digital solid state recorder and an Audio-Technica AT813 microphone. All of the CC words were bisyllabic, with main stress on the first syllable (e.g., [zgámo]). The C^əC words were all trisyllabic, with stress on the second syllable (e.g., [zəgámo]). Participants produced the C^əC without any difficulty, but alternately produced the CC sequence accurately (no vocalic material between the consonants), or with a transitional schwa between the consonants (C^oC).

a. Training stimuli In the current experiment, the training stimuli consisted of eight CC-initial words and eight C^əC-initial words, each of which were paired with a unique cartoon character. For the training stimuli in the current study, if the speaker produced CC with any vocalic material between the consonant cluster, it was excised out in Praat by cutting at zero crossings to prevent acoustic artifacts in the signal. The list of stimuli and some example picture-word pairings are shown in Fig. 1.

In addition to the picture-word pairs, there were also minimal pair distracters in the form of CC-initial words that matched the words learned as C^əC, and C^əC-initial words that matched the CC words. For example, the word [zəmagu] was learned as the name of the dragon cartoon character, so [zmagu] was presented as a minimal pair distracter. The minimal pair distracters were produced by the same speaker as the target items. The minimal pair distracters were intended to teach the participants that words that differed with respect to the presence or absence of a schwa are not variations in production of the same lexical item. However, no picture was ever associated with the minimal pair distracter item.

b. Test stimuli The test stimuli were a set of words related to the learned items. The stimuli were divided into four conditions: Condition A, where the learned word was paired with a totally unfamiliar CC or C^əC word; Condition B, where the C^oC-initial word corresponding to the learned target word was paired with an unfamiliar word; Condition C, where the minimal pair distracter word was paired with an unfamiliar word; and Condition D, where two unfamiliar words were presented. The division of each stimulus item into experimental conditions is given in the Appendix. As explained further in Sec. II A 3, participants were played both words in conjunction with a picture and chose either one of the words or a "neither" option. Each of these test productions, including the C^oC words, were taken from the productions of a different English speaker from Davidson (2006) than the speaker used in the training. This speaker was chosen because all of his repetitions of the CC clusters

used in the current experiment were repaired by inserting a transitional schwa. Thus, the matching CC words were obtained by splicing out the schwa as described earlier. An acoustic analysis of the words demonstrated a substantial difference between the duration of the lexical [ə] and inserted [ə̃] (lexical: $M=50$ ms, $s.d.=15.5$ ms; inserted: $M=41$ ms, $s.d.=12.7$ ms) and the F1 value at the midpoint (lexical: $M=438$ Hz, $s.d.=67$ Hz; inserted: $M=390$ Hz, $s.d.=68$ Hz).

It should be noted that the stimuli shown in Fig. 1 represent a diverse set of consonant sequence types. The particular CC and CəC sequences were chosen for two main reasons. First, since most word-learning studies have small stimulus sets and it takes some time for participants to be adequately trained, we did not want to use extremely similar stimuli (e.g., all /f/-initial sequences) which may have made the task even more difficult. Second, the results of the transcription task in Davidson (2007) showed no effect of the first consonant type on transcribers' accuracy on /f/, /z/, and /v/-initial consonant sequences. While Davidson (2006) did show a significant effect of C1 type on production accuracy, this factor did not seem to carry over to how speakers perceive words with the same non-native sequences. Thus, we chose to use stimuli beginning with a variety of consonant sequence types.

c. Practice stimuli Finally, a set of stimuli were also recorded to use in a practice session intended to familiarize the participants with the task. Four Spanish words were chosen for the training phase: *botas*, *mapa*, *casa*, and *mano*. These words were recorded by a native speaker of English who had studied Spanish in high school and college. The test stimuli, recorded by a different native English speaker who had also studied Spanish in high school and college, consisted of the four training words plus four words that were similar except for a different first phoneme: *notas*, *papa*, *taza*, and *sano*.

3. Procedure

There were three phases in the experiment: A familiarization phase, a training phase, and a test phase. In order to simplify the learning task, participants learned eight word-picture pairings at a time. After the first familiarization-training-test sequence, the participants were given a break and returned to complete the second half. A total of 16 words were learned. The experiment was programmed using E-PRIME.

In order to ensure that participants could complete the task in no more than the hour they were expected to participate for, E-PRIME tracked how long the participants were taking. If they could not complete the first training phase within 45 min, they were thanked for their participation and told they had completed the requirement.

a. Familiarization phase Each of the eight picture-word pairs appeared on the screen twice in random order. While the cartoon character appeared on the screen, the sound file containing the name of the character played once. In the instructions for the familiarization phase, listeners were told that they would hear words spoken by an English speaker, some of which were more plausible English words than others, and that they should try to memorize each character-word pairing. Nevertheless, the familiarization

stage was intended to introduce the picture-word pairs, while learning the names actually occurred during the training phase.

b. Training phase Once the familiarization phase was completed, participants automatically moved on to the training phase. The eight cartoon pictures appeared on the screen in two rows, along with a black X on the top right side of the screen. Subjects then heard a word, and had to use the mouse to click on the picture that corresponded to that word from the familiarization phase. When a participant clicked on the correct picture, feedback was given by displaying a green checkmark visually on the screen. If a participant clicked on the incorrect picture, the feedback displayed a red "X," and the cartoon picture was shown again in conjunction with the correct sound file. When the minimal pair distracters were presented, participants were expected to click on the black X. For example, if the participant heard the word [fnagu], the correct response would be to click on the picture of the yellow star. However, if the participant heard [fə̃nagu], then the correct response was to click on the black X. If they did not choose the black X for the minimal pair distracters, they received the red X feedback.

In addition to the feedback, there was also an "escape hatch" provided to allow participants to review the picture-word pairs. By clicking on the word "Practice" at the bottom of the screen, participants were brought to a screen where they could click on a picture and hear the corresponding name. There was no limit to how long they could stay on the review screen.

The training session continued until all eight picture-word pairs were correctly matched in a row. After getting positive feedback, participants were told how many they had gotten right in a row. Choosing the wrong picture for a learned word or using the escape hatch reset the count, and participants had to start over again to get all 8 pairs in a row. To decrease the time spent in the training phase, incorrectly choosing a picture when presented with a minimal pair distracter did not reset the count, but negative feedback was given.

c. Test phase In this phase, a cartoon character appeared on the screen. Participants heard two words and were told to press the first key on a button box if the first word matched the picture, the second key if the second word matched, and the third key if neither word matched. The stimuli were divided up among the conditions, so that there were two CC items and two CəC items in each condition. Thus, after each of the familiarization and training phases, there are 8 test trials, for a total of 16 trials at the end of the experiment. Each condition was counterbalanced for stimulus order. A complete list of the conditions, learned words, stimuli pairs in the test phases, and the correct answers is in the Appendix.

In addition to the instructions regarding the button box, the participants were told that the English speaker for the test trials would be different from the one in the training phase. Though they had learned in training to reject the minimal pair distracters, they were also provided with the following information: "The choices may include exactly the same word that you learned for the character, may be completely different from any of the words that you learned, or may differ slightly from the word that you learned. If one option does not sound exactly like the word that you learned (e.g. a consonant or vowel sound is different), you should choose 'neither.' "

d. Practice The first familiarization/training/test sequence was preceded by a practice session using real Spanish words and the pictures they corresponded to. This was intended to introduce participants to the structure of the task. The familiarization phase consisted of showing each of the four picture-word pairs twice, followed by a training phase in which the participant heard one of the Spanish words they had just been familiarized with (e.g., *casa*), and had to click on the correct picture (e.g., a house) that is within an array of four pictures. Finally, in the test phase, a picture appeared followed immediately by two words (e.g., [kasa] [tasa]) and participants chose the appropriate button on the button box.

After the practice, the first familiarization/training/test sequence was presented. Participants were then given a short break and then continued to finish the second familiarization/training/test sequence.

In the following section, predictions for the experimental conditions are presented.

B. Predictions

1. *Condition A: Correct Name, Unfamiliar Word.* Participants should choose the correct word on most trials.

2. *Condition B: C^oC, Unfamiliar Word.* There are three possibilities:

- If listeners identify C^oC with CC, they should choose the C^oC utterance more often for words learned as CC than for words learned as C^oC.
- If listeners identify C^oC with C^oC, they should choose the C^oC utterance more often for words learned as C^oC than for words learned as CC.
- If listeners cannot distinguish C^oC from either CC or C^oC, then they should choose C^oC equally often regardless of the phonotactic type of the learned word.

3. *Condition C: Minimal Pair Distracter, Unfamiliar Word.* Since participants should have learned to distinguish between learned words and their minimal pair distracter counterparts in the training phase, they should be reasonably accurate regardless of the phonotactic type of the learned word. However, since minimal pairs are not explicitly presented in the test phase, it is likely that the participants could be led astray in this condition.

4. *Condition D: Two Unfamiliar Words.* Participants should choose “neither” on most trials.

C. Results

Here, we report on results for accuracy in the four conditions presented earlier. Although reaction time data were collected, it is not reported on because the only significant difference was for Condition D (Two Unfamiliar Words), which was significantly faster than all other conditions [$F(3,632)=5.65, p<0.001$].

1. Accuracy

The accuracy on each condition was assessed with a univariate analysis of variance (ANOVA). The independent variables were Condition (A, B, C, D) and Sequence (CC words vs. C^oC words). The dependent variable was the proportion of each subject’s accurate responses for each Condi-

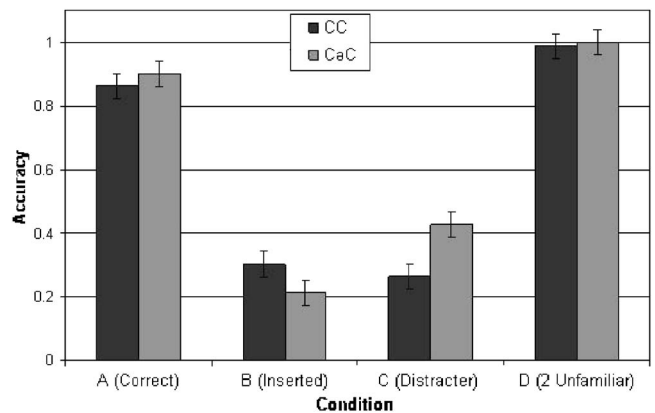


FIG. 2. Accuracy for each of the four conditions of auditory word pairs. The black bars represent the CC words that were learned, and the gray bars are the C^oC words. Error bars indicate standard error.

tion and Sequence, which was arcsin transformed. Results showed a significant main effect of Condition [$F(3,117)=105.57, p<0.001$], no main effect of Sequence [$F<1$], and a significant interaction between Sequence and Condition [$F(3,117)=2.74, p=0.047$]. Accuracy results are shown in Fig. 2. An examination of the incorrect responses shows that when participants chose the incorrect response in Conditions B and C (the correct response is “neither” for both), they always chose the inserted form and the minimal pair distracter, respectively.

Student Newman Keuls post-hoc tests indicated that collapsing over sequence type, Conditions A (Correct) and D (2 Unfamiliar Words) are not significantly different from each other, but both are significantly different from Conditions B (Inserted) and C (Minimal Pair Distracter) ($p<0.001$). Conditions B and C are also significantly different from one another ($p<0.05$). The significant interaction of Condition and Sequence is a result of a near-significant difference between CC and C^oC words in Condition C (Minimal Pair Distracter) [$F(1,78)=2.98, p=0.088$].

These results show that listeners were very consistent in choosing the correct word for Condition A and “neither” for Condition D, as predicted. They were significantly better at rejecting the minimal pair distracters in Condition C than the inserted words in Condition B, but only by a small amount. In Condition C, participants have a tendency to choose the minimal pair distracter more often for words learned as CC than words learned as C^oC.

2. High vs. Low performers

After the data were collected, we decided to examine the individual participants according to their performance on Condition C (Minimal Pair Distracter) to investigate whether all participants could learn and generalize the distinction between phonologically permitted (e.g., [zomagu]) and unattested (e.g. [zmagu]) nonwords to another speaker. There are many indications in the literature suggesting that both first and second language learners have differing phonological abilities that may be attributable to factors such as phonological awareness (deJong *et al.*, 2000; Hu and Schuele, 2005; Speciale *et al.*, 2004) or phonological short term

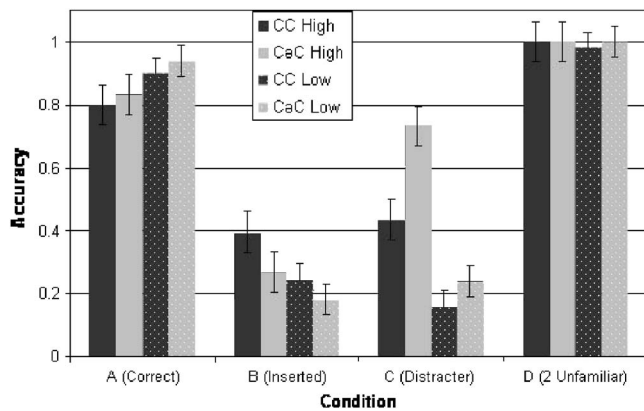


FIG. 3. Accuracy for each of the four conditions of auditory word pairs comparing High and Low performers (solid and dotted bars, respectively). The black bars represent the CC words that were learned, and the gray bars are the C ∂ C words.

memory (Gathercole and Pickering, 1999). Similarly, Wong and Perrachione (2007) showed a distinction between highly successful and less successful English learners of Mandarin tone in a word-learning task. Because of the *a posteriori* nature of this investigation, we have not collected relevant phonological awareness or short term memory data that might allow us to determine why some participants are better at learning the CC/C ∂ C distinction than others. Nevertheless, it is instructive to examine the performance of the participants to determine whether different types of learners are evident. The presence of such a division in these participants would provide motivation for a future study examining what factors might affect second language learners' perceptual discrimination abilities.

All participants who moved on to the test phase passed the training phase, indicating that they could learn the difference between one speaker's productions of target words and minimal pair distracters. However, this does not necessarily mean that mastery of these minimal pairs carried over to the second speaker's productions in the test phase. In dividing the participants into two groups, chance performance in Condition C (Minimal Pair Distracter) was used as the criteria. Since there were three choices for each test trial (word 1, word 2, and "neither"), chance was taken to be 33%. Those participants with average accuracy on Condition C below 33% were considered the Low performers ($N=23$), and those with accuracy above 33% were the High performers ($N=17$).

A univariate ANOVA was performed with the independent variables of Condition, Sequence, and Performance (High performers vs. Low performers). The dependent variable was accuracy. Results show that there is a significant effect of Performance [$F(1,304)=14.29, p<0.001$] and Condition [$F(3,304)=108.15, p<0.001$], but no effect of Sequence [$F<1$]. The two-way interactions between Performance and Condition [$F(3,304)=10.85, p<0.001$] and Sequence and Condition [$F(1,304)=3.62, p=0.014$] were significant. The three-way interaction was not significant. The results comparing High and Low performers are shown in Fig. 3.

Separate ANOVAs for each condition were conducted to examine the main effects and interaction of Sequence and Performance. The results for Condition A (Correct) show a significant main effect only for Performance [$F(1,76)=4.84, p=0.037$]. The Low performers are significantly more accurate on Condition A than the High performers (the latter respond "neither" more often), but average performance on Condition A for all participants is 80% or greater. For Conditions B (Inserted) and D (2 Unfamiliar), there are no significant effects of Sequence or Performance, nor is there an interaction. For Condition C (Minimal Pair Distracter), there is a significant effect of both Sequence [$F(1,76)=4.95, p=0.029$] and Performance [$F(1,76)=24.40, p<0.001$], but no significant interaction [$F(1,76)=1.57, p=0.214$]. For both High and Low performers, participants were significantly more accurate on words learned as C ∂ C than on words learned as CC. Furthermore, collapsing across sequence, Student Newman Keuls post-hoc tests showed that high performers were significantly more accurate on Condition C than on Condition B ($p=0.002$), whereas Low performers were not ($p=0.99$).

D. Discussion

The results of Experiment 1 conform to the predictions expressed in Sec. II B. In Condition A, in which the correct word was paired with an unfamiliar word, all participants chose the correct word with greater than 80% accuracy. Likewise, for Condition D, which paired two unfamiliar words with the learned cartoon character, participants chose "neither" almost 100% of the time. The unexpected finding that Low performers were significantly better than High performers on Condition A may be explained by their being less capable of distinguishing between the correct target word and the minimal pair distracter word that they had learned in the training. That is, High performers may have missed the correct answer because they were more aware of the possibility that what they were hearing was the minimal pair distracter, whereas Low performers were more often correct because they would have accepted a large range of variants for the target word.

In Condition C, participants were presented with the minimal pair distracter for the word that they had learned. Thus, when the star appeared on the screen, participants heard the word [f ∂ nagu] and an unrelated foil, and the correct response was to choose "neither." This distinction had been learned in the training phase, since participants were required to choose a black X for the minimal pair distracter in order to receive positive feedback. When divided into High and Low performers, results for this condition demonstrate that roughly half of the participants could generalize to a new voice with greater than chance accuracy, whereas the other half could not. Still, both High and Low performers were more accurate on choosing "neither" when they were presented with a CC word for a word that they had learned with a schwa as compared to words that had been learned with a consonant cluster.

Finally, in Condition B, in which participants were presented with the C ∂ C "version" of the word they had learned,

both groups were equally likely to choose the C^oC word instead of the “neither” option. High performers were numerically more accurate (more likely to reject C^oC) for words learned as CC. Although it is not significant, this result, which is suggestive of Prediction 2b for the High performers, is consistent with the hypothesis that they associate C^oC words with C^əC and are therefore likely to be more accurate on words learned as CC in Condition B. This point will be revisited in Sec. IV in conjunction with results from Experiment 2.

It is also relevant that there was no significant difference in accuracy between Conditions B and C for the Low performers. This indicates that they are unable to generalize the CC/C^əC distinction that they ostensibly learned in the training phase to a new speaker. The overall low accuracy in both conditions suggests that these participants are willing to accept acoustic variants of learned words regardless of the presence or absence of a vowel, possibly because they cannot overcome the phonotactic restrictions on the CC sequences, leading them to equate CC and C^əC.

In this experiment, participants were explicitly introduced to phonological contrast by having to learn CC and C^əC words which had the same initial consonants but which differed in the remainder of the word. They were exposed to minimal pairs in that they were required to reject those acoustic forms in the training session, but they did not ever learn minimal pairs where each word picked out a unique object. In the next experiment, we examine whether learning minimal pairs, where each of the lexical forms is associated with a meaning, enhances the participants’ ability to discriminate between CC and C^əC forms. We also ask whether learning minimal pairs clarifies how the participants treat the C^oC stimuli.

III. EXPERIMENT 2

A. Methods

1. Participants

The participants were 41 native American English speakers who were recruited from Craigslist and from flyers posted around New York University. None of these participants had taken part in Experiment 1. No participant had any experience with Slavic languages, Hebrew, or any other language that has the initial obstruent clusters used in the experiment. No participants were phonetically trained, and none reported any history of speech or hearing impairments. Eleven more participants were tested but their data were not used; nine failed to complete the training phase, and two failed to respond to more than 15% of the test items.

2. Materials

The materials for Experiment 2 are very similar to those for Experiment 1; however, since the purpose of this experiment is to have the participants learn minimal pairs, the words are CC-C^əC pairs. The stimuli are shown in Fig. 4.

The training and test stimuli were taken from the same two speakers as in the training and test conditions of Experiment 1. The stimuli for the training phase contained only the minimal pairs shown in Fig. 4, and stimuli for the test phase



	CC	C ^ə C		
Block 1	[vbano]	[vəbano]		[vbano]
	[zbasi]	[zəbasi]		
	[zdati]	[zədati]		
	[fnapi]	[fənapi]		
Block 2	[vgane]	[vəgane]		[zmagu]
	[zmagu]	[zəmagu]		
	[vzaku]	[vəzaku]		
	[ftake]	[fətake]		

FIG. 4. (Color online) CC and C^əC minimal pairs (left) and example picture-word pairings (right) used in the training and test phase.

also included the corresponding C^oC words. The same practice stimuli used in Experiment 1 were also used in this experiment.

3. Procedure

As in Experiment 1, there were three phases in the experiment: A familiarization phase, a training phase, and a test phase. Participants again learned 8 word-picture pairings at a time (4 CC/C^əC pairs) for a total of 16 words learned. After the first familiarization-training-test sequence, the participants were given a break and returned to complete the second half.

The familiarization phase was the same as for Experiment 1. The training phase differed in that there was only a 4 × 2 array of cartoon characters but no black X, since there were no minimal pair distracters in this experiment. Participants had to accurately match the CC stimulus (e.g., /zəmagu/) with its picture (the fish) and the C^əC stimulus (e.g., /zəmagu/) with its picture (the dragon). Once all 8 pictures were correctly matched in a row, participants moved into the test phase. Again, the count of correctly matched picture-word pairs was reset when a participant made an incorrect choice, and the escape hatch was available for the participant to review the cartoon-name pairings.

In the test phase, all learned pictures were presented in three different conditions. This is different from Experiment 1, where each picture was only presented once with an auditory pair of options. After completing Experiment 1, it was determined that the test conditions should be changed for Experiment 2 to make the task more statistically robust (more observations per participant) and more comparable to previous AX discrimination tasks which did not include a word-learning component (Davidson, 2007). The conditions are shown in Table I. In Condition A trials, participants heard the correct word that they had learned and the distracter word, or the minimal pair which was the name of another cartoon character. In Condition B, participants heard the correct word and the matching word which contained the in-

TABLE I. Test conditions and examples of auditory stimuli Experiment 2. See Fig. 4 for the pictures represented here by text descriptions.

Learned word	Condition	Picture	Auditory stimuli
C ^o C: [zəmagu]	A: Correct vs. Distracter	Dragon	[zəmagu] [zmagu]
	B: Correct vs. Inserted		[zəmagu] [z ^ə magu]
	C: Distracter vs. Inserted		[zmagu] [z ^ə magu]
CC: [zmagu]	A: Correct vs. Distracter	Fish	[zmagu] [zəmagu]
	B: Correct vs. Inserted		[zmagu] [z ^ə magu]
	C: Distracter vs. Inserted		[zəmagu] [z ^ə magu]

served schwa. Condition C, consisting of the distracter minimal pair and the inserted word, was a mismatch condition which did not contain the correct name at all.

Participants were instructed to press the first button on the button box if the correct answer was the first response and the second button on the button box if the correct answer was the second response. There was no “neither” option in this experiment. Each subject heard all conditions in the test phase; the stimuli were presented in random order and the experimental halves were counterbalanced. There were a total of 48 test trials, with 24 in each half. The practice phase for Experiment 2 was the same as in Experiment 1.

B. Predictions

In this experiment, speakers must learn that minimal pairs of CC/CəC items cannot be acoustic variants of one another because they denote different pictures. The conditions given in Table I lead to several hypotheses. The participants in this experiment are also divided into High and Low performers on the basis of Condition A (Correct vs. Distracter), which is the counterpart of Condition C (Minimal Pair Distracter) in Experiment 1.

If learning minimal pairs where each word is associated with meaning does not help Low performers discriminate between the lexical items, then it is assumed that they will essentially perform at chance for all conditions. However, if learning minimal pairs improves their performance, then all participants should conform to the following predictions (based on the results of Experiment 1).

1. *Condition A: Correct vs. Distracter.* Accuracy should be greater than chance. It is also possible that accuracy on Condition A for words learned as CəC is significantly greater than for words learned as CC [cf. Condition C (Minimal Pair Distracter) in Fig. 3].

2. *Condition B: Correct vs. Inserted.* It is not expected that participants will be very accurate in this condition. The results of Experiment 1 suggest the following two as the most likely outcomes (i.e., it is not likely that listeners will identify C°C only with CC):

- a. If listeners identify C°C with CəC, they will choose Correct more often for pictures learned as CC. They will be at chance for pictures learned as CəC because they will be confused by the presentation of C°C and CC on the same trial.
- b. If C°C is treated as a variant of both types of words, they should be at chance regardless of whether the target word was CC or CəC.

3. *Condition C: Distracter vs. Inserted.* This is a “mismatch” condition that was intended to provide further insight both into how participants categorize C°C utterances and how robust the representation of the CC and CəC words that they learned is.

- a. If listeners identify C°C with CəC, they will choose Inserted more often than chance for pictures learned as CəC. They will be at chance for pictures learned as CC.
- b. If C°C is treated as a variant of both types of words, and participants recognize that the Distracter word is not the right

match for the picture, then the C°C answer should be chosen more often regardless of whether the target word was CC or CəC.

C. Results

Following Experiment 1, the first step was to divide the participants into High and Low performing groups on the basis of Condition A (Correct vs. Distracter), which is the counterpart of Condition C (Minimal Pair Distracter) in Experiment 1. Again, it was expected that since the participants could not leave the training without accurately matching the CC and CəC minimal pairs with the correct picture, they should do well on Condition A. In this experiment, chance was taken to be 50% since there were only two choices for each test trial. There were 19 participants in the Low performing group and 22 participants in the High performing group.

Because there is no “correct” response in Condition C, it is impossible to directly compare Conditions A and B to Condition C on the basis of accuracy. Thus, an omnibus ANOVA with condition as an independent variable cannot be carried out. Instead, each condition is examined in a separate ANOVA to examine the effects of sequence and performance (High vs. Low). For Conditions A and B, the dependent variable is accuracy. For Condition C, the response type reflects how often the participants chose the “Inserted” (CəC) word instead of the distracter word in the minimal pair. Proportions of responses for Experiment 2 are shown in Table II.

In Condition A (Correct vs. Distracter), there was a significant main effect of Performance [$F(1,78)=81.78, p < 0.001$] but no significant effect of Sequence [$F < 1$]. There was no significant interaction between Sequence and Performance [$F < 1$]. Regardless of whether the learned target word was CC or CəC, the High performing participants chose the correct word more often than the minimal pair word.

In Condition B (Correct vs. Inserted), there was no main effect of either Sequence [$F < 1$] or Performance [$F(1,78)=2.60, p=0.11$]. The interaction between Sequence and Performance is marginally significant [$F(1,78)=3.43, p=0.07$]. The interaction results are due to slightly more accurate performance by High performers for words learned as CC as compared to words learned as CəC, but it is not significant ($p=0.27$).

In Condition C (Distracter vs. Inserted), there was a significant main effect of Sequence [$F(1,78)=21.81, p < 0.001$], but no main effect of Performance [$F < 1$]. The interaction between Sequence and Performance was significant [$F(1,78)=4.96, p=0.029$]. The interaction is due to High performers choosing the C°C stimulus significantly more often for words learned as CəC than for words learned as CC, and more often than Low learners ever choose C°C ($p < 0.001$).

D. Discussion

The results for Experiment 2 indicate that there are differences in performance for High performers vs. Low performers. Collapsing over Sequence, the Low performers are

TABLE II. Response types for each of the three conditions of auditory word pairs comparing High and Low performers. Bold numbers represent the accurate response for Conditions A and B. Although there is no “correct” response for Condition C, C°C is the more expected response. The dash (—) indicates that the response type was not available for that condition.

		Response type for High and Low learners					
		CC		C∅C		C°C	
		High	Low	High	Low	High	Low
Condition A (Correct vs. Distracter)	CC	0.71	0.46	0.29	0.54	—	—
	C∅C	0.36	0.52	0.74	0.48	—	—
Condition B (Correct vs. Inserted)	CC	0.64	0.52	—	—	0.36	0.48
	C∅C	—	—	0.59	0.59	0.41	0.41
Condition C (Distracter vs. Inserted)	CC	—	—	0.53	0.47	0.47	0.53
	C∅C	0.27	0.37	—	—	0.73	0.63

at chance for all conditions except Condition C (Distracter vs. Inserted) (comparison to chance using the binomial distribution and an alpha of 0.05, the probability for Condition A=0.86, Condition B=0.34, Condition C=0.003). That is, although they were able to learn the distinctions between the minimal pairs and associate them with the correct cartoon characters during the training phase, they could not easily generalize that knowledge to the parallel distinction produced by a different speaker. In Condition C, the mismatch condition, the Low performers pattern like the High performers, preferring to choose the C°C token for words learned as C∅C.

With an accuracy of greater than 70% in Condition A (Correct vs. Distracter), the High performers are fairly successful in generalizing the distinction between CC and C∅C that they learned in the training phase. The lack of any difference between words learned as CC and words learned as C∅C indicates that High performers’ ability to learn and encode both native and minimally different non-native phonotactics is enhanced by learning actual minimal pairs. This finding is in contrast to Experiment 1, where performance on Condition C, in which participants were expected to choose “neither” when presented with a minimal pair distracter (which had not been learned in conjunction with a picture), was better for words learned as C∅C than for words learned as CC.

The results for Condition B (Correct vs. Inserted) indicate that the High performing participants had some difficulty distinguishing between the words they had learned and the C°C variant for both types of sequences. However, there was a small tendency for the High performing participants to choose the correct word more often for words that had been learned as CC. This is consistent with Hypothesis 2a in Sec. III B, which stated that if participants associated C°C with C∅C, then they would be more confused by trials in which these words were paired together.

This trend is better confirmed in Condition C (Distracter vs. Inserted), in which High performers chose the C°C option much more often for words learned as C∅C than for words learned as CC. It was hypothesized that if participants rejected the distracter word equally for both CC and C∅C words, then they should choose the C°C option at equal (and

perhaps above chance) rates. However, the results demonstrate that the participants equate C°C with C∅C much more often, and they are at chance for the words learned as CC. In the latter case, it is possible that neither utterance is an acceptable version of CC for the participants, and therefore they randomly choose between the two substandard options.

IV. GENERAL DISCUSSION

These experiments were designed to address three main questions: (1) Does learning sound-meaning pairs assist learners in perceiving and distinguishing the fine-grained acoustic differences between phonotactically native C∅C and non-native CC sequences, (2) do minimal pairs further enhance the ability to distinguish these sequences, and (3) how does the picture-matching task affect the interpretation of a speaker’s repair of non-native phonotactics?

A. Learning to distinguish CC from C∅C

With respect to the first and second questions, the experimental results indicate that it is primarily when listeners are explicitly taught sound-meaning relationships with minimal pairs that they improve their ability to distinguish between CC and C∅C tokens across the board. Thus, this study extends Hayes-Harb’s (2007) finding regarding the usefulness of minimal pairs in learning to distinguish non-native phonemes to phonotactic sequences. The results from Experiment 1, which required participants to reject the minimal pair distracter words in the training phase, show that Low performers are unable to generalize the distinction between CC and C∅C to a new speaker’s voice. They only correctly choose “neither” about 20% of the time in Condition C in the test phase, where the minimal pair distracter is paired with a totally unrelated word. High performers are somewhat more accurate, but show a large difference between words learned as CC and those learned as C∅C. While they are able to reject the CC distracter for words learned as C∅C approximately 73% of the time, they only reject the C∅C distracter about 40% of the time.

This result is the first indication that the phonetic encoding or representation formed for phonotactically legal words learned as C∅C is more robust and leads to greater accuracy

when generalizing to another voice. For words learned as CC, participants seem to accept a greater variety of productions, suggesting that the native language phonological prohibition on the CC sequences used in this study hampers a detailed phonetic encoding of these items, at least at an early stage of exposure. This experiment indicates that simply introducing participants to the phonological contrast between CC and CəC does not sufficiently underscore the lexical relevance of the ə ~ ∅ alternation. If listeners do not need to prevent lexical confusion because there are no minimal pairs, then perhaps they are willing to treat CC and CəC as variants of the same word.

In Experiment 2, which explicitly relies on minimal pairs, participants show an overall improvement in accuracy, though the results are more evident for the High performers. In Condition A, which directly pits CC against CəC words that were learned in association with a picture, the accuracy of High performers is the same for both CC and CəC words and is similar to their performance for words learned as CəC in Condition C (Minimal Pair Distracter) in Experiment 1. In order to further evaluate whether the minimal pairs contributed to better discrimination than phonological contrast alone, it is appropriate to examine whether an accuracy of 71% in Condition A of Experiment 2 where chance is 50% and 43% when chance performance is 33% (Condition C of Experiment 1) are both above chance. Using the binomial distribution to calculate the cumulative probability of the participants' performance in both studies, the probability of obtaining an average accuracy of 43% or greater given a chance level of 33% in Experiment 1 is 0.16. In Experiment 2, the probability of scoring 71% or greater when chance is 50% is less than 1.17×10^{-8} . Thus, while the results for Experiment 1 were not above chance, the probability of scoring 71% in Experiment 2 is much less.

These findings suggest that when minimal pairs are matched with meaning, participants have greater incentive to attend to the subtle differences between the words, even when spoken by a different talker. These results suggest that effectiveness of minimal pairs in adult word learning tasks is different than for infants, who have more difficulty discriminating unfamiliar minimal pairs when paired with meaning than in purely auditory tasks (Pater *et al.*, 2004; Stager and Werker, 1997). The difference between adult and infant performance is consistent with Werker and colleagues' interpretation that greater task demands prevent infants from integrating the acoustic information with their newly formed lexical representation; adults are not assumed to have infant-like limitations on their processing resources, so the picture matching task helps focus attention on the phonetic differences rather than obscure them. Likewise, minimal pair comparisons in older children with phonological delay have been used successfully to help children learn to correctly produce new sounds, especially when learning two new sounds which were previously both erroneously produced [see Gierut (1998) for an overview of this issue].

Two limitations of the current study should be taken into consideration when interpreting the results. First, there is no baseline condition showing how well the participants would have been able to identify CC or CəC tokens without the

benefit of lexical training with pictures. However, the AX discrimination task from Davidson (2007) provides information about English speakers' performance in the absence of training. In the AX experiment, English speaking participants were presented with the same types of CCVCV and CəCVCV pairs produced by the same English speaker as in the test phase of the current experiments (there were also CC/CəC and CəC/CəC conditions). To encourage a response based on an acoustic comparison of the contrastive sequences, the discrimination pairs were presented with a small ISI of 250 ms and the participants were asked to determine whether the sound files were "exactly the same" or not. Results for the CCVCV/CəCVCV comparison showed that participants labeled them as "different" only 49% of the time; thus, when asked to determine whether the CəC and CC sound files were exactly the same, participants were at chance (chance level=50%).

The second limitation is that the experimental conditions for Experiment 2 are not exactly the same as those for Experiment 1. This makes it difficult to directly compare the results of the two studies, although the probabilities of the outcomes determined using the binomial distribution assist in interpreting the results. This limitation is particularly evident when comparing Experiment 1 with the AX discrimination task, because there was no condition explicitly pitting CC against CəC. Still, it is useful to evaluate the performance on Condition C (Minimal Pair Distracter) in light of the AX discrimination task. While the Low performers in the current study correctly rejected the minimal pair distracter for both CC and CəC stimuli only approximately 20% of the time, the High performers accurately rejected the minimal pair distracter 43% of the time for CC stimuli and 73% of the time for CəC stimuli. In the absence of direct comparison between the two types of tokens (as in the AX task), the Low performers in this study simply accept the minimal pair distracter as a variant for the word they learned. The High performers, however, already show some benefit from the lexical training for CəC as compared to the AX performance results, but not for the phonotactically unattested CC.

Condition A (Correct vs. Distracter) from Experiment 2 is more directly comparable to the AX discrimination task, since participants were presented with the minimal pair in the trial. High performers were significantly better than chance for both CC and CəC stimuli, and these were not different from one another. Low performers did not improve over the AX discrimination results, despite the addition of the minimal pair training. Although not reported in Davidson (2007), a post-hoc examination of individual subject performance for the AX discrimination task showed that only 8 out of 31 participants performed above chance, so dividing the group into High and Low performers would not have been motivated.

Comparing Experiments 1 and 2 from this study to the AX discrimination task confirms that while forming sound-meaning pairs might improve accuracy in discrimination for the words that already have a more detailed phonetic representation (i.e., CəC words), it is only when lexical contrast

is added that participants are really required to attend to the presence versus absence of the vowel. This result is consistent with the general idea outlined by Hawkins (2003) and Coleman (2003) that subtle but systematic phonetic differences play an important role in phonemic (and by extension, phonotactic) differentiation, and that meaning and experience affect the interpretation of fine phonetic detail. While the idea that minimal pairs are crucial for the formation of phonemic and phonotactic inventories is hardly a new idea in phonology, this study does provide new insight into the amount of phonetic detail that speakers might be willing to overlook—such as a reduced vowel—in the absence of a minimal pair.

Though the current study suggests that there are conditions under which participants can acquire the CC/CəC contrast, studies of the lexical representations of highly fluent bilinguals suggest that whether or not contrasts can be learned and robustly encoded in the lexicon may depend on the specific phonemes or phonotactic sequences in question (e.g., Curtin *et al.*, 1998; Weber and Cutler, 2004). For example, Pallier *et al.* (2001) showed that Spanish-dominant Spanish/Catalan bilinguals showed repetition priming for Catalan minimal pairs differing only in [ɛ]/[e], [o]/[ɔ], or [s]/[z], whereas Catalan-dominant bilinguals did not (e.g., [nɛta]/[neta], “clean/granddaughter”). They concluded that a lack of sensitivity to these contrasts for Spanish-dominant bilinguals leads to representation of these minimal pairs as homophones.

The current study, taken together with previous findings, suggests that the phonetic content of some kinds of phonological elements may be either more or less salient to the learner than other types are. Whereas the distinction between vowels may be particularly difficult to represent lexically due to the perceptual organization of the vowel space (Best, 1995; Flege and MacKay, 2004; Kuhl and Iverson, 1995), it may be easier for learners to ultimately learn new phonotactic contrasts, especially where the consonants in question are in the inventory of the listeners but are not allowed in CC sequences in word-initial position. Furthermore, the phonetic detail in the representation of non-native items may not be as robustly encoded as English-possible items (cf. Gathercole *et al.*, 1999 for phonotactic effects in short-term recall by children). As a result, even the High performing listeners accept CəC as a variant of CC more often than they accept CC for CəC.

The division between participants who can distinguish between CC and CəC after training and those who cannot is another issue raised by this study. The Low performers in this study never manage to perform better than chance even in Experiment 2, and it is not clear from this study whether they would continue to have difficulty in distinguishing the CC/CəC contrast, or whether they would be able to carry out the task at the same level as the High performers if they were given more practice. The results of these experiments suggest that not all adult language learners are necessarily equipped with the same ability to perceive and distinguish

between non-native sounds, at least at the earliest stages of acquisition (cf. Speciale *et al.*, 2004). Why this is the case, and whether all participants would eventually show equal ability on these tasks over time, is an interesting question for future research.

B. The interpretation of fine phonetic detail

The third question addressed by this study pertained to the effect that associating meaning with the learned words would have on the interpretation of C²C stimuli. Although results of the AX discrimination task for comparisons containing the C²C tokens were universally poor in Davidson (2007), there was nevertheless a significant difference between performance on the C²C/CəC discrimination (20% “different” responses) versus the C²C/CC discrimination (34% “different” responses). Similarly, in the transcription task, participants wrote a vowel between the consonants for over half of the C²C tokens. These findings already suggested a bias for perceiving the inserted schwa as a vowel, despite the fact that several acoustic properties of the inserted schwa, including duration and F1 midpoint, are significantly different from a lexical vowel.

The results of both Experiments 1 and 2 confirm that while confusability with both CC and CəC is generally high, listeners are more likely to interpret C²C as containing a vowel. In Experiment 1, High performers in Condition B, in which C²C was presented with an unrelated word, were more likely to choose “neither” for words that were learned as CC than for words learned as CəC. This indicates that they were more likely to equate C²C with the lexical vowel, and thus were less capable of rejecting the C²C foil for words learned as CəC. Similarly, in Condition C of Experiment 2, which paired the distracter word with the C²C word, High performers were more likely to choose C²C over CC for words that had been learned as CəC, as compared to words that had been learned as CC. Low performers showed a similar tendency.

Although there are inevitable differences between this type of laboratory study and real language contact, the results of these experiments can help inform predictions about loanword adaptation. If we assume that these stimuli represent the actual types of utterances that a recipient might hear from disseminators during the process of loanword adaptation, then the findings of this study, in conjunction with Davidson (2006), suggest that two phonotactic categories distinguishing between CC and CəC might eventually be established over time if English were to borrow large numbers of words with initial obstruent-obstruent sequences. That is, when the conditions are right—when learners must provide some kind of representation to a new word, and if they are fortunate enough to learn minimal pairs—they are better able to perceive and encode CC sequences. As noted in Sec. IA, it was not the purpose of these studies to attempt to teach listeners a three-way contrast between CC, CəC, and C²C. In a borrowing situation, C²C sequences would likely be divided up among the CC and CəC categories, with a larger proportion of them being assigned a CəC representation. Furthermore, results of Davidson (2007) demonstrate

that disseminators sometimes produce fricative-initial CC sequences accurately, so some proportion of these sequences would be both represented and produced accurately by borrowers. It is when these items are passed down to the next generation of children that they will ultimately lead to phonological change such that the relevant CC sequences become permitted by the phonology (Lahiri *et al.*, in press).

V. CONCLUSION

The process of learning new sounds and new contrasts, whether segmental or phonotactic, is affected by factors both internal and external to the phonetic string itself. It has been shown that certain phonemes or phonotactic sequences are more difficult than other to accurately perceive, but regardless of what kind of unit is being acquired, the acquisition of a new sound system or the interpretation of loanword adaptations does not simply require listeners to distinguish between two auditory strings. In language learning and contact situations, lexical items are acquired with their associated meanings. In this sense, a picture matching task may be a better laboratory analogy for loanword adaptation or second language learning than simple discrimination tasks, since learning meaning is an integral part of language contact. The results of these experiments suggest that the more incentive that adult listeners have to distinguish fine phonetic detail, the more they will attend to it and incorporate it into their lexical representations.

ACKNOWLEDGMENTS

The authors thank the three anonymous reviewers, and audiences at Yale University and the New York University Phonetics/Phonology Lab for their comments on this work. We would also like to thank David Goldberg for creating the cartoon characters used in the experiments. This work was supported by the National Science Foundation CAREER Grant No. BCS-0449560.

APPENDIX

Stimuli used in each experimental condition in Experiment 1. Block refers to whether the stimulus was presented in the first or second half of the experiment. See Fig. 1 for examples of the pictures represented here by text descriptions.

Condition	Picture	Learned word	Stimulus pair in test	Correct response
A (Correct)	Alien	vbano	vbano zgade	vbano
	Fly	zdati	zdati fāsaga	zdati
	Octopus	vəbaki	vəbaki fmasa	vəbaki
	Turtle	zədanu	zədanu vəgalu	zədanu

B (Inserted)	Robot	vgane	v ^ə gane fpami	neither
	Helmet	zmapi	z ^ə mapi vəmabu	neither
	Fish	vəgapo	v ^ə gapo fətano	neither
	Dragon	zəmagu	z ^ə magu vdapi	neither
C (Minimal Pair Distracter)	Star	fnagu	fənagu znaʃo	neither
	Skull	vzaku	vəzaku zədaba	neither
	Flower	fənapi	fnapi zbatu	neither
	Snake	vəzaba	vzaba zəgano	neither
D (2 Unfamiliar)	Ghost	zbagi	fkada vənali	neither
	Bird	ftake	zmafo vəzamo	neither
	Cat	zəbasi	vəmala fkabe	neither
	Flame	fətalū	zəvapa vdagu	neither

- Berent, I., Steriade, D., Lennertz, T., and Vahnin, V. (2007). "What we know about what we have never heard: Evidence from perceptual illusions," *Cognition* **104**, 591–630.
- Best, C. (1995). "A direct-realist view of cross-language perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* edited by W. Strange (York, Baltimore), pp. 171–204.
- Bradlow, A., and Pisoni, D. (1999). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," *J. Acoust. Soc. Am.* **106**, 2074–2085.
- Broselow, E. (2004). "Language contact phonology: Richness of the stimulus, poverty of the base," in *Proceedings of NELS 34* edited by K. Moulton and M. Wolf (Booksurge, North Charleston, SC), pp. 1–21.
- Broselow, E., and Finer, D. (1991). "Parameter setting in second language phonology and syntax," *Second Lang. Res.* **7**, 35–59.
- Cebrian, J. (2000). "Transferability and productivity of L1 rules in Catalan-English interlanguage," *Stud. Second Lang. Acquis.* **22**, 1–26.
- Coleman, J. (2003). "Discovering the acoustic correlates of phonological contrasts," *J. Phonetics* **31**, 351–372.
- Curtin, S., Goad, H., and Pater, J. (1998). "Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers," *Second Lang. Res.* **14**, 389–405.
- Davidson, L. (2006). "Phonology, phonetics, or frequency: Influences on the production of non-native sequences," *J. Phonetics* **34**, 104–137.
- Davidson, L. (2007). "The relationship between the perception of non-native phonotactics and loanword adaptation," *Phonology* **24**, 261–286.
- Davidson, L., Jusczyk, P., and Smolensky, P. (2004). "The initial and final states: Theoretical implications and experimental explorations of richness of the base," *Fixing Priorities: Constraints in Phonological Acquisition*, edited by R. Kager, W. Zonneveld, and J. Pater (Cambridge University Press, Cambridge), pp. 321–368.
- Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (2000). "Electrophysiological correlates of phonological processing: A cross-linguistic study," *J. Cogn. Neurosci.* **12**, 635–647.
- deJong, P. F., Seveke, M.-J., and van Veen, M. (2000). "Phonological sensitivity and the acquisition of new words in children," *J. Exp. Child Psychol.* **76**, 275–301.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). "Epenthetic vowels in Japanese: A perceptual illusion?," *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 1568–1578.
- Dupoux, E., Pallier, C., Kakehi, K., and Mehler, J. (2001). "New evidence for prelexical phonological processing in word recognition," *Lang. Cognit.*

- Processes **16**, 491–505.
- Flege, J. E. (1995). "Second-language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD) pp. 229–273.
- Flege, J. E., and MacKay, I. (2004). "Perceiving vowels in a second language," *Stud. Second Lang. Acquis.* **26**, 1–34.
- Flege, J. E., Schirru, C., and MacKay, I. (2003). "Interaction between the native and second language phonetic subsystems," *Speech Commun.* **40**, 467–491.
- Gathercole, S., and Pickering, S. J. (1999). "Estimating the capacity of phonological short-term memory," *Int. J. Psychol.* **34**, 378–382.
- Gathercole, S., Frankish, C., Pickering, S., and Peaker, S. (1999). "Phonotactic influences on short-term memory," *J. Exp. Psychol. Learn. Mem. Cogn.* **25**, 84–95.
- Gierut, J. (1998). "Treatment efficacy: Functional phonological disorders in children," *J. Speech Lang. Hear. Res.* **41**, 85–101.
- Hallé, P., Segui, J., Frauenfelder, U., and Meunier, C. (1998). "Processing of illegal consonant clusters: A case of perceptual assimilation?," *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 592–608.
- Hawkins, S. (2003). "Roles and representations of systematic fine phonetic detail in speech understanding," *J. Phonetics* **31**, 373–405.
- Hayes-Harb, R. (2007). "Lexical and statistical evidence in the acquisition of second language phonemes," *Second Lang. Res.* **23**, 61–90.
- Hu, C.-F., and Schuele, M. (2005). "Learning nonnative names: The effect of poor native phonological awareness," *Appl. Psycholinguist.* **26**, 343–362.
- Imai, S., Walley, A., and Flege, J. E. (2005). "Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners," *J. Acoust. Soc. Am.* **117**, 896–907.
- Kabak, B., and Idsardi, W. (2007). "Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints?," *Lang Speech* **50**, 23–52.
- Kang, Y. (2003). "Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean," *Phonology* **20**, 219–273.
- Kuhl, P., and Iverson, P. (1995). "Linguistic experience and the perceptual magnet effect," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, edited by W. Strange (York, Baltimore).
- Lahiri, A., Wetterlin, A., and Jönsson-Steiner, E. (in press). "Tones and loans in the history of Scandinavia," in *Phonetic and Behavioural Studies in Word and Sentence Prosody, Tones and Tunes*, Vol. II, edited by C. Gussenhoven and T. Riad (Mouton de Gruyter, Berlin).
- Lively, S., Logan, J. S., and Pisoni, D. (1993). "Training Japanese listeners to identify English /r/ and /l/. The role of phonetic environment and talker variability in learning new perceptual categories," *J. Acoust. Soc. Am.* **94**, 1242–1255.
- MacKay, I., Meador, D., and Flege, J. E. (2001). "The identification of English consonants by native speakers of Italian," *Phonetica* **58**, 103–125.
- Massaro, D., and Cohen, M. (1983). "Phonological context in speech perception," *Percept. Psychophys.* **34**, 338–348.
- Moreton, E. (2002). "Structural constraints in the perception of English stop-sonorant clusters," *Cognition* **84**, 55–71.
- Pallier, C., Colomé, A., and Sebastián-Gallés, N. (2001). "The influence of native-language phonology on lexical access: Exemplar-based vs. abstract lexical entries," *Psychol. Sci.* **12**, 445–449.
- Pater, J., Stager, C., and Werker, J. (2004). "The perceptual acquisition of phonological contrasts," *Language* **80**, 384–402.
- Peperkamp, S. (2007). "Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vakinin," *Cognition* **104**, 631–637.
- Peperkamp, S., and Dupoux, E. (2003). "Reinterpreting loanword adaptations: The role of perception," *Proceedings of the 15th International Congress of Phonetic Sciences*, in edited by M. J. Solé, D. Recasens, and J. Romero, Barcelona, Universitat Autònoma de Barcelona, pp. 367–370.
- Pitt, M. (1998). "Phonological processes and the perception of phonotactically illegal consonant clusters," *Percept. Psychophys.* **60**, 941–951.
- Sebastián-Gallés, N., and Soto-Faraco, S. (1999). "Online processing of native and non-native phonemic contrasts in early bilinguals," *Cognition* **72**, 111–123.
- Speciale, G., Ellis, N., and Bywater, T. (2004). "Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition," *Appl. Psycholinguist.* **25**, 293–321.
- Stager, C., and Werker, J. (1997). "Infants listen for more phonetic detail in speech perception than in word-learning tasks," *Nature (London)* **388**, 381–382.
- Storkel, H., Armbruster, J., and Hogan, T. P. (2006). "Differentiating phonotactic probability and neighborhood density in adult word learning," *J. Speech Lang. Hear. Res.* **49**, 1175–1192.
- Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (2001). "Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners," *J. Acoust. Soc. Am.* **109**, 1691–1704.
- Strange, W., Bohn, O.-S., Trent, S. A., and Nishi, K. (2004). "Acoustic and perceptual similarity of North German and American English vowels," *J. Acoust. Soc. Am.* **115**, 1791–1807.
- Weber, A., and Cutler, A. (2004). "Lexical competition in non-native spoken-word recognition," *J. Mem. Lang.* **50**, 1–25.
- Werker, J., and Tees, R. (1984). "Phonemic and phonetic factors in adult cross-language speech perception," *J. Acoust. Soc. Am.* **75**, 1866–1878.
- Wong, P., and Perrachione, T. (2007). "Learning pitch patterns in lexical identification by native English-speaking adults," *Appl. Psycholinguist.* **28**, 565–585.

Intra- and intergroup vocal behavior in resident killer whales, *Orcinus orca*

Brigitte M. Weiß^{a)}

Department of Behavioural Biology, University of Vienna, A-1090 Vienna, Austria

Helena Symonds and Paul Spong

OrcaLab, P.O. Box 258, Alert Bay, B.C., V0N 1A0, Canada

Friedrich Ladich^{b)}

Department of Behavioural Biology, University of Vienna, A-1090 Vienna, Austria

(Received 13 June 2007; revised 21 September 2007; accepted 25 September 2007)

Vocal communication within and between groups of individuals has been described extensively in birds and terrestrial mammals, however, little is known about how cetaceans utilize their sounds in their natural environment. Resident killer whales, *Orcinus orca*, live in highly stable matrilineal units and exhibit group-specific vocal dialects. Single call types cannot exclusively be associated with particular behaviors and calls are thought to function in group identification and intragroup communication. In the present study call usage of three closely related matrilineal units of the Northern resident community was compared in various intra- and intergroup contexts. In two out of the three matrilineal units significant changes in vocal behavior depending both on the presence and identity of accompanying whales were found. Most evidently, family-specific call subtypes, as well as aberrant and variable calls, were emitted at higher rates, whereas “low arousal” call types were used less in the presence of matrilineal units from different pods, subclans, or clans. Ways in which the observed changes may function both in intra- and intergroup communication.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799907]

PACS number(s): 43.80.Ka [WWA]

Pages: 3710–3716

I. INTRODUCTION

Vocal communication within and between groups of individuals has been described extensively in birds and terrestrial mammals (birds: e.g., Kroodsma and Miller, 1996; Radford, 2004; Beecher and Campbell, 2005; mammals: e.g., Seyfarth, 1987; Boughman and Wilkinson, 1998; McComb *et al.*, 2000). However, little is known about how cetaceans utilize their sounds in their natural environment (Janik, 2000), in particular, vocal interactions with conspecifics. Vocal signals relate to behavioral contexts in several species, mainly in humpback whales, *Megaptera novaeangliae*, bottlenose dolphins, *Tursiops truncatus* (see review by Tyack, 2000; Janik, 2000), southern right whales, *Eubalaena australis* (Clark, 1982), and beluga whales, *Delphinapterus leucas* (Belikov and Belkovich, 2003). Recently, Saulitis *et al.* (2005) reported context-specific calls also in the AT1 subpopulation of mammal-eating killer whales or orcas (*Orcinus orca*) in southern Alaska. Unlike “transient” orcas, the fish-eating, “resident” orcas of the northeast Pacific live in exceptionally stable matrilineal units (hereafter termed “matrilines”), where offspring of both sexes travel with their mothers lifelong (Bigg *et al.*, 1990; Ford *et al.*, 2000). They are frequently vocal and possess a complex vocal system with group-specific dialects (Ford, 1989, 1991; Yurk *et al.*, 2002) that remain stable over decades (Deecke *et al.*, 2000)

and reflect genetic relatedness (Barrett-Lennard 2000). Closely related matrilineal units are referred to as “pods” (Bigg *et al.*, 1990) and share most or all of their call repertoire. Ford (1991) grouped all pods that share any call types or subtypes into acoustic clans; subclans further define clans through use of subclan-specific call types. Relative production rates of different call types and whistles vary with broad behavioral states of the entire group (Ford, 1989), but in contrast to the AT1 transients, none of the residents’ call types correlate exclusively with any particular activity (Ford, 1989). Rather than reflecting behavioral states, the discrete call repertoires are thought to function primarily to maintain cohesion and coordinate activities in intragroup contexts (Ford, 1989, 1991).

There is increasing evidence that the individually distinct signature whistles of bottlenose dolphins function as cohesion calls when individuals of a social group are separated (Janik and Slater, 1998; Watwood *et al.*, 2005). Signature whistles may even facilitate reunions between separated individuals, especially between calves and their mothers (Smolker *et al.*, 1993). Similarly, resident orca matrilineal units were recently found to increase the usage of family-specific call types immediately after the births of calves, suggesting that family-specific call types are of profound importance for maintaining cohesion within the matriline, in particular between mothers and their dependent offspring (Weiß *et al.*, 2006). Also, call type matching in vocal exchanges within matrilineal units suggests that the discrete call types of residents function in intragroup communication (Miller *et al.*, 2004b).

^{a)}Current address: Konrad Lorenz Forschungsstelle, Fischerau 11, A-4645 Grünau im Almtal, Austria. Electronic mail: a9400355@unet.univie.ac.at

^{b)}Electronic mail: Friedrich.Ladich@univie.ac.at

In social species vocal signals are commonly found to not only serve communication within, but also between groups, and call usage and structure frequently change with the social context (e.g., Elowson and Snowdon, 1994; Smolker and Pepper, 1999; Hopp *et al.*, 2001; Snowdon and de la Torre, 2002; Baker, 2004; Radford, 2005). Group size and composition are known to affect the use of stereotyped calls in several highly social species, e.g., African elephants (Payne *et al.*, 2003) and Northern right whales (Parks and Tyack, 2005). Resident orcas are very social and matriline regularly travel and interact together irrespective of relatedness or degree of call sharing, yet, intergroup communication has received little attention and has only come into focus recently. Riesch *et al.* (2006) described stereotyped whistle types that are shared throughout the Northern resident population and potentially serve in vocal communication even between members of different acoustic clans. Also, the call design of several discrete call types suggests that they are long-range communication signals with an active space exceeding by far the distances across which members of a matriline usually separate (Miller, 2006) and the existence of multiple long-range call types suggests a role in intergroup communication. We thus suggest a significant role of discrete calls not only in intragroup, but also in intergroup, communication of resident orcas. To test this hypothesis, we analyzed call use of three Northern resident matriline in intra- and intergroup contexts, i.e., matriline traveling alone or with other matriline of different relatedness. In particular, we tested the following predictions: (1) Call use of focal matriline changes with the presence or absence of other matriline and (2) changes depend on the identity of the other matriline(s).

II. MATERIAL AND METHODS

A. Study animals and data collection

Johnstone Strait and adjacent waters off Vancouver Island, British Columbia, form the summer “core area” for the Northern resident community of orcas, which consists of more than 200 individually known orcas in three acoustic clans (Bigg *et al.*, 1990). The focus in this study was on three closely related matriline, A12, A30, and A36, comprising the most commonly encountered pod, A1 (Ford *et al.*, 2000). In October 2002, they consisted of 7, 7, and 3 individuals, respectively (Table I).

Visual data were obtained at OrcaLab, located centrally in the study area (50°34'N and 126°42'W), and through a network of observers: OrcaLab volunteers stationed at field stations, other independent researchers, and whale watch operators. Data from all sources were integrated and summarized on a daily basis. The waterways were routinely surveyed with spotting scopes; visual observations were done on an opportunistic basis, whenever whales were seen or heard within the vicinity of a station. Upon sighting, the number and identity of individuals [based on ID catalog (Ford *et al.*, 2000)], group composition, group cohesion, direction of movement, and behavioral state (travel, motionless, forage, or socialize) were recorded. As long as whales

TABLE I. Life history parameters of the individuals belonging to the three matriline within the A1 pod in the studied timeframe. ID numbers and demographic data according to Ford *et al.* (2000).

Matriline	ID	Sex	Born–Died	Mother	
A12	A12	Female	1941	unknown	
	A31	Male	1958–1997	A12	
	A33	Male	1971	A12	
	A34	Female	1975	A12	
	A55	Male	1989	A34	
	A62	Female	1993	A34	
	A67	Unknown	1996	A34	
	A74	Unknown	2000	A34	
	A30	A30	Female	1947	A2
		A6	Male	1964–1999	A30
A38		Male	1970	A30	
A39		Male	1975	A30	
A50		Female	1984	A30	
A54		Female	1989	A30	
A72		Unknown	1999	A50	
A75		Unknown	2001	A54	
A36	A36	Female	1947–1997	A1	
	A32	Male	1964	A36	
	A37	Male	1977	A36	
	A46	Male	1982	A36	

were within visual range, changes in any of the previous parameters, as well as times, when the whales passed key landmarks, were noted.

Acoustic data were collected with a hydrophone network monitored at OrcaLab 24 h a day and year round. Whales were recorded on a two-channel audio cassette recorder (Sony Professional Walkman WM-D6C or Sony TCD-D3) with up to six radio-transmitting, custom-made hydrophone stations (overall system frequency response 10 Hz–15 kHz) whenever they were vocal (see Weiß *et al.*, 2006). Data collection was strictly land based and thus did not interfere with or disturb the whales.

B. Acoustic analyses

Focal matriline were frequently observed and recorded with matriline from different pods (closely related matriline), subclans, and clans. For investigating the intra- and intergroup vocal behavior we selected recordings where focal matriline were encountered in one of five clearly defined social contexts: (1) alone, (2) together with the other two A1 matriline (“same pod”), (3) in the company of matriline belonging to a different pod within the same acoustic subclan (“same subclan”), (4) in the company of matriline belonging to a different subclan within the same clan (“other subclan”), or (5) in the company of matriline belonging to a different clan (“other clan”). “Alone” referred to situations in which only the focal matriline was seen or heard within the same or adjacent hydrophone range(s). A focal matriline was considered to be in the company of another matriline when both were observed within acoustic range of each other, were heading in the same direction and were engaged in the same behavior. Distances between matriline traveling in company were estimated with the help

TABLE II. Number of samples, calls and recording days of focal matriline in varying social contexts.

Matriline	With	<i>n</i> samples	<i>n</i> calls	<i>n</i> days
A12	...	10	983	7
A12	Other pod	6	570	3
A12	Other subclan	2	181	1
A12	Other clan	2	159	1
A30	...	15	1500	14
A30	Other pod	4	353	3
A30	Other subclan	2	186	1
A30	Other clan	6	550	4
A36	...	10	965	7
A36	Other pod	3	258	3
A36	Other subclan	4	366	2
A36	Other clan	9	810	6
A12+A30+A36		8	783	4

of landmarks and were typically well below 1000 m. We only used recordings for further analysis during which the spacing, direction of travel, and behavioral states of the involved matriline were observed from shore or were reported from whale watching boats, and that allowed definite attribution of calls to the matriline in a defined situation. This excluded night-time recordings as well as those where one or more additional matriline were seen and/or heard within range of the same hydrophone as the defined matriline(s). The selected recordings were obtained between 1989 and 2002 except for one recording of the A30 and B7 matriline, that was obtained in August 2005. The predominant behaviors were traveling and/or foraging.

Calls were classified according to Ford (1987, 1989, 1991) by simultaneous acoustic and visual inspection of sonagrams, generated with Cool Edit 2000 (Syntrillium Software Corporation) or Raven 1.2 (Cornell Lab of Ornithology). Two call subtypes, N5iii and N9iv (Weiß *et al.*, 2006), were additionally distinguished because they were family specific to focal matriline.

C. Statistical analyses

If more than 5% of calls were not both visually and acoustically recognizable because of poor signal-to-noise ratio, recordings were excluded from statistical analysis to

avoid a bias towards call types of higher amplitude (see Miller and Tyack, 2001). The remaining data were split into samples of 100 calls. Preferably, samples were chosen from different recording days. However, because selection criteria strongly reduced the number of usable samples in some of the defined social contexts, we also included recordings with less than 100, but a minimum of 75 calls. For the same reason, we sometimes used multiple samples from the same day, but as widely separated in time as possible and never more than three to maximize statistical independence of the data (Table II). For each sample, we determined percentages of call use per call (sub-)type as well as the call rate (*n* calls/minute/individual) and the number of different call types used.

Data were analyzed using the SPSS® statistical program. As data clearly deviated from normal distribution (Shapiro-Wilk, all parameters $p < 0.02$), they were tested nonparametrically. Also, data were tested separately for each matriline, as basic call use differs somewhat between the three focal matriline (Miller and Bain, 2000; Weiß *et al.*, 2006). Frequencies of call types and numbers were compared between single focal matriline and focal matriline in company using Mann-Whitney-U tests. Because calls could not be reliably attributed to the producing matriline when all three focal matriline were recorded together, we did not compare recordings of the three matriline together with those of the single matriline, but rather with several averaged samples (frequency of a given call type for A12 + A30 + A36 divided by 3). In those cases, where call use did differ between the single and the company contexts, we conducted Kruskal-Wallis tests to further test for differences in call use depending on the identity of the company. Comparisons of call use between each single social context were not feasible due to an *n* below 5 in 6 of 9 of the company contexts. We did not consider alpha correction for multiple testing, because of an increased risk of type-II error due to small sample sizes (Nakagawa, 2004). Call types with rates of occurrence below 1% in any context were included in the category “other” for the given matriline. All statistical tests were two-tailed.

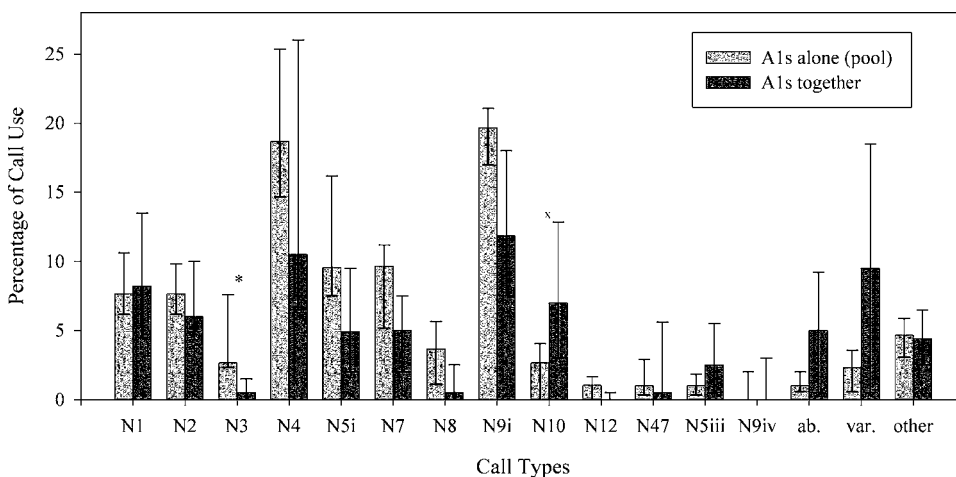


FIG. 1. Call use of all three focal matriline traveling on their own or with each other. ab, aberrant and var, variable. Bars show median percentage of total calls and first and third quartiles. Asterisks mark significance levels: * = $p < 0.05$; x = $p < 0.06$; and $n = 8$ (alone) and 9 (together).

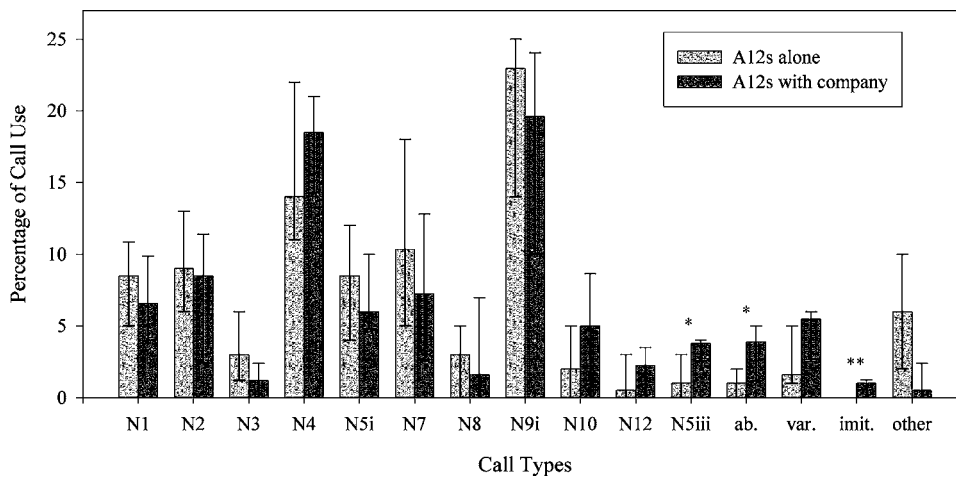


FIG. 2. Call use of the A12 matriline traveling on its own or with whales from different pods or (sub-)clans. ab., aberrant; var., variable; and imit., imitation. Bars show median percentage of total calls and first and third quartiles. Asterisks mark significance levels: * = $p < 0.05$; ** = $p < 0.01$; and $n = 10$ (alone) and 10 (with company).

III. RESULTS

Altogether, 81 samples totaling 7664 calls of focal matrilines on their own or in the company of other groups were of sufficient quality for statistical analysis (Table II). Call patterns when matrilines were alone were comparable to those described by Miller and Bain (2000) and Weiß *et al.* (2006) for the three focal matrilines.

With the exception of N3 calls and N10 calls, mean call use of the single A1 matrilines was very similar to that of the A1 matrilines traveling together (Fig. 1). N3 calls made up a significantly smaller proportion of calls when all three A1 matrilines were together (Mann-Whitney U test, $n=17$, $U=9$, $p=0.008$), while the use of N10 calls tended to increase in such situations (Mann-Whitney U test, $n=17$, $U=16$, $p=0.059$).

In the presence of more distantly or unrelated matrilines from other pods/clans, the A12 matriline was found to increase the use of the family-typical N5iii, as well as aberrant and imitation calls (Fig. 2, Mann-Whitney U test, N5iii: $n=20$, $U=22.5$, $p=0.035$, aberrant: $n=20$, $U=20$, $p=0.023$, imitation: $n=20$, $U=15$, $p=0.007$). In each case, changes in call use were below 5%. Although we did not find any significant differences in call use of the A30 matriline traveling with or without company (Fig. 3), they used the family typical N47 call 2–3 times more often in the presence of B-subclan for which two samples were available. The A36 matriline showed a number of distinct changes (Fig. 4). As in

the A12 matriline, the family-typical call subtype, N9iv, was used significantly more often in the presence of other groups, which was also the case for variable calls (Mann-Whitney U test, N9iv: $n=26$, $U=25$, $p=0.003$, variable: $n=26$, $U=21.5$, $p=0.001$). On the other hand, N1, N3, N7, N8 and N9i calls made up significantly higher portions of the call repertoire when the matriline was recorded on its own (Mann-Whitney U test, N1: $n=26$, $U=41.5$, $p=0.041$, N3: $n=26$, $U=24$, $p=0.002$, N7: $n=26$, $U=33$, $p=0.012$, N8: $n=26$, $U=32$, $p=0.01$, N9i: $n=26$, $U=34.5$, $p=0.014$). With up to 10% change in call use, differences were more pronounced in the A36 matriline than in the A12 matriline.

The use of some call types also differed depending on which group accompanied a focal matriline (Fig. 5, Kruskal-Wallis test, A12—aberrant: $n=10$, $H=5.595$, $p=0.061$, A36—N7: $n=10$, $H=5.854$, $p=0.054$, A36—variable: $n=10$, $H=6.014$, $p=0.049$), which suggests that changes in call use may not only have been affected by the presence of other groups, but also by their identity. However, sample sizes in the different conditions were low and more detailed analyses were not feasible.

Focal matrilines did not differ in their call rate (n calls/minute/individual) nor in the number of different call types used in any of the contexts (Kruskal-Wallis test, all $p > 0.1$, Mann-Whitney U test, all $p > 0.1$).

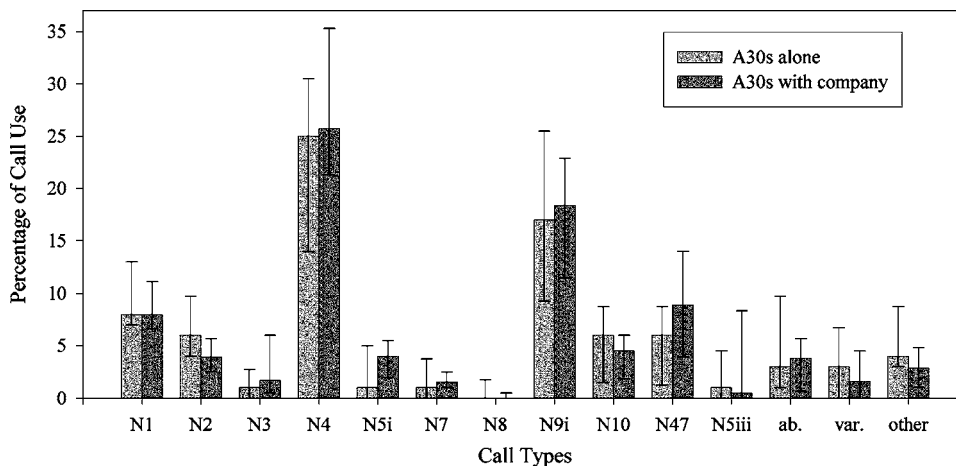


FIG. 3. Call use of the A30 matriline traveling on its own or with whales from different pods or (sub-)clans. ab., aberrant and var., variable. Bars show median percentage of total calls and first and third quartiles. $n=15$ (alone) and 12 (with company).

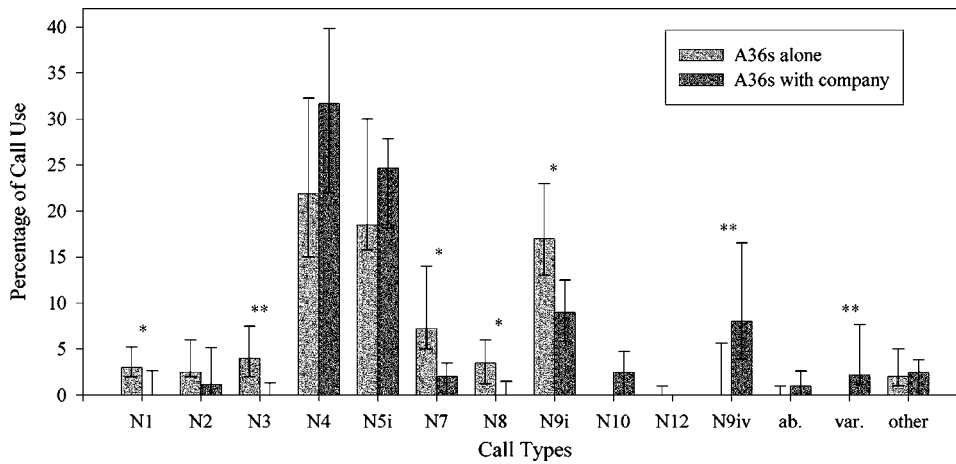


FIG. 4. Call use of the A36 matriline traveling on its own or with whales from different pods or (sub-)clans. ab., aberrant and var., variable. Bars show median percentage of total calls and first and third quartiles. Asterisks mark significance levels: * = $p < 0.05$. ** = $p < 0.01$; and $n = 10$ (alone) and 16 (with company).

IV. DISCUSSION

This study presents evidence that resident orca matriline change their vocal behavior in intergroup contexts. Consistent with our predictions, percentages of call types used by focal matriline depended on the presence or absence of additional matriline in two out of three studied matriline; findings for the third matriline were also consistent in key respects, though sample sizes were small and the changes were not statistically significant. The most consis-

tent changes were increases in the use of family-specific call subtypes as well as variable and aberrant calls in the presence of orcas from other groups. Changes in call use were typically well below 10% and thus less pronounced than the changes observed in call use in another social context, i.e., after the birth of a calf (Weiß *et al.*, 2006).

The focal matriline differed considerably in the number of call types with emission rates affected by the social context. The A36 matriline showed significant changes in call use in half of their repertoire (7 of 14 call types), the A12 matriline to a considerably lesser extent (3 of 15 call types), and there were no statistically significant changes at all in the A30 matriline. However, their family typical N47 call was used 2–3 times as often in the presence of B-subclan matriline and interestingly, changes in the A36s' acoustic behavior were also strongest in the presence of matriline from B-subclan.

To some extent, the differences in call use may reflect low sample sizes for some matriline in some contexts; however, it is also possible that they reflect different social roles arising from differences in associations and movement patterns. At least one of the three focal matriline is present in the study area, almost daily, during each summer and fall (Symonds and Spong, private communication). Changes in call use exhibited by the A1 matriline thus may reflect differences in the manner in which each of the matriline responds to intergroup situations involving the other, less commonly visiting Northern resident groups in the Johnstone Strait area, and they may underscore possible differing social roles within their own pod and community. Just how each of the A1 matriline performs its role and specifically what these roles might be is beyond the scope of this study, however, the indication that matriline within a pod have possible different roles should encourage further investigation.

It also seems possible that some of the changes we observed reflect differences in the age and sex composition of individual matriline. Indeed, the three focal matriline do differ in this respect; the A36 matriline consisted of only one female and her three adult sons prior to the matriarch's death in 1997, and since then only of adult males, whereas the other two matriline each added a new generation during the study period. Studies of other species show that vocal signals can convey not only individual or group-specific informa-

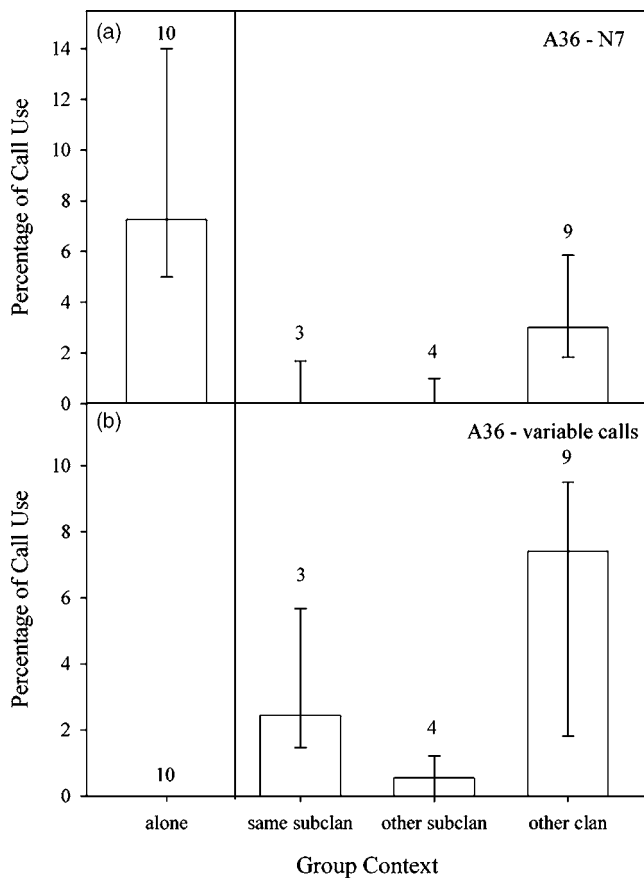


FIG. 5. Use of (a) N7 calls and (b) variable calls by the A36 matriline in different social contexts (focal matriline alone, with matriline from the same subclan, another subclan or another clan). Bars show median percentage of calls and first and third quartiles. Numbers of samples appear above the bars.

tion, but also age and sex-specific cues that may remain distinguishable even when transmitted over greater distances (e.g. Green, 1981; Rendall *et al.*, 2004; Blumstein and Munos, 2005). Animals may respond differently to calls of males and females (e.g., Vicario *et al.*, 2001; Miller *et al.*, 2004a) or may respond differently to signals depending on their own sex (Rogers *et al.*, 2006). It is plausible, therefore, that the differences in call use among our focal matriline stemmed partly from different age or sex distributions in the focal or the accompanying matriline. However, as virtually nothing is known about age or sex differences in the call use of wild orcas, this idea remains purely speculative.

At least in some call types, changes in use seemed not only dependent on the mere presence of nonfocal matriline, but to some degree on whether these nonfocal whales were from a pod within the same subclan, from a different subclan, or from an entirely different clan, i.e., whales that differ both in the degree of relatedness and of call type sharing. Although the focal matriline share almost all call types with whales within the same subclan, they share few with those from a different subclan and none with whales from a different clan (Ford, 1987, 1991). Studies in other species have shown that the degree of vocal sharing may play an important role in acoustic communication between groups and individuals. For instance, song sparrows, *Melospiza melodia*, were more likely to perceive a song as directed at them if the song was shared than if it was unshared (Beecher and Campbell, 2005); and great tits, *Parus major*, responded differently to song types shared with neighbors and strangers (Stoddard, 1996). Observations of primates are consistent with those for birds. Lemasson and Hausberger (2004) observed more vocal exchanges in Campbell's monkeys, *Cercocebus campbelli*, that shared calls and suggested sharing to be important in advertising bonds.

Complex vocal signals can serve multiple functions in inter-group behavior, such as cooperation, inter- and intra-sexual assessment between groups (e.g., Seddon, 2002), and call repertoire and complexity are often parameters used in mate choice (see McGregor, 1992). The higher percentages of variable, aberrant and in part also imitation calls in inter-group contexts may reflect some of these vital aspects of social interactions. Variable and aberrant calls have been associated with situations of high arousal, e.g. during socializing (Ford, 1989, 1991) and may reflect a more sexually charged situation, since whales from disparate groups are more likely to mate (Barrett-Lennard, 2000). Ford (1989) also reported whales to be highly vocal when engaged in socializing (i.e. showing physical contact, aerial displays etc.). However, the mere presence of other groups that were not engaged in active socializing, did not lead to an increase in the call rates of our focal matriline. In contrast to states of high arousal, "low arousal" calls like N3 and N7 calls are heard most often during behaviours such as resting (Ford, 1989; Symonds and Spong, private communication), so their tendency to occur more frequently in matriline on their own might reflect a similar state of "low arousal" in this particular social context. Finally, the lack of changes in the majority of discrete call types may indicate that they serve the same

functions in intra- and intergroup communication, such as directionality cueing and thus indicating one's location and direction of movement (Miller, 2002).

In conclusion, the presence and identity of accompanying matriline significantly affected calling behavior of resident orca matriline. The observed changes seem to reflect call functions in both intragroup, as well as intergroup, communication, and differences between matriline hint at possible different social roles within the community. To get a better understanding of these roles, we will need extensive data sets of individual calling behavior along with precise behavioral observations. Both are currently extremely difficult to obtain, but provide potentially fruitful challenges for future research.

ACKNOWLEDGMENTS

The authors wish to thank Anna Spong and OrcaLab assistants for continuous recording and observation efforts. Also, they are grateful for the contributions of observers, researchers, and whale-watchers in the Johnstone Strait area, especially the late Michael Bigg, Graeme Ellis, and John Ford. The study was supported by a doctoral scholarship of the Austrian Academy of Sciences to B.M.W.

- Baker, M. C. (2004). "The chorus song of cooperatively breeding laughing kookaburras (Coraciiformes, Halcyonidae: *Dacelo novaeguineae*): characterization and comparison among groups," *Ethology* **110**, 21–35.
- Barrett-Lennard, L. G. (2000). "Population structure and mating patterns of killer whales, *Orcinus orca*, as revealed by DNA analysis," Ph.D. thesis, University of British Columbia, Vancouver.
- Beecher, M. D., and Campbell, S. E. (2005). "The role of unshared songs in singing interactions between neighbouring song sparrows," *Anim. Behav.* **70**, 1297–1304.
- Belikov, R. A., and Belkovich, V. M. (2003). "Underwater vocalization of the beluga whales (*Delphinapterus leucas*) in a reproductive gathering during different behavioral situations," *Okeanologiya* **43**, 118–126.
- Bigg, M. A., Olesiuk, P. F., and Ellis, G. M. (1990). "Social organization and genealogy of resident killer whales (*Orcinus orca*) in the coastal waters of British Columbia and Washington State," *Rep. Int. Whal. Comm., Spec. Issue*, **12**, 383–405.
- Blumstein, D. T., and Munos, O. (2005). "Individual, age- and sex-specific information is contained in yellow-bellied marmot alarm calls," *Anim. Behav.* **69**, 353–361.
- Boughman, J. W., and Wilkinson, G. S. (1998). "Greater spear-nosed bats discriminate group mates by vocalizations," *Anim. Behav.* **55**, 1717–1732.
- Clark, C. W. (1982). "The acoustic repertoire of the southern right whale, a quantitative analysis," *Anim. Behav.* **30**, 1060–1071.
- Deecke, V. B., Ford, J. K. B., and Spong, P. (2000). "Dialect change in resident killer whales (*Orcinus orca*): Implications for vocal learning and cultural transmission," *Anim. Behav.* **60**, 629–638.
- Elowson, A. M., and Snowdon, C. T. (1994). "Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment," *Anim. Behav.* **47**, 1267–1277.
- Ford, J. K. B. (1987). "A catalogue of underwater calls produced by killer whales (*Orcinus orca*) in British Columbia," *Canadian Data Report of Fisheries and Aquatic Science* No. 633.
- Ford, J. K. B. (1989). "Acoustic behaviour of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia," *Can. J. Zool.* **67**, 727–745.
- Ford, J. K. B. (1991). "Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia," *Can. J. Zool.* **69**, 1454–1483.
- Ford, J. K. B., Ellis, G. M., and Balcomb, K. C. (2000). *Killer Whales: The Natural History and Genealogy of Orcinus Orca in British Columbia and Washington*, 2nd ed. (UBC Press, Vancouver).
- Green, S. M. (1981). "Sex differences and age gradations in vocalizations of Japanese and lion-tailed monkeys (*Macaca fuscata* and *Macaca silenus*)," *Am. Zool.* **21**, 165–183.
- Hopp, S. L., Jablonski, P., and Brown, J. L. (2001). "Recognition of group

- membership by voice in Mexican jays, *Aphelocoma ultramarina*,” *Anim. Behav.* **62**, 297–303.
- Janik, V. M. (2000). “Food-related bray calls in wild bottlenose dolphins (*Tursiops truncatus*),” *Proc. R. Soc. London, Ser. B* **267**, 923–927.
- Janik, V. M., and Slater, P. J. B. (1998). “Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls,” *Anim. Behav.* **56**, 829–838.
- Kroodsma, D. E., and Miller, E. H. (1996). *Ecology and evolution of acoustic communication in Birds* (Cornell University Press, Ithaca, N.Y.).
- Lammers, M. O., Schotten, M., and Au, W. W. L. (2006). “The spatial context of free-ranging Hawaiian spinner dolphins (*Stenella longirostris*) producing acoustic signals,” *J. Acoust. Soc. Am.* **119**, 1244–1250.
- Lemasson, A., and Hausberger, M. (2004). “Patterns of vocal sharing and social dynamics in a captive group of Campbell’s monkeys (*Cercopithecus campbelli campbelli*),” *J. Comp. Psych.* **118**, 347–359.
- McComb, K., Moss, C., Sayialel, S., and Baker, L. (2000). “Unusually extensive networks of vocal recognition in African elephants,” *Anim. Behav.* **59**, 1103–1109.
- McGregor, P. K. (1992). *Playback and studies of animal communication*,” NATO ASI Series, Series A: Life Sciences, Vol. **228** (Plenum, New York).
- Miller, P. J. O. (2002). “Mixed-directionality of killer whale stereotyped calls: a direction of movement cue?,” *Behav. Ecol. Sociobiol.* **52**, 262–270.
- Miller, P. J. O. (2006). “Diversity in sound pressure levels and estimated active space of resident killer whale vocalizations,” *J. Comp. Physiol. [A]* **192**, 449–459.
- Miller, P. J. O., and Bain, D. E. (2000). “Within-pod variation in the sound production of a pod of killer whales, *Orcinus orca*,” *Anim. Behav.* **60**, 617–628.
- Miller, P. J. O., and Tyack, P. L. (2001). “Mixed-directionality of *O. orca* stereotyped calls: a design feature promoting social cohesion,” *Proceedings of the 14th Biennial Conference on the Biology of Marine Mammals*, Vancouver.
- Miller, C. T., Scarl, J., and Hauser, M. D. (2004a). “Sensory biases underlie sex differences in tamarin long call structure,” *Anim. Behav.* **68**, 713–720.
- Miller, P. J. O., Shapiro, A. D., Tyack, P. L., and Solow, A. R. (2004b). “Call-type matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca*,” *Anim. Behav.* **67**, 1099–1107.
- Nakagawa, S. (2004). “A farewell to Bonferroni: the problems of low statistical power and publication bias,” *Behav. Ecol. Sociobiol.* **15**, 1044–1045.
- Parks, S. E., and Tyack, P. L. (2005). “Sound production by North Atlantic right whales (*Eubalaena glacialis*) in surface active groups,” *J. Acoust. Soc. Am.* **117**, 3297–3306.
- Payne, K. B., Thompson, M., and Kramer, L. (2003). “Elephant calling patterns as indicators of group size and composition: The basis for an acoustic monitoring system,” *Afr. J. Ecol.* **41**, 99–107.
- Radford, A. N. (2004). “Vocal Coordination of Group Movement by Green Woodhoopoes (*Phoeniculus purpureus*),” *Ethology* **110**, 11–20.
- Radford, A. N. (2005). “Group-specific vocal signatures and neighbour-stranger discrimination in the cooperatively breeding green woodhoopoe,” *Anim. Behav.* **70**, 1227–1234.
- Rendall, D., Owren, M. J., Weerts, E., and Hienz, R. D. (2004). “Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners,” *J. Acoust. Soc. Am.* **115**, 411–421.
- Riesch, R., Ford, J. K. B., and Thomsen, F. (2006). “Stability and group specificity of stereotyped whistles in resident killer whales, *Orcinus orca*, off British Columbia,” *Anim. Behav.* **71**, 79–91.
- Rogers, A. C., Mulder, R. A., and Langmore, N. E. (2006). “Duet duels: sex differences in song matching in duetting eastern whipbirds,” *Anim. Behav.* **72**, 53–61.
- Saulitis, E. L., Matkin, C. O., and Fay, F. H. (2005). “Vocal repertoire and acoustic behavior of the isolated AT1 killer whale subpopulation in southern Alaska,” *Can. J. Zool.* **83**, 1015–1029.
- Seddon, N. (2002). “The structure, context and possible functions of solo, duets and choruses in the subdesert mesite (*Monias benschi*),” *Behaviour* **139**, 645–676.
- Seyfarth, R. M. (1987). “Vocal communication and its relation to language,” in *Primate Societies*, edited by B. B. Smuts, D. L. Cheney, R. M. Seyfarth, R. W. Wrangham, and T. T. Struhsaker (University of Chicago Press, Chicago), pp. 440–451.
- Smolker, R. A., Mann, J., and Smuts, B. B. (1993). “Use of signature whistles during separations and reunions by wild bottlenose dolphin mothers and infants,” *Behav. Ecol. Sociobiol.* **33**, 393–402.
- Smolker, R. A., and Pepper, J. W. (1999). “Whistle convergence among allied male bottlenose dolphins (Delphinidae, *Tursiops* sp.),” *Ethology* **105**, 595–617.
- Snowdon, C. T., and de la Torre, S. (2002). “Multiple environmental contexts and communication in pygmy marmosets (*Cebuella pygmaea*),” *J. Comp. Psychol.* **116**, 182–188.
- Stoddard, P. K. (1996). “Vocal recognition of neighbors by territorial passerines,” in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca, NY), pp. 356–374.
- Tyack, P. L. (2000). “Functional aspects of cetacean communication,” in *Cetacean Societies: Field Studies of Dolphins and Whales*, edited by J. Mann, R. C. Connor, P. L. Tyack, and H. Whitehead (University of Chicago Press, Chicago), pp. 270–307.
- Vicario, D. S., Naqvi, N. H., and Raksin, J. N. (2001). “Sex differences in discrimination of vocal communication signals in a songbird,” *Anim. Behav.* **61**, 805–817.
- Watwood, S. L., Owen, E. C. G., Tyack, P. L., and Wells, R. S. (2005). “Signature whistle use by temporarily restrained and free-swimming bottlenose dolphins, *Tursiops truncatus*,” *Anim. Behav.* **69**, 1373–1386.
- Wei, B. M., Ladich, F., Spong, P., and Symonds, H. (2006). “Vocal behavior of resident killer whale matriline with newborn calves: The role of family signatures,” *J. Acoust. Soc. Am.* **119**, 627–635.
- Yurk, H., Barrett-Lennard, L., Ford, J. K. B., and Matkin, C. O. (2002). “Cultural transmission within maternal lineages: Vocal clans in resident killer whales in southern Alaska,” *Anim. Behav.* **63**, 1103–1119.

Flying big brown bats emit a beam with two lobes in the vertical plane

Kaushik Ghose and Cynthia F. Moss

Neuroscience and Cognitive Science Program, Institute for Systems Research, Dept. of Psychology, University of Maryland, College Park, Maryland 20742

Timothy K. Horiuchi

Neuroscience and Cognitive Science Program, Institute for Systems Research, Dept. of Electrical and Computer Engineering, University of Maryland, College Park, Maryland 20742

(Received 18 January 2007; revised 12 September 2007; accepted 18 September 2007)

The sonar beam of an echolocating bat forms a spatial window restricting the echo information returned from the environment. Investigating the shape and orientation of the sonar beam produced by a bat as it flies and performs various behavioral tasks may yield insight into the operation of its sonar system. This paper presents recordings of vertical and horizontal cross sections of the sonar beam produced by *Eptesicus fuscus* (big brown bats) as they fly and pursue prey in a laboratory flight room. In the horizontal plane the sonar beam consists of one large lobe and in the vertical plane the beam consists of two lobes of comparable size oriented frontally and ventrally. In level flight, the bat directs its beam such that the ventral lobe is pointed forward and down toward the ground ahead of its flight path. The bat may utilize the downward directed lobe to measure altitude without the need for vertical head movements. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2799491]

PACS number(s): 43.80.Ka, 43.66.Qp, 43.80.Lb [JAS]

Pages: 3717–3724

I. INTRODUCTION

Insectivorous echolocating bats, such as the big brown bat, *Eptesicus fuscus*, navigate and forage for airborne insects in darkness. They produce intermittent pulses of directed ultrasound and use the information contained in the returning echoes to detect, localize and track flying insect prey, relying on hearing to guide complex spatial behaviors.^{1,2} The sonar beam is directional¹ and may serve as a spatial window, restricting the region of space where the bat gathers information. Knowing the shape and direction of the sonar beam as the bat performs sonar tasks is therefore important to our understanding of bat echolocation.

The sonar beam in stationary, head-fixed bats has been studied by several investigators.^{3,4} Hartley and Suthers performed detailed studies of the sonar beam in *E. fuscus* by measuring the shape of the beam in both elevation and azimuth in multiple frequency bands.⁵ The bats were anesthetized and positioned on a platform that controlled the position of the head. The midbrain was electrically stimulated to elicit ultrasonic vocalizations. The sonar beam pattern recorded in that condition consisted of a large main lobe. At frequencies higher than 60 kHz a second lobe appeared, 6 dB less intense than the main lobe. This second lobe was located ventral to the main lobe, at an angle of 30°.

It is not known whether the sounds elicited by brain stimulation in a stationary, head-fixed bat produce beam patterns similar to that in a flying bat. The sonar beam pattern depends not only on the frequency content of the signal, but also on the shape of the vocal cavity and, potentially, the configuration of other body structures, such as wings, during the time of emission. It is important, therefore, to study the sonar beam patterns produced by bats executing natural

echolocation behaviors in flight such as insect capture. In this study we recorded vertical and horizontal cross sections of the sonar beam produced by echolocating bats as they flew and pursued prey in a laboratory flightroom. We were limited by technical considerations to recording a narrow band of frequencies. We chose a frequency band centered around 35 kHz as this is roughly the peak of the first-harmonic sweep, and we found that, within this band, we could record signals from all foraging stages. In contrast to the study by Hartley and Suthers⁵ we find evidence for a second lobe, directed ventrally to the main lobe, at these lower frequencies. The size of the ventral lobe is comparable to the frontal “main” lobe. We speculate that the ventral lobe may serve to generate a ground-return that helps the bat to measure its altitude in flight without the need for head movements.

II. METHODS

A. Behavioral experiments

We trained two bats of the species *E. fuscus* to fly individually in a large (L7.3 m × W6.4 m × H2.5 m) laboratory flightroom (Fig. 1). The bats were trained to catch insects (mealworms) suspended from a tether. The tether was attached to a motorized boom placed at random points under the ceiling. The insect could be hung stationary from the boom, or swung in horizontal arcs by the boom to present a moving target to the bat. The walls and ceiling of the flight room were lined with sound-absorbent acoustic foam (Sonex One, Acoustical Solutions, Inc., Richmond, VA) to reduce reverberations. The room was illuminated by dim, long wavelength light (>650 nm, light from normal incandescent

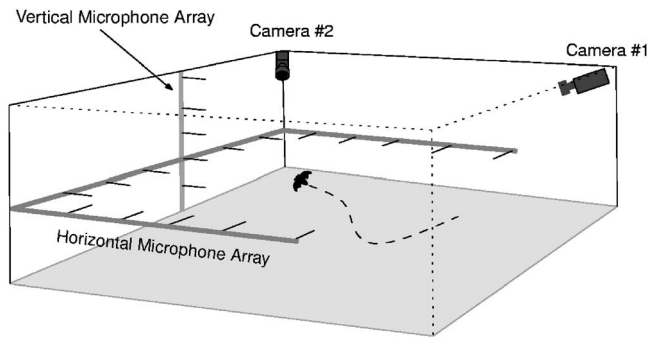


FIG. 1. Laboratory flight room.

bulbs passed through a infrared filter plate—Plexiglas G #2711, Atofina Chemicals, Philadelphia, PA) to which *E. fuscus* is insensitive.⁶ Stereo images from two high speed video cameras [Kodak MotionCorder, CCD based cameras operating at 240 frames/s (4.16 ms sampling interval), synchronized to 1/2 frame accuracy] were used to reconstruct the three-dimensional flight path of the bat and the trajectory of the prey. Simultaneously, an array of 20 microphones (Knowles Co, FG 3329, electret) recorded the sonar beam pattern produced by the bat. 16 microphones were arranged in a horizontal U-shape around the room to extract the horizontal aspect of the sonar beam. The remaining 4 microphones were arranged vertically on one wall of the room. These four microphones and one microphone from the horizontal arrangement formed a five microphone vertical array that enabled us to record vertical cross-sections of the emitted sonar beam. The Knowles microphones have a small membrane diameter (~ 1 mm) and are omnidirectional at 35 kHz ($\lambda_{35 \text{ kHz}}=9.7$ mm) enabling recording of the beam pattern as the bat flies around the room.⁷ The microphone signals were amplified, band-pass filtered ($f_c=35$ kHz, $Q_{-3 \text{ dB}}=2.5$, half-power band ranging from 28 to 42 kHz) and then processed through an envelope extractor circuit. The envelopes of the bat vocalizations were digitized at 20 kHz and stored on a computer for later analysis.⁷

B. Computation of the sonar beam pattern

The envelope traces were used to compute the received intensity of the sonar beam, $I_{r,m}$ at each microphone m . This intensity was corrected for spherical spreading loss and atmospheric attenuation to obtain the corrected, normalized intensity, $I_{c,m}$ at each microphone. Previous experiments on stationary anesthetized bats⁵ and from flying bats^{7,8} have indicated that the sonar beam is horizontally symmetrical. In this study we fit the horizontal sonar beam pattern to a Gaussian shape in order to extract the sonar beam-axis direction. The fit is performed by adjusting the direction and width of the Gaussian beam pattern to obtain the least-mean-square error to the observed beam pattern. We have previously used a vector averaging method⁷ to obtain the sonar beam direction. We observed that fitting the sonar beam to a Gaussian shape before computing the axis direction made the computation less sensitive to edge effects⁷ (Sec. III C) and less sensitive to variations in the recorded sonar beam profiles from vocalization to vocalization.

C. Microphone array calibration

The microphone array was calibrated for gain and frequency response by playing ultrasonic frequency sweeps of a fixed amplitude from an ultrasonic speaker with a circular aperture of 1.5 cm placed at a fixed distance (40 cm) and orientation to each microphone. A gain factor was computed for each array channel (microphone and signal processing circuit) that normalized the recordings across all the channels. This speaker was also used as a control emitter to test the microphone array. Recordings of cross sections of the sonar beam produced by the control emitter appeared as single lobes on the horizontal and vertical arrays.

III. RESULTS

We analyzed flight paths, vocalization timing patterns and sonar-beam patterns from a total of 15 flights from two bats as they intercepted insects. Horizontal sonar beam patterns appeared as a single lobe, consistent with previous recordings.⁷ In all trials, however, we observed notched beam shapes recorded by the vertical array.

A. The sonar beam has a prominent notch in the dorso-ventral plane

We observed a prominent notch in the vertical sonar beam profile. Figure 2 shows the sonar beam pattern of a single vocalization. The top panel shows the top view of the flight room. The bottom panel shows the sonar beam profile recorded by the vertical array. The dotted lines show the “field-of-view” of the vertical array. These lines show the angular limits of the vertical array, beyond which we cannot measure the sonar beam pattern. As can be seen, in contrast to the single prominent lobe shape of the horizontal cross section, the vertical cross section presents a forked shape. This shape is suggestive of two lobes of comparable size arranged dorso-ventrally. At this stage the bat was producing pulses at a low repetition rate (<20 Hz) which are associated with searching behaviors. The bat’s flight maneuvers suggest it was not responding to the target yet.

Figure 3 shows the sonar beam pattern of a vocalization approximately one second later for the trial shown in Fig. 2. At this time, the bat was producing pulses at a high repetition rate (>100 Hz) and was responding to the target. The vertical sonar beam profile shows no evidence of a notch for this vocalization. If the bat’s sonar beam had two prominent lobes in the vertical plane and if the bat was directing its sonar beam down at the target (black cross) such that the location of the lower lobe is below the lower angular limit of the array (black dotted lines) this is the expected shape of the recorded sonar beam cross section. Figures 4–6 present further evidence that the bat moves the sonar beam in the vertical plane, perhaps in order to track the vertical position of the target.

Figure 4 shows six vocalizations from a sequence of calls made by bat #1 as it intercepted a tethered insect. To reduce clutter in the diagram, not all vocalizations in the pulse train produced by the bat are shown. Dotted lines in the bottom panel show the field-of-view of the vertical array beyond which the shape of the vertical beam could not be

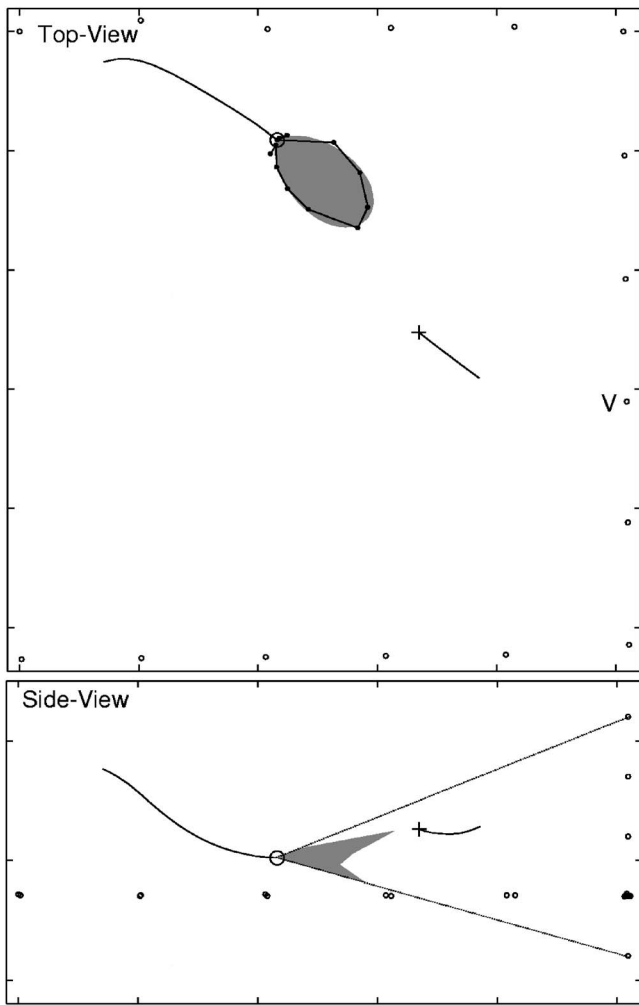


FIG. 2. The sonar beam of flying *E. fuscus* has a vertical notch. (Top) The small black circles are the positions of the array microphones, the black line is the flight path of the bat, the black circle is the current position of the bat. The dark gray polar plot represents the smoothed profile of the horizontal sonar beam, with radius proportional to normalized intensity. The black dots joined by a line are the unsmoothed measured intensities. The black line is the trajectory of the tethered insect and the black cross is the present position of the target. The position of the vertical array is indicated by the label "V." (Bottom) Polar plot of the vertical section of the sonar beam recorded by the vertical array (gray polygon). Dotted lines show the angular limits of the vertical array measurement. The bat is producing pulses at a slow repetition rate (<20 Hz) at this stage.

recorded. Vocalization samples 1–5 show a prominent notch in the vertical cross section, whereas the notch is absent in 6. The bat is responding to the target during 5 and 6, and is swooping down on the target during 6. The vertical sonar beam patterns are consistent with the hypothesis that the bat's sonar beam has two lobes arranged dorso-ventrally. During vocalization samples 1–4 the bat is directing the beam such that one lobe points up and the other down to the ground. The absence of a notch in 6 is consistent with the bat directing its beam downward, with the main lobe pointing in the direction of the target.

Figures 5 and 6 show vocalization sequences from bats 1 and bat 2, respectively, each showing the notched appearance of the sonar beam. Initially the bat is producing signals at a low rate (<20 Hz). In later stages the bat locks its beam onto the target (black cross) in the horizontal plane and produces

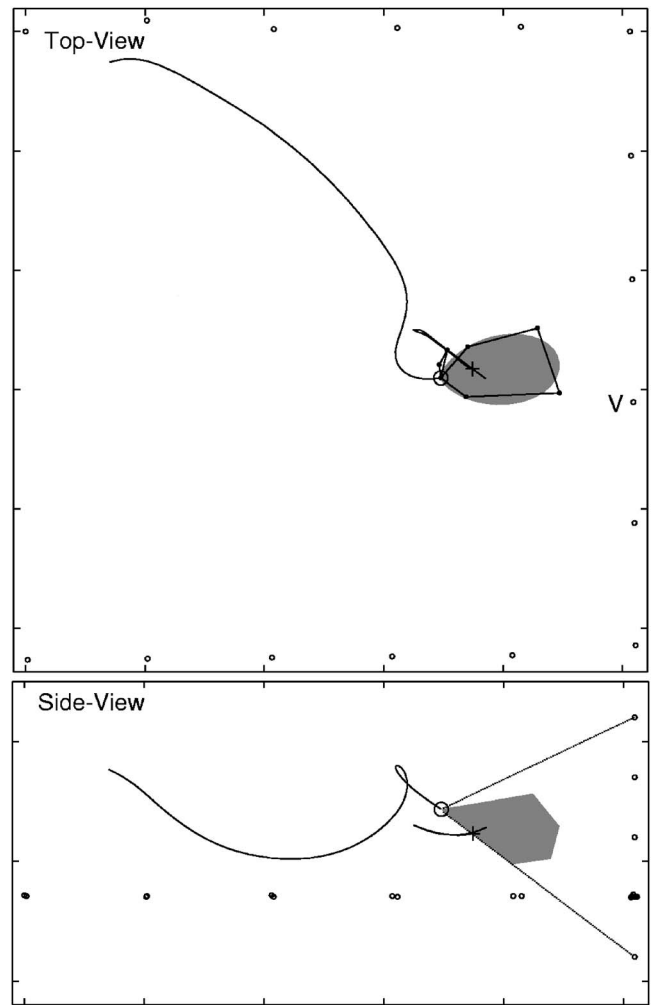


FIG. 3. The bat may move its sonar beam up and down. A sonar beam pattern approximately 1 s after the vocalization shown in Fig. 2. (Top-view) Bat is responding to the target by adjusting its flight path and that the horizontal aspect of the beam is locked to the target. (Side-view) No evidence of a notch in the beam. If the bat's sonar beam had two prominent lobes in the vertical plane and if the bat was directing its sonar beam down at the target (black cross) such that the location of the lower lobe is below the lower angular limit of the array (black dotted lines) this is the expected shape of the recorded sonar beam cross section.

vocalizations at a high rate (>100 Hz). At the end of both sequences, the vertical array records one lobe. At this point in the pursuit, the target is higher than the bat. The last pattern recorded is consistent with the bat directing its beam upwards, so that only part of one lobe is picked up by the vertical array (also see discussion). This observation suggests that the bat may also track the position of the target in the vertical plane with its sonar beam. Webster and Brazier⁹ performed strobe photography on bats chasing targets and suggested that bats track targets in the vertical as well as horizontal plane by adjusting head direction.

We analyzed a total of 279 vocalizations produced by 2 bats, from 13 flights. Of these 61 (22%) vocalizations had a notch deeper than 3 dB (half-power). A notch was defined as a dip in the sonar beam intensity profile flanked on both sides by intensities greater than or equal to twice the notch intensity. Most of the notched calls appear during the lower repetition rate (approach) phase of insect capture. At pulse rates

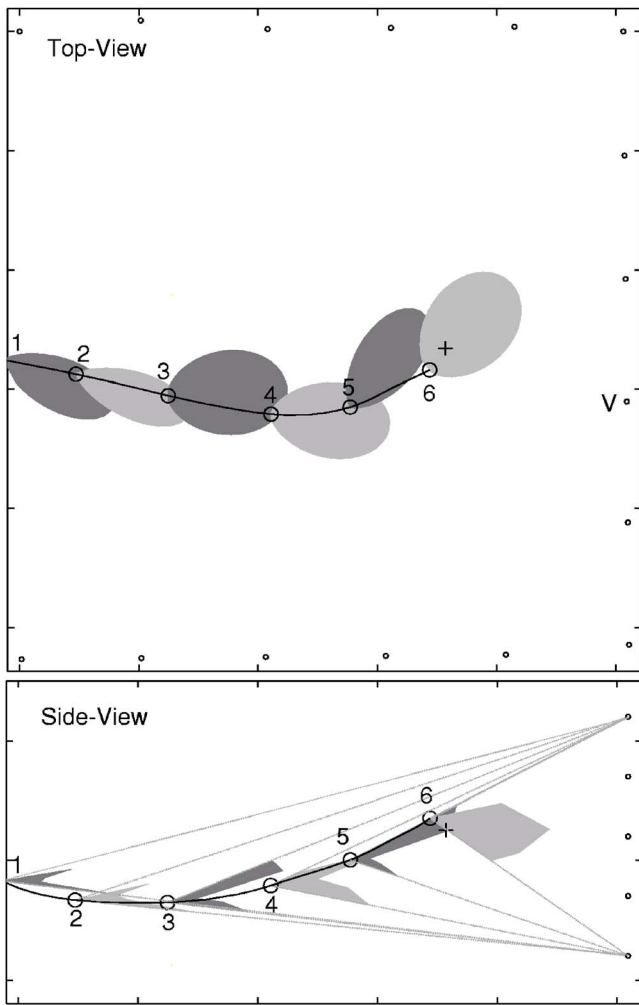


FIG. 4. (Color online) The bat may move its beam up and down to track the target. Some vocalizations from bat #1 during a prey interception. To improve clarity not all vocalizations in the train produced by the bat are shown in the diagram. (top) Horizontal cross sections of the sonar beam and (bottom) vertical cross sections. Dotted lines in the bottom panel show the field-of-view of the vertical array. Vocalization samples 1–4 were produced at a low rate (<20 Hz), whereas 5 and 6 are taken from a train of pulses produced at a high rate (>100 Hz). Samples 1–5 show a prominent notch in the vertical cross-section, whereas the notch is absent in 6. The bat is responding to the target during 5 and 6, and is swooping down on the target during 6. In the side-view the horizontal array is omitted for clarity.

below 50 Hz, 50% (58/117) of the calls had 3 dB deep notches, while at pulse rates above 50 Hz, only 2% (3/162) of the calls had notches 3 dB deep. During the higher pulse repetition rate phases the big brown bat increases the bandwidth of its vocalizations and shifts the peak energy of its calls.¹⁰ This change in energy at different frequencies should change the overall sonar beam pattern shape, assuming a constant aperture size. Our apparatus measures energy from a fixed, narrow band of frequencies centered at 35 kHz and our reconstructed sonar beam shape should not vary simply due to different energy distributions.

B. Control measurements

We controlled for the possibility that a systematic error in either the hardware or the calibration procedure lead to one or more of the vertical array microphones giving a con-

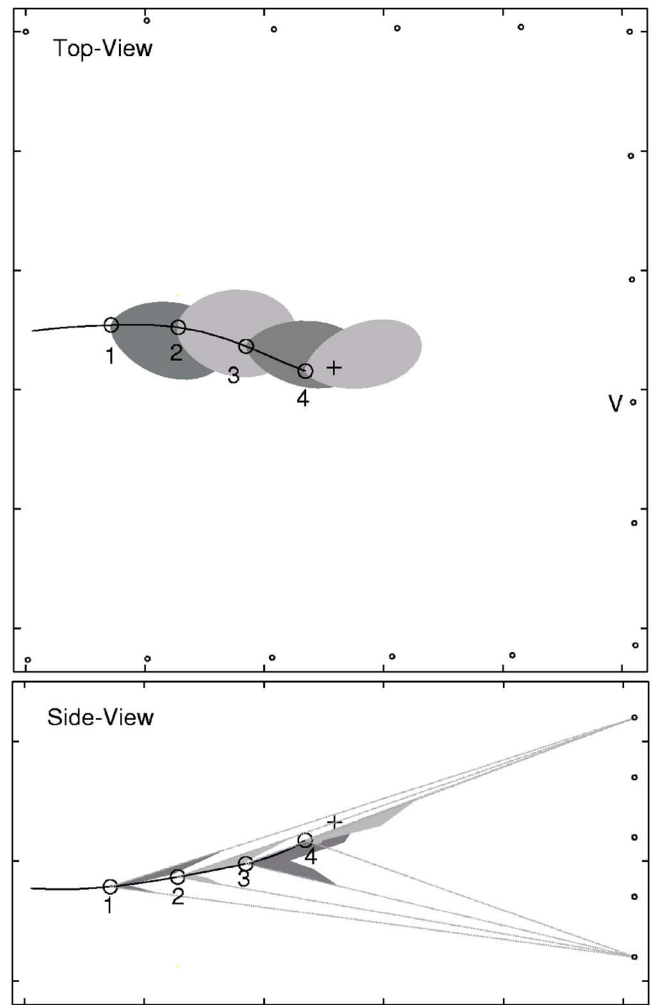


FIG. 5. (Color online) A sequence of beam-patterns from bat #2. Beam pattern samples 1–3 show the notched beam pattern in the vertical array. The bat is producing signals at a low rate (<20 Hz) during patterns 1 and 2, and a high rate (>100 Hz) during patterns 3 and 4. The notch disappears from the recording for sample 4, consistent with the hypothesis the bat is directing its sonar beam up toward the target (black cross).

sistently low or high reading. This erroneous reading could have given rise to a notch artifact on the array recordings. We used an emitter with a circular aperture to ensonify the microphone array. The emitter was stimulated by a frequency sweep of the same bandwidth as bat sonar vocalizations. The recorded beam patterns on both the vertical and horizontal array appeared as single lobes. Instances (like Fig. 3) where only one lobe is recorded on the vertical array also serve as control measurements to eliminate the possibility of systematic errors with the gain.

If the different envelope detectors had band-pass characteristics that were sufficiently mismatched, then a given vocalization, if sufficiently narrow band, may result in a notch artifact being observed on the microphone channel that is outside the bandwidth of the signal. We controlled for this possibility by testing the frequency response of the envelope detector circuits and ensuring that the peak response did not vary from the designed value by more than 2 kHz.

We also inspected the power spectra of full-bandwidth recordings of vocalizations that produced a notch (Fig. 7). We verified that notched beam patterns were observed for

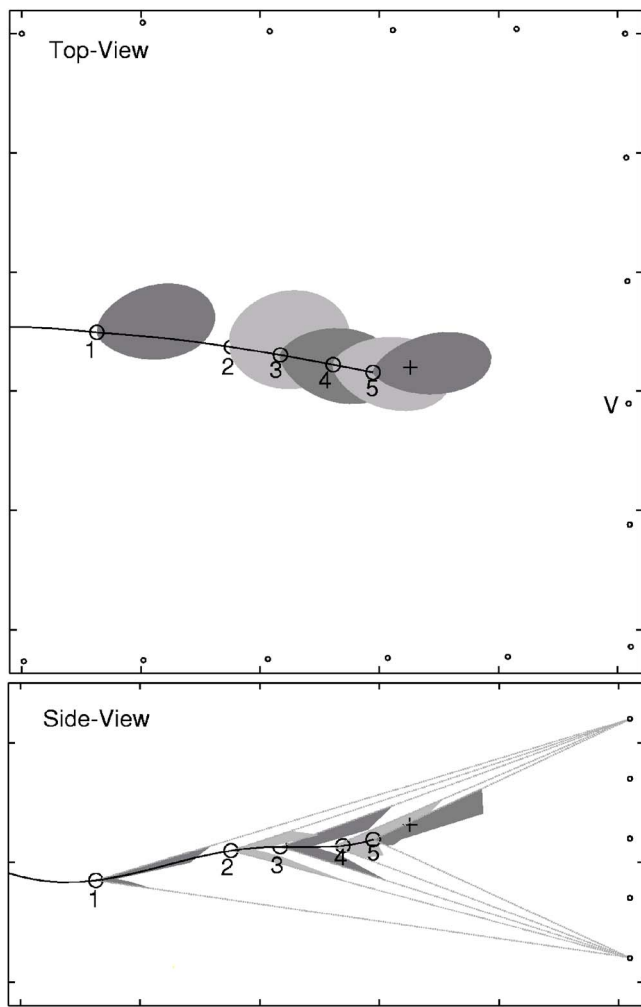


FIG. 6. (Color online) A sequence of beam-patterns from bat #1. As in Fig. 5, for the last beam pattern shown, when the bat is in the terminal stage of insect capture, no notch is recorded by the vertical array, consistent with the hypothesis the bat is directing its sonar beam up toward the target.

vocalization spectra that spanned the entire bandwidth of the filters [Fig. 7(b)]. We also verified that notched beam patterns were observed for vocalizations with power in the lower [Fig. 7(d)], middle [Fig. 7(f)], and higher [Fig. 7(h)] ends of the filter bandwidth. Though the microphone used for broad band ultrasonic recordings did not record the exact signal received at the vertical array (The power spectra of the signals recorded at the array are likely to be different due to the frequency dependent directionality of the sonar beam), the examples show that the notch is unlikely to be due to any mismatch of the filters.

Echoes of the original bat vocalization that reached the microphone array to create destructive interference may also create a notched beam for specific positions of the bat. We, however, observed the vertical notch consistently as the bat traveled in different parts of the room, so we believe it to be unlikely that such a consistent observation from different emitter locations would result from multipath interference.

C. The sonar beam may consist of two lobes arranged dorso-ventrally

We constructed a three-dimensional model of the bat's sonar beam in an effort to replicate the experimental obser-

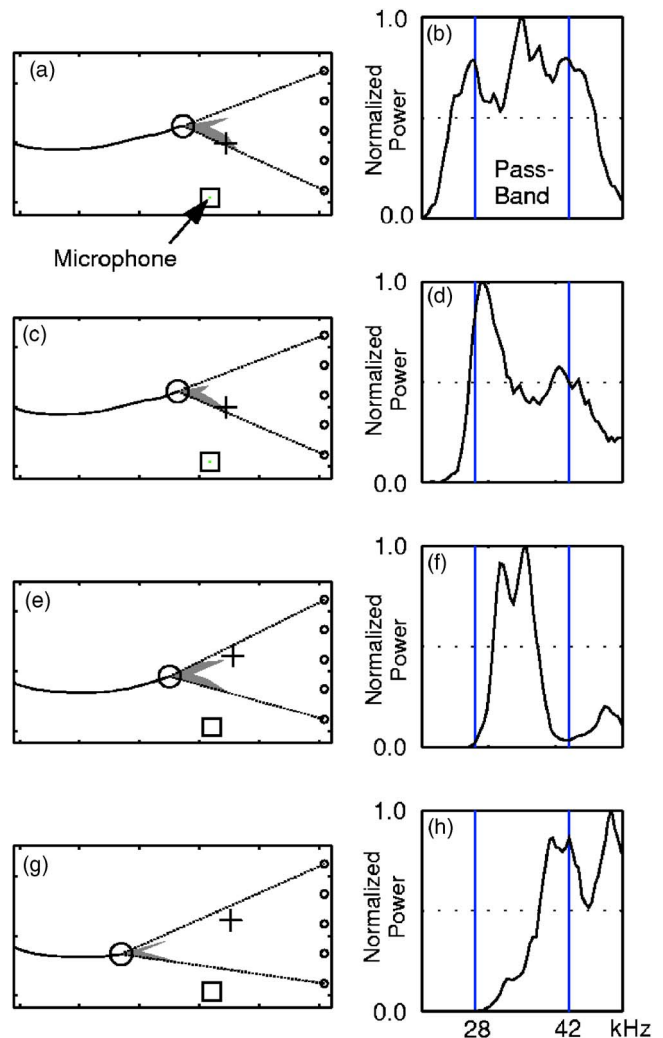


FIG. 7. (Color online) Control for bandwidth. (a,c,e,g) Examples of notched sonar beam patterns emitted by the bat. (b,d,f,h) Power spectra (linear scale) of the corresponding bat calls. The bandwidth of the filter used for the envelope extraction is indicated by vertical lines on the power spectra plots. Power spectra of the call spans: (b) The entire pass-band, (d) the lower end, (f) the middle, and (h) the higher end of the envelope detectors. The black square shows the position of the ultrasonic microphone from which the spectra are generated.

vations via numerical simulations. We found that a model of the sonar beam consisting of two equally sized lobes oriented dorso-ventrally (Fig. 8) could replicate the experimental findings. We modeled each lobe as the emission pattern of a single piston source^{5,11} extended to three dimensions. Let P be the ratio of the sound pressure in the direction θ (azimuth), ϕ (elevation) to the on axis sound pressure. J_1 is the Bessel function of order one. Let k be the wavelength constant, given by $2\pi f/c$, where f is the frequency of the sound source, and $c=340$ m/s, the speed of sound. Let a_h and a_v be the horizontal and vertical diameters of the emitter. A single lobe is modeled as

$$P = \left(2 \frac{J_1\{ka_h \sin(\theta)\}}{ka_h \sin(\theta)} \right) \left(2 \frac{J_1\{ka_v \sin(\phi)\}}{ka_v \sin(\phi)} \right). \quad (1)$$

Our model consists of two lobes oriented at an angle ζ apart in the vertical plane. They are represented by the equations

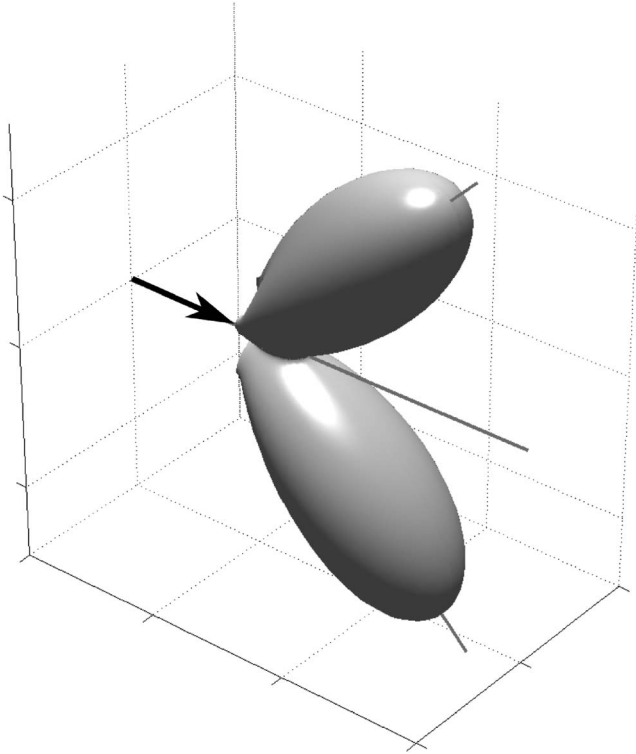


FIG. 8. Phenomenological three-dimensional model of the bat sonar beam. In this model the sonar beam consists of two equal sized lobes, directed ζ degrees apart in the dorso-ventral plane. Each isolated, individual lobe is modeled as the beam from a piston source. The drawing of the bat is merely schematic. No inference is made about the relation between the vertical orientation of the bat's head and the vertical orientation of the hypothesized beam.

$$P_{\text{dorsal}} = \left(2 \frac{J_1\{ka_h \sin(\theta)\}}{ka_h \sin(\theta)} \right) \left(2 \frac{J_1\{ka_v \sin(\phi + \zeta/2)\}}{ka_v \sin(\phi + \zeta/2)} \right) \quad (2)$$

and

$$P_{\text{ventral}} = \left(2 \frac{J_1\{ka_h \sin(\theta)\}}{ka_h \sin(\theta)} \right) \left(2 \frac{J_1\{ka_v \sin(\phi - \zeta/2)\}}{ka_v \sin(\phi - \zeta/2)} \right). \quad (3)$$

The equations for sound pressure for two lobes are combined to obtain the sound intensity of the complete model as

$$H = P_{\text{dorsal}}^2 + P_{\text{ventral}}^2. \quad (4)$$

For the simulations we used $f=35$ kHz, $a_h=4$ mm and $a_v=6$ mm and $\zeta=90^\circ$. Hartley and Suthers⁵ used $a=4.7$ mm from some measurements of the open mouth of *E. fuscus* in their experiment. We arbitrarily chose two values for a_h and a_v that reflect both the asymmetry of the open mouth (which opens wider vertically than horizontally) and the asymmetry of the beam (which is wider horizontally than vertically⁵).

This is a phenomenological model of the beam and we do not propose any specific structure in the bat that creates this pattern. In the vertical plane it approximates the emission pattern obtained by placing two isotropic sources close together,¹² but in the horizontal plane it resembles the emission pattern of a piston source.^{5,11} We define the horizontal

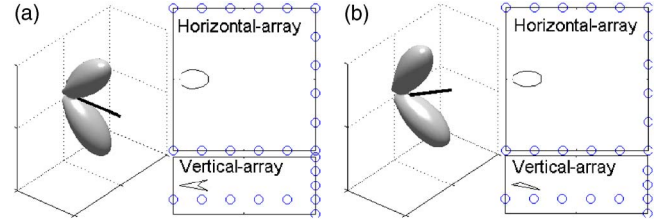


FIG. 9. (Color online) Simulated array measurements of the three-dimensional model. (a) The model beam is oriented at 0° , such that the two lobes are symmetrically above and below the horizontal plane (three-dimensional plot). The simulated measurement from the horizontal array (top) single lobe. The simulated measurement from the vertical array (bottom) notched pattern. (b) The model beam is oriented up by 30° . The simulated measurement from the vertical array only captures part of the lower lobe.

position of the beam (0 degrees) to be such that the two lobes are symmetrically above and below the horizontal plane (as shown in Fig. 8).

Figures 9 and 10 show how this three-dimensional model of the sonar beam is able to replicate the sonar beam cross sections recorded by both the horizontal and vertical segments of the array. Fig. 9 shows simulations of the sonar beam cross sections that would be recorded by the horizontal and vertical segments of the microphone array. In Fig. 9(a) the model beam is oriented at 0° to the horizontal, such that the two lobes are symmetrically above and below the horizontal plane. The simulated measurement from the vertical array shows a prominent notched shape. In Fig. 9(b) the model beam is directed upwards by 30° . The simulated measurement from the vertical array only captures part of the lower lobe.

Figure 10 shows the results of sweeping the model beam from -60° to $+60^\circ$ in the vertical plane. When the beam is oriented within $\pm 5^\circ$ of the horizontal the vertical cross section presents a notched appearance. Larger angles result in only part of one lobe being captured, giving rise to a single lobed appearance. The simulated vertical cross-sections shown in Fig. 10 are comparable to the measured vertical cross sections from flying bats shown in Figs. 2–6.

In our simulation we chose a model with two equal sized lobes directed 90° apart. This model is one representative from a class of multiple lobed models that fit the observa-

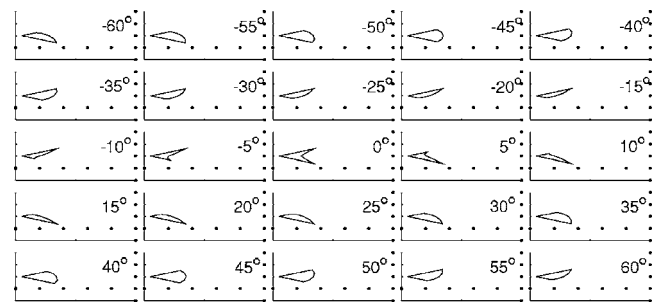


FIG. 10. Effect of sweeping the model beam in the vertical plane. The model sonar beam is swept from -60° to $+60^\circ$ in 5° increments and the cross-section pattern that would be obtained by the vertical-array is simulated. When the sonar beam orientation is within $\pm 5^\circ$ of the horizontal the vertical cross section presents a notched appearance. Larger angles result in only part of one lobe being captured.

tions. It is possible to obtain similar results with two lobes 60° apart and with lobes that differ in intensity by a factor of two. Lobes that are closer together and/or have larger intensity differences fail to produce the prominent notch observed in our experiments. Lobes that are much further apart than 90° would produce a wider “valley” than what we observe in our experiments. The exact beam shape may vary across vocalizations and may depend on the physical characteristics of each individual bat.

IV. DISCUSSION

The sonar beam pattern of an echolocating bat restricts the spatial region sampled by its echolocation system. Study of the shape and direction of the beam during flight may yield insight into how the bat gathers information to guide behavior. We measured horizontal and vertical cross sections of the sonar beams emitted by big brown bats flying in a laboratory flight room. We observed a prominent notch in the vertical cross-section of the beam. Our observations are consistent with the hypothesis that the bat can emit a sonar beam with at least two lobes of comparable size arranged dorso-ventrally.

A previous study on the beam patterns emitted by electrically stimulated anesthetized *E. fuscus*⁵ reported that the sonar beam of these bats consisted of a main lobe and, at frequencies above 60 kHz, a ventral lobe 6 dB less intense than the main lobe. The ventral lobe was directed 30° below the main lobe. The authors concluded that, at higher frequencies, the width of the main lobe could be explained by a piston model. This model, however, broke down at frequencies below 25 kHz. The bat’s beam remained directional at these frequencies, while the model predicted an almost omnidirectional beam. The authors also report that the ventral lobe is not explained by any simple model.

In our study, conducted on flying bats, we also observe a splitting of the sonar beam into a ventral and dorsal lobe. Unlike Hartley’s study, we observe that this split occurs in the 35 kHz frequency range and infer that the lobes are of comparable size and separated in direction by a larger angle (ranging from 60° to 90°). We cannot currently infer the full shape of the sonar beam at multiple frequencies because of sampling limitations in the measurement apparatus. A more detailed analysis of the sonar beam pattern cross-sections, combined with a denser and wider experimental sampling of the sonar beam, is required to determine the exact shape and variability of the sonar beam in flight. Methods that are designed to record the sonar beam pattern in one shot are likely to be more successful than methods that attempt to combine multiple partial snapshots of the sonar beam into a composite whole because the beam pattern can vary greatly from vocalization to vocalization.¹³

The observation that the notch is prominent in about 50% of our samples taken at low call repetition rates suggests that the bat may rotate a multilobed sonar beam up and down in the vertical plane. This observation is also consistent with the hypothesis that the bat is capable of producing sonar beams with or without prominent multiple lobes. Since the

extent of the vertical array was limited we can not eliminate the possibility that the bats adjusted not only the direction, but also the shape of the emitted beam.

The multiple sonar beam lobes of a bat may serve specialized roles in the bat’s echolocation system. We suggest that the forward directed lobe of a bat’s sonar beam gathers information about insects and obstacles, whereas the ventrally directed lobe provides the bat with information about the ground, such as altitude. Such specialization within a sensory system is seen, for example, in the visual system of raptors. Raptors have two foveae that “look” in different directions.¹⁴ The temporal fovea receives information from the frontal (binocular) visual field while the central fovea receives information from the lateral visual field and is connected to the visual streak (a narrow, linear region of tightly packed photoreceptors, providing high-spatial acuity). Observations indicate that a bird tracks its prey with the temporal fovea,¹⁵ whereas the visual streak is hypothesized to help the bird locate the horizon when flying.

Griffin and Buckler¹⁶ addressed the very interesting possibility that migrating birds can sense the type of surface below them from echoes of their calls. From aerial measurements of echoes from different surfaces at frequencies below 6 kHz the authors suggested that birds, assuming their calls radiated omnidirectionally, could hear echoes from the ground, and potentially determine the nature of the terrain they were flying over. Compared to birds, bats typically emit and are sensitive to much higher frequencies (>20 kHz). These higher frequency sounds form more directional beams and are subject to greater attenuation by the atmosphere.¹⁷ In order to sense the ground from high altitudes, bats emitting a single lobed sonar beam may need to scan their head downwards to obtain a detectable echo from the ground. By emitting a sonar beam with two lobes, one directed forward and the other downward, a bat would be able to sense the terrain ahead while scanning for prey in the air without making additional head movements.

ACKNOWLEDGMENTS

The authors thank the Knowles company for donating several samples of their FG3329 electret microphone, which were used to build the array used in the experiments. They thank Chen Chiu for help with some of the experiments. Thanks go to Murat Aytekin for suggesting inclusion of the field-of-view lines in the sequence plots to aid in their interpretation. This work was supported by National Institutes of Health (NIH) P-30 Center Grant No. DC04664, NSF Grant No. IBN0111973 to C.F.M., NIH-NIBIB Grant No. 1 R01 EB004750-01 to T.K.H. and C.F.M. (as part of the NSF/NIH Collaborative Research in Computational Neuroscience Program), AFOSR Grant No. FA95500410130 to T.K.H. and the UMD-CP Psychology Dept. Jack Bartlett fellowship to K.G.

¹D. R. Griffin, *Listening in the Dark* (Yale University Press, New Haven, CT, 1958).

²D. R. Griffin, F. A. Webster, and C. R. Michael, “The echolocation of flying insects by bats,” *Anim. Behav.* **8**, 141–154 (1960).

³A. D. Grinnell and H.-U. Schnitzler, “Directional sensitivity of echolocation in horse-shoe bat, *Rhinolophus ferrumequinum* 2. Behavioral directionality of hearing,” *J. Comp. Physiol. [A]* **116**, 63–76 (1977).

- ⁴H.-U. Schnitzler and A. D. Grinnell, "Directional sensitivity of echolocation in horseshoe bat, *Rhinolophus ferrumequinum* 1. Directionality of sound emission," J. Comp. Physiol. [A] **116**, 51–61 (1977).
- ⁵D. J. Hartley and R. A. Suthers, "The sound emission pattern of the echolocating bat, *Eptesicus fuscus*," J. Acoust. Soc. Am. **85**, 1348–1351 (1989).
- ⁶G. M. Hope and K. P. Bhatnagar, "Electrical response of bat retina to spectral stimulation: comparison of four microchiropteran species," *Experientia* **35**, 1189–1191 (1979).
- ⁷K. Ghose and C. F. Moss, "The sonar beam pattern of a flying bat as it tracks tethered insects," J. Acoust. Soc. Am. **114**, 1120–1131 (2003).
- ⁸K. Ghose and C. F. Moss, "Steering by hearing: A bat's acoustic gaze is linked to its flight motor output by a delayed, adaptive linear law," J. Neurosci. **26**, 1704–1710 (2006).
- ⁹F. A. Webster and O. G. Brazier, "Experimental studies on target detection, evaluation and interception by echolocating bats," Technical Report, Aerospace Medical Research Labs, Wright-Patterson (1965).
- ¹⁰A. F. Surlykke and C. F. Moss, "Echolocation behavior of big brown bats, *Eptesicus fuscus*, in the field and the laboratory," J. Acoust. Soc. Am. **108**, 2419–2429 (2000).
- ¹¹L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics*, 2nd edition (Wiley, New York, 1962).
- ¹²G. K. Strother and M. Mogus, "Acoustical beam patterns for bats: Some theoretical considerations," J. Acoust. Soc. Am. **48**, 1430–1432 (1970).
- ¹³M. Aytekin, "Sound localization by echolocating bats," Ph.D. thesis, University of Maryland, College Park, MD 2007.
- ¹⁴G. R. Martin, "Form and function in birds," *Eye* **3**, 311–373 (1988).
- ¹⁵M. F. Land, "The roles of head movements in the search and capture strategy of a tern (Aves, Laridae)," J. Comp. Physiol., A **184**, 265–272 (1999).
- ¹⁶D. R. Griffin and E. R. Buckler, *Echolocation of Extended Surfaces* (Springer, Berlin, 1978), 201–208.
- ¹⁷B. D. Lawrence and J. A. Simmons, "Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocation by bats," J. Acoust. Soc. Am. **71**, 585–590 (1982).

Short- and long-term changes in right whale calling behavior: The potential effects of noise on acoustic communication

Susan E. Parks^{a)} and C. W. Clark

Bioacoustics Research Program, Cornell Laboratory of Ornithology, 159 Sapsucker Woods Road, Ithaca, New York 14850

P. L. Tyack

Biology Department, MS #50, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 20 February 2007; revised 31 August 2007; accepted 25 September 2007)

The impact of anthropogenic noise on marine mammals has been an area of increasing concern over the past two decades. Most low-frequency anthropogenic noise in the ocean comes from commercial shipping which has contributed to an increase in ocean background noise over the past 150 years. The long-term impacts of these changes on marine mammals are not well understood. This paper describes both short- and long-term behavioral changes in calls produced by the endangered North Atlantic right whale (*Eubalaena glacialis*) and South Atlantic right whale (*Eubalaena australis*) in the presence of increased low-frequency noise. Right whales produce calls with a higher average fundamental frequency and they call at a lower rate in high noise conditions, possibly in response to masking from low-frequency noise. The long-term changes have occurred within the known lifespan of individual whales, indicating that a behavioral change, rather than selective pressure, has resulted in the observed differences. This study provides evidence of a behavioral change in sound production of right whales that is correlated with increased noise levels and indicates that right whales may shift call frequency to compensate for increased band-limited background noise.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2799904]

PACS number(s): 43.80.Nd, 43.80.Ka, 43.66.Dc [WWA]

Pages: 3725–3731

I. INTRODUCTION

Marine mammals rely upon sound for communicating, navigating, and foraging and the potential impact of anthropogenic sources of sound on marine mammals has received increasing attention in the past two decades (Richardson *et al.*, 1995; N.R.C., 2005). Most human activities in the ocean generate sound either intentionally (e.g., sonar or seismic exploration) or inadvertently as noise (e.g., ship engines and propeller cavitation). Low-frequency sound can propagate over great distances in the ocean, and intense low-frequency sources can be detected halfway around the globe (Munk *et al.*, 1994). There has been increased scrutiny of the impacts of noise on marine mammal species. The acute impacts of high amplitude sound sources have been a major focus of marine mammal acoustics research (Miller *et al.*, 2000; Fristrup *et al.*, 2003; Jepson, 2003). Several mass stranding incidents have been linked to military exercises using mid-frequency active naval sonars (Frantziis, 1998; Cox *et al.*, 2006). Other studies have investigated short-term changes in behavior related to increased background noise levels from vessels (Lesage *et al.*, 1999; Buckstaff, 2004; Foote *et al.*, 2004; Scheifele *et al.*, 2005). Further research is needed to determine if there are long-term effects from a combination of these sources of noise. The goal of this study is to investigate if long-term changes in ambient noise levels have led to changes in the sound production behavior of North and

South Atlantic right whales (*Eubalaena glacialis* and *E. australis*, respectively).

Payne and Webb (1971) were the first to raise concern over the effects of anthropogenic sound on marine mammals and identified noise from commercial shipping as the major source of noise in the 5–200 Hz band. Since that initial alarm, shipping has been increasing steadily as demand for transport has grown. The distribution of global trade is uneven, with significantly more ship traffic in the northern hemisphere than in the southern hemisphere (Wang and Corbett, 2005). A comparison of the tonnage of seaborne trade between countries in the North and South Atlantic indicates that from 1970 to 2003 trade has increased in both regions (North Atlantic: 2.5×10^9 – 5.5×10^9 t, South Atlantic: 8.0×10^8 – 1.3×10^9 t) but that over this period North Atlantic totals have consistently been 3–4 times greater than those in the South Atlantic (U.N., 2004). Three studies have demonstrated an increase in noise level below 100 Hz for the Northeast Pacific from the 1960s to the present, where that increase is coincident with the documented increase in shipping traffic (Ross, 1974; Andrew *et al.*, 2002; McDonald *et al.*, 2006). We assume that a similar increase has occurred in the North Atlantic where shipping has increased by similar levels.

Distant shipping has been suggested as the most likely source for the observed increases in low-frequency noise. Although close passage of nearby ships can elevate noise at frequencies above 1 kHz (Aguilar Soto *et al.*, 2006), energy from distant shipping is primarily below 100 Hz (Wenz, 1962). In all three of these studies reporting on deep-water

^{a)} Author to whom correspondence should be addressed. Current address: The Pennsylvania State University, Applied Research Laboratory, P.O. Box 30, State College, PA 16804-0030. Electronic mail: sep20@psu.edu

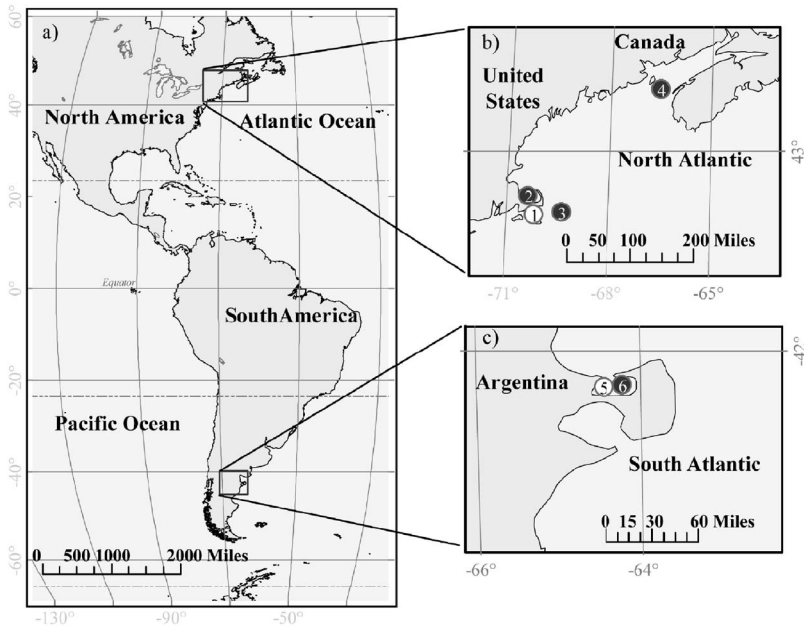


FIG. 1. Map showing locations where right whale recordings were made for this study. The (a) Black squares mark the recording areas. (b) North Atlantic right whale area off the east coast of the United States. 1—location of 1956 recordings and 2–4—locations of 2000–2004 recordings. (c) South Atlantic right whale recording area with 5 and 6 marking the historic and contemporary recordings respectively in Golfo San Jose.

noise measurements, where distant shipping noise predominates, ambient noise levels have increased more for frequencies between 20 and 100 Hz than for higher frequencies (Ross, 1974; Andrew *et al.*, 2002; McDonald *et al.*, 2006). In shallow coastal waters, where many species of marine mammals are located, low-frequency noise levels are dominated by local ship traffic patterns (Zakarauskas *et al.*, 1990). A study of noise levels in the North Atlantic on the Canadian continental shelf noted noise peaks generally around 80 Hz (Zakarauskas *et al.*, 1990). The study noted an increase of up to 15 dB re $1 \mu\text{Pa}^2/\text{Hz}$ for frequencies between 30 and 80 Hz when compared to earlier deeper water measurements on the Grand Banks (Piggott, 1964), but suggested that propagation differences might explain the Zakarauskas *et al.* (1990) observation.

This paper investigates whether increased noise levels have resulted in any behavioral changes in sound production by North and South Atlantic right whales. Right whales are a long-lived species of baleen whale, with some individuals known to live at least 65 years (Hamilton *et al.*, 1998). Right whale populations worldwide were reduced to low levels by whaling. Southern right whales (*Eubalaena australis*) have shown remarkable recovery since protection from whaling, but North Atlantic right whales remain on the brink of extinction (Kraus *et al.*, 2005). Right whales produce sounds to communicate, using variable tonal calls in social interactions and a stereotyped tonal upswEEP as a contact call (Clark, 1982; Parks and Tyack, 2005).

As a population declines, it may become increasingly difficult for individuals to locate each other for social interaction and mating due to dilution effects. If ambient noise levels increase, the range over which individuals can locate one another will be further reduced unless the whales can modify their calling behavior to improve the probability of detection by another whale. Five potential mechanisms that animals can use to increase the detectability of their calls include (1) increasing the intensity of their calls (Brumm, 2004; Scheifele *et al.*, 2005), (2) increasing the rate of call-

ing (Penna *et al.*, 2005), (3) increasing the duration of their calls (Foote *et al.*, 2004), (4) shifting the frequencies (Hz) of their calls to a frequency band with lower noise levels (Lesage *et al.*, 1999; Slabbekoorn and Peet, 2003), or (5) waiting until the noise decreases before calling (Lesage *et al.*, 1999; Sun and Narins, 2005).

This paper uses a combination of three different analyses to investigate whether right whales change their calling behavior in times of increased background noise. The first analysis investigates short-term changes in response to transient increases in noise from nearby vessels using recordings of right whale surface active groups from the Bay of Fundy, Canada (Parks and Tyack, 2005). The second and third analyses rely on larger scale comparisons. The second analysis compares calls recorded from North and South Atlantic right whales. Shipping traffic is lower in the South Atlantic than in the North Atlantic, therefore Southern right whales should be calling in a quieter ambient noise environment than North Atlantic right whales. Shipping levels, and presumably noise levels, have increased significantly over the past 40 years in both hemispheres. The third analysis compares calls recorded from 1956 (North Atlantic) and 1977 (South Atlantic) with contemporary recordings (2000–2004) for another comparison of low noise (historical recordings) with high noise (contemporary recordings) situations. These three comparisons allow us to test whether there are changes in calling behavior that are consistent with predicted mechanisms to compensate for chronic changes in ocean noise.

II. MATERIAL AND METHODS

A. Comparison 1—Sound production in Surface Active Groups

1. Recordings

Acoustic recordings were made of right whale surface active groups (SAGs) in the Bay of Fundy, Canada (Fig. 1) from 1999 to 2002 using a Hi-Tech HTI-94-SSQ hydrophone and a TASCAM DA-PI DAT recorder [system flat (± 1 dB re

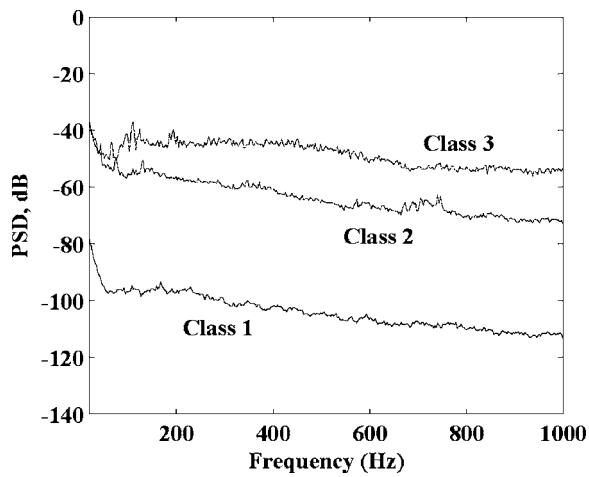


FIG. 2. Illustration of the relative noise power spectral density estimate using the Bartlett method for the three classes of noise levels [Class 1 (low)-Class 3 (high)] used in Comparison 1 (16 384 pt. FFT, no overlap, $F_s = 44.1$ kHz).

1 μ Pa) frequency response, 20 Hz–20 kHz, 44.1 kHz sampling rate] (Parks and Tyack, 2005). These recordings consist of scream calls produced only by the focal female whale and upcalls produced by other whales in the group (Parks and Tyack, 2005). Noise levels for these recordings were defined by presence or absence of vessel noise in the recording. A quiet recording (Class 1) contained no audible vessel noise. A midrange recording (Class 2) had audible vessel noise from at least one distant vessels (e.g., tanker at >5 nm), but the level of noise received at the whales was moderate and variable. This category was primarily used to separate low noise, quiet, recordings from high-noise recordings. A high noise (Class 3) recording contained relatively continuous noise from at least one vessel less than 1 nm from the whales and included small research or whale watch vessels and larger commercial ships (Fig. 2).

2. Analysis

Individual scream calls and upcalls were manually selected from longer surface active group (SAG) recordings using XBAT (<http://www.xbat.org>), a custom designed MATLAB (Mathworks, Inc.) program developed by the Cornell Bioacoustics Research Program for browsing acoustic files. Scream calls are produced by the single focal female in SAGs (Parks and Tyack, 2005), making it possible to measure individual call rates and responses to noise in the frequency and time domains. Timing of scream call production was measured as the interval between the start time of one scream and the start time of the next scream, including all detectable scream calls. The minimum frequency and duration of the scream calls were measured by visual inspection of the spectrograms [2048 point fast Fourier transform (FFT) and frame length, 50% overlap, Hanning window, $F_s = 44.1$ kHz] [Fig. 3(a)]. All upcalls were down sampled to 2000 Hz (the lowest original sampling rate for upcall datasets) before measurements were made. Start frequency, end frequency, and duration were measured from spectrograms (512 pt. FFT and frame length, 50% window overlap, Hanning window, $F_s = 2$ kHz) using a custom designed MAT-

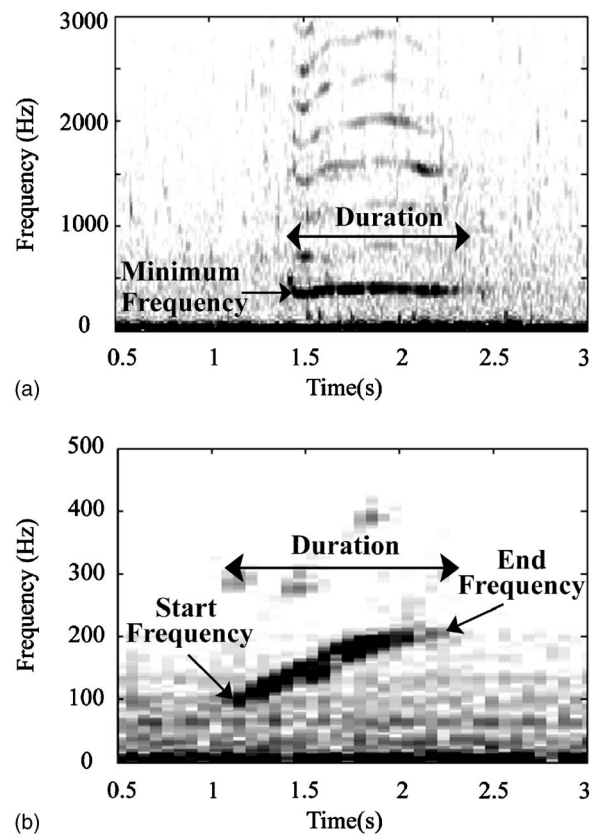


FIG. 3. Example of (a) a scream call from a SAG and (b) an upcall with measurements used in these experiments are indicated.

LAB 6.5 script (Mathworks, Inc.) [Fig. 3(b)]. Intercall intervals were not measured because multiple individuals in the SAG may have produced the upcalls. The source level of the calls was not measured due to lack of data on the range to the whale producing the calls.

Mixed models were used to determine if noise level (Classes 1–3) had an effect on the minimum frequency of the call, the call duration, and the inter-call interval of scream calls in SAGs (SAS 9.1, SAS Institute Inc., Cary, NC). We used least squares means and multiple contrasts for pair-wise comparisons (Littell *et al.*, 2006). JMP 6.0.2 (SAS Institute Inc.) was used to compare start frequency of upcalls in low and high noise conditions in SAGs using a heteroskedastic t-test.

B. Comparison 2—North Atlantic right whale versus southern right whale upcalls

1. Recordings

a. North Atlantic right whales. Recordings for this analysis include upcalls recorded from SAGs in the Bay of Fundy, Canada (1999–2002), additional upcalls recorded from SAGs in the Bay of Fundy in 2004 with the same equipment, and upcalls recorded on bottom-mounted recording devices referred to as “pop-ups” (Clark *et al.*, 2002) deployed in Cape Cod Bay, MA and the Great South Channel, east of Massachusetts from 2000 to 2002 (system flat frequency response 20–400 Hz, 2 kHz sampling rate) (Fig. 1).

b. Southern right whales. South Atlantic (SA) (2000) recordings were made in Golfo San Jose, Argentina in the spring of 2000 by A. Purgue and R. MacCurdy from the

TABLE I. Summary of measurements of scream calls from North Atlantic right whale surface active groups. Least squares means (LSM) \pm the standard error (SE) values for each noise level [low(Class 1), medium (Class 2), and high (Class 3)], and the contrast values for each comparison.

Call parameter/noise level	Low (Class 1)	Medium (Class 2)	High (Class 3)	Low (1) vs. Medium (2)	Low (1) vs. High (3)	Medium (2) vs. High (3)
Minimum frequency (Hz)	381.4 \pm 16.50	390.3 \pm 15.14	422.4 \pm 15.55	$F=0.8$ $p=0.37$	$F=9.0$ $p=0.003$	$F=7.2$ $p=0.008$
Call duration (s)	1.18 \pm 0.08	1.11 \pm 0.07	1.22 \pm 0.08	$F=3.58$ $p=0.06$	$F=0.41$ $p=0.52$	$F=5.09$ $p=0.02$
Intercall Intervals (s)	17.9 \pm 5.06	18.5 \pm 4.55	28.1 \pm 4.63	$F=0.03$ $p=0.86$	$F=5.39$ $p=0.02$	$F=6.35$ $p=0.01$

Cornell Bioacoustics Research Program using a TEAC DAP20 DAT recorder and custom hydrophones (system flat frequency response 50 Hz–5 kHz, 44.1 kHz sampling rate).

2. Analysis

Upcalls were selected from longer sequences of recordings using XBAT (NA 1999–2004, SA 2000). All upcalls were down-sampled to 2000 Hz (the lowest original sampling rate for upcall datasets) before measurements were made. Start frequency, end frequency, and duration were measured from spectrograms (512 pt. FFT and frame length, 50% window overlap, Hanning window, $F_s=2$ kHz) using a custom designed MATLAB 6.5 script (Mathworks, Inc.) [Fig. 3(b)].

C. Comparison 3—Historic versus contemporary upcall recordings

1. Recordings

a. North Atlantic right whales. Historic sound data consist of 1956 recordings made by William Schevill from the Woods Hole Oceanographic Institution during four recording sessions in April, 1956 in Vineyard Sound, MA (Fig. 1) using a Magnecorder (Model Pt6J) and an AX-58 hydrophone (Brush Development Corporation) (system flat frequency response 30 Hz–10 kHz). The recordings were obtained with permission from the library of William Watkins at WHOI. Contemporary recordings consist of the dataset described for Comparison 2.

b. Southern right whales. South Atlantic recordings were made in Golfo San Jose, Argentina in 1977 using a Nagra IV-S stereo tape recorder and a custom built hydrophone array (system flat frequency response 50 Hz–5 kHz) (Clark, 1982) (Fig. 1). Contemporary recordings consist of the dataset described for Comparison 2.

2. Analysis

Historic upcalls used for this analysis were extracted from longer sequences of recordings using XBAT (SA 1977) or were obtained as individual clips (SA 1956). Upcalls were measured with the same methods described for the analysis in Comparison 2. JMP 6.0.2 (SAS Institute Inc.) was used to compare start frequency, end frequency, and duration of upcalls between species and between historic and contemporary recordings in both species using a two-way ANOVA.

III. RESULTS

A. Comparison 1—Sound production in surface active groups

A total of 3,486 calls recorded from 32 surface active groups (SAGs) in the Bay of Fundy from 1999 to 2004 were analyzed to determine if calling parameters changed in response to noise. Of these, 3,213 had signal-to-noise levels adequate to obtain measurements of minimum frequency, start frequency, and duration. The calls were broken down into two categories, scream calls ($n=2795$) and upcalls ($n=418$). The scream calls were produced by a single animal in the group, the focal female, and were the focus of the surface active group study (Parks and Tyack, 2005). The identity, age and reproductive status of the focal female was known for 20 of 32 groups. The upcalls were produced by an unknown number of individuals within each group. The duration of the recordings varied from 2 to 134 min ($\bar{x}=28\pm26.6$ min). The number of scream calls from each group ranged from 9 to 616 ($\bar{x}=110\pm119.9$), and the number of upcalls ranged from 0 to 67 ($\bar{x}=13.34\pm17.42$). The numbers of scream calls in each of the three noise classes were Class 1: 623, Class 2: 1157, Class 3: 1015. The numbers of upcalls in each of the three noise classes were Class 1: 25, Class 2: 244, Class 3: 149.

Mixed models were used to determine if noise level (Classes 1–3) had an effect on the minimum frequency of the scream calls, the call duration, and the intercall interval in SAGs. The mixed model analysis blocked for focal individual identification and recording session to account for recording session variability. The initial mixed models included two main factors (age and noise level) and the interaction term. The mixed models were refined by stepwise elimination of nonsignificant factors [i.e., age ($F_{2,1751}=0.00$, $p=0.9470$) and then the interaction term (age \times noise $F_{2,1751}=0.91$, $p=0.4016$) for minimum frequency]. Thus, age of the vocalizing whale was found to have no significant effect on any of the measured variables. Noise level affected the minimum frequency (mixed model, $F_{2,1751}=4.79$, $p=0.0085$), the call duration (mixed model, $F_{2,1751}=3.57$, $p=0.0283$), and the intercall interval (mixed model, $F_{2,2120}=3.54$, $p=0.0291$) of scream calls (Table I). Minimum frequency was significantly lower in lower noise (Class 1) (Table I). Call duration was significantly shorter in midlevel (Class 2) noise than in high noise situations (Class 3), but

TABLE II. Summary of measurements of upcalls from the North Atlantic (NA) and southern (SA) right whales. The table shows the number of calls measured (n), and the mean \pm the standard deviation for the call duration in seconds, start frequency, end frequency, and bandwidth in Hz.

Species	Year	n	Duration (s)	Start Frequency (Hz)	End Frequency (Hz)	Bandwidth (Hz)
NA	1956	19	0.74 \pm 0.26	70 \pm 16	171 \pm 32	105 \pm 34
NA	2000–2004	929	0.87 \pm 0.27	101 \pm 22	195 \pm 38	100 \pm 37
SA	1977	846	0.84 \pm 0.25	69 \pm 12	137 \pm 25	78 \pm 23
SA	2000	78	0.82 \pm 0.23	78 \pm 15	156 \pm 29	86 \pm 23

duration was not significantly different between low (Class 1) and high noise recordings (Class 3) (Table I). Intercall intervals were significantly shorter in lower noise (Class 1) (Table I).

In SAGs the start frequency of upcalls recorded in SAGs was significantly higher in high noise (Class 3) than in low noise (Class 1) (t-test, $t=4.06$, $df=36.9$, $p=0.0002$). There was no significant difference in upcall duration between high (Class 3) and low (Class 1) noise situations (t-test, $t=1.33$, $df=40.3$, $p=0.190$). Inter-upcall intervals were not measured because the number of different individuals producing the upcalls was unknown.

B. Comparison 2—North Atlantic right whale versus southern right whale upcalls and Comparison 3—Historic versus contemporary upcall recordings

A total of 948 upcalls from North Atlantic (NA) right whales and 924 upcalls from southern right whales (SA) were analyzed. Measurements included the start frequency, end frequency, bandwidth, and duration for all upcalls (Table II).

1. Start frequency

The two-way ANOVA found a main effect of species, $F(1, 1872)=28.8$, $p\leq 0.0001$, indicating that North Atlantic right whales call at a higher start frequency than southern right whales (Fig. 4). There was also a main effect of time, $F(1, 1872)=77.5$, $p\leq 0.0001$. This effect showed that start frequency has increased over time for both species. There was an interaction between species and time, $F(1, 1872)=24.8$, $p\leq 0.0001$, that showed that there was a greater change in frequency through time in the North Atlantic right whale and a greater difference in frequency between the two species from recent recordings than historic recordings.

2. End frequency

The two-way ANOVA for end frequency found a main effect of species, $F(1, 1872)=75.7$, $p\leq 0.0001$, indicating that North Atlantic right whales calls have a higher end frequency than southern right whales. There was also a main effect of time, $F(1, 1872)=26.0$, $p\leq 0.0001$. This effect showed that end frequency has increased over time for both species. There was no interaction between species and time, $F(1, 1872)=0.39$, $p\leq 0.53$.

3. Bandwidth

The two-way ANOVA for call bandwidth found a main effect of species, $F(1, 1872)=26.5$, $p\leq 0.0001$, indicating that North Atlantic right whale upcalls have a larger bandwidth than southern right whales. There was no effect of time, $F(1, 1872)=0.18$, $p\leq 0.67$ and no interaction between species and time, $F(1, 1872)=2.73$, $p\leq 0.09$ for this parameter.

4. Duration

The two-way ANOVA for duration found no effect of species, $F(1, 1872)=0.63$, $p\leq 0.43$ or time, $F(1, 1872)=2.35$, $p\leq 0.13$. However, there was an interaction between species and time, $F(1, 1872)=4.63$, $p\leq 0.03$ because North

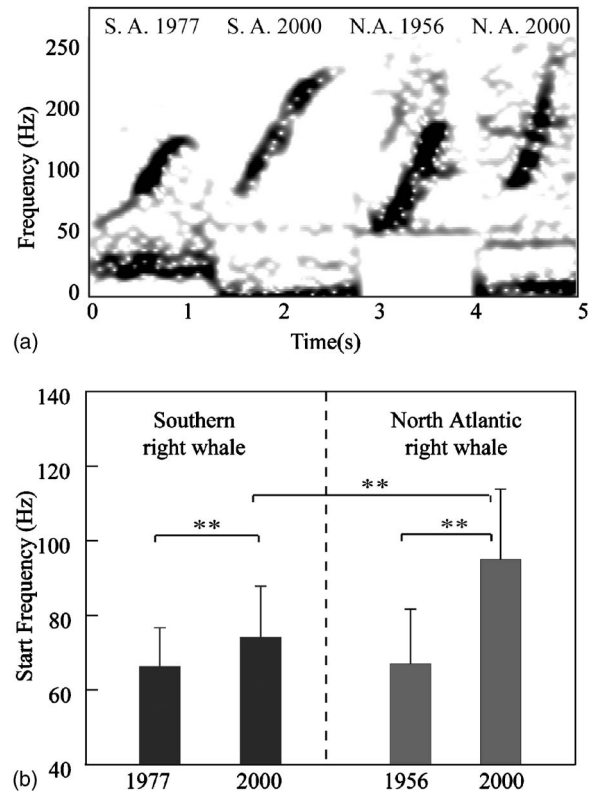


FIG. 4. (a) Examples of representative right whale contact calls from the South Atlantic in 1977 and 2000 and the North Atlantic in 1956 and 2000. Notice the upwards shift in frequency for the calls from contemporary recordings. (b) Summary of start frequency differences between species and frequency differences over time for both species. (**—two-way ANOVA, $p<0.001$).

Atlantic right whale call duration increased slightly through time, whereas southern right whale call duration decreased through time.

IV. DISCUSSION AND CONCLUSIONS

This study was designed to investigate how right whales change their calling behavior in response to increased noise in their environment. The first part of this study investigated changes in SAGs where a single individual produces the majority of the tonal sounds. In this situation it was possible to investigate individual responses to noise relating to the minimum frequency, call duration, and call rate. Our results indicate that there is a significant difference in each of these parameters under high noise conditions. The observed changes include an increase in minimum frequency and a decrease in call rate under high noise conditions. While changes in duration were found to be significant between noise level categories, there was no clear trend with increasing noise. It is interesting to note that although increasing the frequency (Hz) could improve detectability of a call in band limited low-frequency noise, a lower call rate may not. It is possible that the whales decrease their call rate due to increased effort, such as an increase in call amplitude, when calling in higher noise. Alternatively, whales may be waiting to call until there is a temporary reduction in the background noise level which would be a reasonable strategy for whales to adopt in the presence of continuous vessel noise. This dataset did not allow for measurement of call amplitude as a function of short-term variability in noise level and further studies should be conducted to investigate this aspect of calling behavior.

The results from investigating upcalls in SAGs followed a similar pattern. The starting frequencies of the upcalls were higher in periods of high noise. In these cases, it was not possible to determine which individual produced each call; therefore it could not be determined if the observed changes were the result of a single individual changing its calls or if a different individual was calling. This also prevented us from measuring the intercall interval.

To further test the hypothesis that right whales call at higher frequencies in increased noise, we measured recordings of upcalls in two other paired noise situations. We first compared the upcalls of right whales in the North Atlantic, an area with significantly higher ship traffic than the South Atlantic (U.N., 2004), with upcalls of right whales from the South Atlantic. As predicted, North Atlantic right whales produced higher frequency calls on average than Southern right whales. The second comparison measured changes in calling behavior over time for both species. Studies of ambient noise in the ocean have documented an increase in ambient noise in the 40–100 Hz frequency range (Ross, 1974; McDonald *et al.*, 2006), the same frequency range of the start frequency of right whale upcalls. The comparison of recordings from 1956 in the North Atlantic and from 1999 to 2004 showed a significant increase in the start and end frequency of the calls in the contemporary, high noise recordings. The calls from 2000 were approximately 2/3 octave higher than the average values for calls recorded 1956. Given the small sample size collected over four different days, it is

unclear how representative the 1956 data are of the entire population at that time, so these results must be treated with some caution. However, a similar increase was found between 1977 and 2000 in the South Atlantic, with the contemporary calls being higher in frequency than the historic calls.

Alternative hypotheses exist for each individual piece of evidence. These include individual variation for the observed shift of scream calls in SAGs, inherited species differences from random genetic drift or selective pressure for the between-species differences, and a change in population demographics (i.e., with greater proportion of younger animals producing a greater proportion of higher frequency calls) for the observed changes over 23–48 years. In the case of scream calls in SAGs, there are individual differences in calling behavior. However, there were no clear differences between calls produced by juveniles or adults (i.e., age was a nonsignificant factor in the mixed model).

Genetic differences could have caused the observed divergence in the minimum frequencies between right whales of the North and South Atlantic, which have been reproductively isolated for a significant period of time, long enough to be considered separate species based on genetic studies (Gaines *et al.*, 2005). During speciation for North and South Atlantic right whales, there may have been a selection pressure for distinctive reproductive advertisement displays to further isolate the populations. However, we specifically selected the upcall, or contact call, of both right whale species for the comparison, under the assumption that individual recognition signals may be subject to less sexual selection pressure than calls, such as screams, thought to be associated with mating behavior.

Finally, though it is possible that the demographics of both populations have changed over time, with an increasing number of young animals as the populations have been protected from whaling for the past 70 years, there was no evidence of differences in calls based on age in the surface active group study. It is therefore unlikely that a demographic shift could explain the overall observed frequency shift within and across each species.

There has been no direct test of right whale calling behavior in response to controlled noise exposure experiments, which would be necessary to confirm that these changes are caused by a response to increased noise. Further, the recordings used here were not suitable to determine if right whales increase the amplitude of their calls in high noise levels, which has been shown for other marine species exposed to masking signals (Scheifele *et al.*, 2005). We recommend that calibrated recordings be made, particularly in multiple locations in the Southern hemisphere, for a more robust comparison between species. Controlled exposure experiments should be carried out to determine what intensity and duration of noise are required to cause the changes in frequency, intercall interval, and duration.

Despite the potential alternate explanations and the need for further study, in all three comparisons presented here, right whales produced higher frequency calls in the higher noise situation. The most parsimonious explanation is that these observed changes reflect a behavioral response to compensate for increased noise. Whether the whales receive a

significant improvement in signal to noise ratio by this shift in frequency has yet to be demonstrated. Previous studies in other species have demonstrated frequency shifts in response to noise in other marine and terrestrial species (Lesage *et al.*, 1999; Slabbekoorn and Peet, 2003; Slabbekoorn and den Boer-Visser, 2006; Gillam *et al.*, 2007). It is likely that a frequency shift is a common response to band-limited noise, a response that has remained underappreciated and undetected in baleen whales until now. We expect that other species will demonstrate similar shifts of frequency between low and high noise environments. This is a significant result because it indicates there could be a species-wide behavioral change in response to gradual but chronic increases in ambient noise over time. Given the observed behavioral response, the impacts on right whale communication need to be considered to determine what, if any, role increased noise may have on limiting reproduction and recovery of the species.

ACKNOWLEDGMENTS

The authors dedicate this paper to the memory of William Watkins, a pioneer of marine mammal bioacoustics, who provided access to the 1956 Schevill recordings from the North Atlantic. We gratefully acknowledge the contribution of all who assisted with the collection of the recordings used in this paper. M.A. Daher and K. Amaral from the Woods Hole Oceanographic Institution provided the 1956 clips from the Watkins collection. The Bay of Fundy recordings were collected under U.S. Permit No. NMFS No. 1014. Contemporary recordings (1999–2004) were collected with approval of the Cornell University and Woods Hole Oceanographic Institution Animal Care and Use Committees. Funding for the recordings in Cape Cod Bay and the Great South Channel was provided by the International Fund for Animal Welfare (2000 and 2001) and NOAA. Elvira Rawson Paz provided funds to support the 2000 recordings in Argentina. T. J. Balsby, I. Urazghildiiev, and two anonymous reviewers provided helpful suggestions that improved this manuscript. Funding for this analysis was provided by NOAA Grant No. NA04NMF4720413.

Aguilar 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.

Andrew, R. K., Howe, B. M., and Mercer, J. A. (2002). "Ocean ambient sound: Comparing the 1960s with the 1990s for a receiver off the California coast," *ARLO* **3**, 65–70.

Brumm, H. (2004). "The impact of environmental noise on song amplitude in a territorial bird," *J. Anim. Ecol.* **73**, 434–440.

Buckstaff, K. C. (2004). "Effects of watercraft noise on the acoustic behavior of bottlenose dolphins, *Tursiops truncatus*, in Sarasota Bay, Florida," *Marine Mammal Sci.* **20**, 709–725.

Clark, C. W. (1982). "The acoustic repertoire of the southern right whale, a quantitative analysis," *Anim. Behav.* **30**, 1060–1071.

Clark, C. W., Borsani, J. F., and Notarbartolo-di-Sciara, G. (2002). "Vocal activity of fin whales, *Balaenoptera physalus*, in the Ligurian Sea," *Marine Mammal Sci.* **18**, 281–285.

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. L., Wartzok, D., Gisiner, R., Mead, J., and Benner, L. (2006). "Understanding the impacts of anthropo-

genic sound on beaked whales," *J. Cetacean Res. Manage.* **7**, 177–187.

Foote, A. D., Osborne, R. W., and Hoelzel, A. R. (2004). "Whale-call response to masking boat noise," *Nature (London)* **428**, 910.

Frantzis, A. (1998). "Does acoustic testing strand whales?," *Nature (London)* **392**, 29.

Fristrup, K. M., Hatch, L. T., and Clark, C. W. (2003). "Variation in humpback whale (*Megaptera novaeangliae*) song length in relation to low-frequency sound broadcasts," *J. Acoust. Soc. Am.* **113**, 3411–3424.

Gaines, C. A., Hare, M. P., Beck, S. E., and Rosenbaum, H. C. (2005). "Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species," *Proc. R. Soc. London, Ser. B* **272**, 533–542.

Gillam, E., Ulanovsky, N., and McCracken, G. (2007). "Rapid jamming avoidance in biosonar," *Proc. R. Soc. B* **274**, 651–660.

Hamilton, P. K., Knowlton, A. R., Marx, M. K., and Kraus, S. D. (1998). "Age structure and longevity in North Atlantic right whales (*Eubalaena glacialis*) and their relation to reproduction," *Mar. Ecol.: Prog. Ser.* **171**, 285–292.

Jepson, P. D. (2003). "Gas-bubble lesions in stranded cetaceans," *Nature (London)* **425**, 575–576.

Kraus, S. D., Brown, M. W., Caswell, H., Clark, C. W., Fujiwara, M., Hamilton, P. K., Kenney, R. D., Knowlton, A. R., Landry, S., Mayo, C. A., McLellan, W. A., Moore, M. J., Nowacek, D. P., Pabst, D. A., Read, A. J., and Rolland, R. M. (2005). "North Atlantic right whales in crisis," *Science* **309**, 561–562.

Lesage, V., Barrette, C., Kingsley, M. C. S., and Sjare, B. (1999). "The effect of vessel noise on the vocal behavior of belugas in the St. Lawrence River Estuary, Canada," *Marine Mammal Sci.* **15**, 65–84.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS for Mixed Models*, 2nd Ed. (SAS Publishing, Cary, N.C.).

McDonald, M. A., Hildebrand, J. A., and Wiggins, S. M. (2006). "Increase in deep ocean ambient noise in the Northeast Pacific west of San Nicolas Island, California," *J. Acoust. Soc. Am.* **120**, 711–718.

Miller, P. J. O., Biassoni, N., Samuels, A., and Tyack, P. L. (2000). "Whale songs lengthen in response to sonar," *Nature (London)* **405**, 903.

Munk, W. H., Spindel, R. C., Baggeroer, A., and Birdsall, T. G. (1994). "The Heard Island feasibility test," *J. Acoust. Soc. Am.* **96**, 2330–2342.

N.R.C. (2005). *Marine Mammal Populations and Ocean Noise: Determining when Noise Causes Biologically Significant Effects*, National Academies Press, Washington, D.C.

Parks, S. E., and Tyack, P. L. (2005). "Sound production by North Atlantic right whales (*Eubalaena glacialis*) in surface active groups," *J. Acoust. Soc. Am.* **117**, 3297–3306.

Payne, R. S., and Webb, D. (1971). "Orientation by means of long range acoustic signaling in baleen whales," *Ann. N.Y. Acad. Sci.* **188**, 110–141.

Penna, M., Pottstock, H., and Velasquez, N. (2005). "Effect of natural and synthetic noise on evoked vocal responses in a frog of the temperate austral forest," *Anim. Behav.* **70**, 639–651.

Piggott, C. L. (1964). "Ambient sea noise at low frequency in shallow water of the Scotian Shelf," *J. Acoust. Soc. Am.* **36**, 2152–2163.

Richardson, W. J., Greene, C. R., Jr., Malme, C. I., and Thomson, D. H. (1995). *Marine Mammals and Noise* (Academic, San Diego).

Ross, D. (1974). "Ship sources of ambient noise. Proceedings of the International Workshop on Low-Frequency Propagation and Noise. Reprinted in 2005," *IEEE J. Ocean. Eng.* **30**, 257–261.

Scheifele, P. M., Andrew, S., Cooper, R. A., Darre, M., Musiek, F. E., and Max, L. (2005). "Indication of a Lombard vocal response in the St. Lawrence River beluga," *J. Acoust. Soc. Am.* **117**, 1486–1492.

Slabbekoorn, H., and den Boer-Visser, A. (2006). "Cities change the songs of birds," *Curr. Biol.* **16**, 2326–2331.

Slabbekoorn, H., and Peet, M. (2003). "Birds sing at a higher pitch in urban noise," *Nature (London)* **424**, 267.

Sun, J. W. C., and Narins, P. M. (2005). "Anthropogenic sounds differentially affect amphibian call rate," *Stud. Math.* **121**, 419–427.

U.N. (2004). *Review of maritime transport, 2004. UNCTAD/RMT/2004* (Geneva, Switzerland).

Wang, C., and Corbett, J. J. (2005). "Geographical characterization of ship traffic and emissions," *Transp. Res. Rec.* **1909**, 90–99.

Wenz, G. M. (1962). "Acoustic ambient noise in the ocean: spectra and sources," *J. Acoust. Soc. Am.* **34**, 1936–1956.

Zakarauskas, P., Chapman, D. M. F., and Stall, P. R. (1990). "Underwater acoustic ambient noise levels on the eastern Canadian continental shelf," *J. Acoust. Soc. Am.* **87**, 2064–2071.

Application of metrics constructed from vibrotactile thresholds to the assessment of tactile sensory changes in the hands

A. J. Brammer^{a)}

Ergonomic Technology Center, University of Connecticut Health Center, Farmington, Connecticut 06030-2017 and Institute for Microstructural Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada

P. Sutinen

Department of Physical Medicine and Rehabilitation, North Karelia Central Hospital, FIN-80210 Joensuu, Finland and Department of Otorhinolaryngology, Tampere University Hospital, FIN-33521 Tampere, Finland

U. A. Diva

Department of Statistics, University of Connecticut, Storrs Connecticut 06269

I. Pyykkö

Department of Otorhinolaryngology, Tampere University Hospital, FIN-33521 Tampere, Finland

E. Toppila

Department of Otorhinolaryngology, Tampere University Hospital, FIN-33521 Tampere, Finland and Finnish Institute of Occupational Health, FIN-00250 Helsinki, Finland

J. Starck

Finnish Institute of Occupational Health, FIN-00250 Helsinki, Finland

(Received 13 March 2007; revised 14 September 2007; accepted 23 September 2007)

Two tools for assessing tactile sensory disturbances in the hands have been constructed from mechanoreceptor-specific vibrotactile threshold shifts, and thresholds changes with time, and employed in a prospective study of forest workers (N=18). Statistically significant positive threshold shifts (i.e., reductions in sensitivity compared to the hands of healthy persons) were found in five hands at study inception (13.9%), and 15 hands at follow-up (41.7%). Four patterns of threshold shift could be identified, involving selectively the median and/or ulnar nerve pathways and/or end organs. Statistically significant positive threshold changes (i.e., reductions in sensitivity with time) were recorded in 69.4% of the hands over a five-year period, even though a majority of the workers remained symptom free. If the thresholds recorded from subjects not working with power tools are used to control for aging, lifestyle, and environmental factors during the five year period, then 40% of the remaining subjects are found to be experiencing work-related threshold changes in their hands. The ability of the threshold shift metric to predict the numbness reported by these subjects shows that it is closely associated with the tactile sensory changes occurring in their hands. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2799506]

PACS number(s): 43.80.Qf, 43.66.Wv, 43.80.Vj [FD]

Pages: 3732–3742

I. INTRODUCTION

A physiological and psychophysical basis for combining vibrotactile perception thresholds recorded at the fingertips into metrics that are potentially related to changes in tactile acuity has recently been described (Brammer *et al.*, 2007). The purpose of the present work was to construct, and then employ, conservative tools for assessing sensory neurological disturbances in individual hands from these metrics, in order to predict the potential for reduced tactile function in the case of deteriorating thresholds, or for recovery in the case of improving thresholds. With this objective a prospective study appeared necessary, and an open cohort of manual

workers was identified that has been followed since 1972. The health of individuals within the cohort has been the subject of numerous publications, with particular attention having been paid to hand and arm function (Sutinen *et al.*, 2006; Farkkila *et al.* 1980, 1986, 1988). This cohort was selected for the present study as symptoms suggestive of changing tactile function have been reported by almost 30% of the forest workers. Moreover, the predominantly sensory nature of the peripheral neuropathy in forest workers who operate chain saws has been confirmed elsewhere by electro-neuromyography (Bovenzi *et al.*, 2000; Giannini *et al.*, 1999).

The sense of touch in the glabrous skin of the hands is mediated by neural activity in up to four populations of mechanoreceptors, which may be differentiated physiologically on the basis of their morphology, and functionally on

^{a)}Author to whom correspondence should be addressed. Electronic mail: tony.brammer@nrc-cnrc.gc.ca

the basis of their responses to static and dynamic skin indentation (Vallbo and Johansson 1984; Johnson, 2001). The roles of two of these populations in tactile function, the slowly adapting type I receptors (SAI), and the fast adapting type I receptors (FAI), are well established, and alterations in their acuity at the fingertips can be expected to impede perception of the surface features of objects, as well as the gripping and holding of small objects (Johansson and Westling 1984; Johansson and Cole, 1992). Statistically significant associations between vibrotactile thresholds believed mediated, separately, by the SAI and the FAI receptors, as well as the fast-adapting type II receptors (FAII), and symptoms of deterioration in tactile acuity have been reported in a group of manual workers, thus confirming a link between vibrotactile thresholds recorded at the fingertips and tactile performance (Coutu-Wakulczyk *et al.*, 1997). The determination of mechanoreceptor-specific vibrotactile perception thresholds involved a methodology and apparatus that have been described elsewhere (Brammer and Piercy, 1991, 2000; Brammer *et al.*, 2007; Gescheider *et al.*, 2002; Bolanowski *et al.*, 1988).

In this paper, vibrotactile thresholds believed mediated by the SAI and FAI receptor populations at the fingertips are combined in order to facilitate the detection of threshold changes occurring over a period of years. Similarly, the shifts in threshold from the mean value of those recorded at the fingertips of healthy persons are combined with the same objective. In the former metric each subject serves as his own control. The justification for constructing threshold changes and threshold shifts, as defined here, has been described elsewhere (Brammer *et al.*, 2007). A primary consideration is to mitigate the effects of aging on the results. Aging causes thresholds mediated by the FAII receptor population at the fingertips to increase (i.e., decrease in acuity) by, on average, as much as 3 dB/year. The threshold changes with age for the receptor populations included in this analysis are considerably smaller (typically, on average, 0.03 and 0.07 dB/year for the SAI and FAI populations, respectively). The limited effect of aging permits the thresholds for these receptor populations to be followed for intervals of up to at least ten years without introducing errors comparable in magnitude with those customarily associated with psychophysical measurement procedures (Brammer *et al.*, 1992).

After introducing tests to control for potential measurement errors, tools for assessing sensory neurological function are developed by comparing the metrics with statistically based expected values for the threshold shifts, and with the range of threshold changes to be expected from the repeatability of threshold measurements. The purpose of these comparisons is to establish conservative limits for the responses applicable to healthy persons. The metrics are then employed to characterize the threshold shifts, and threshold changes, recorded in the fingers of the subjects on two occasions separated by approximately five years. Patterns of threshold shift are identified, involving either one or both hands, and either the median and/or ulnar sensory nerve pathways and/or end organs. The median and ulnar nerves innervate different fingers (i.e., digits 1–3, and digit 5, re-

spectively), traverse different pathways and anatomical structures from fingers to wrist, and are thus potentially affected differently by mechanical trauma. Statistically significant threshold changes over the five-year period are identified by comparison with the test-retest repeatability for either healthy persons or persons experiencing the same lifestyle and environment as the subjects but who did not operate power tools. After extending the metrics for individual fingers to hands, and to subjects (i.e., combining data for both hands), the threshold changes are compared with estimates of the time working with power tools between studies. The threshold shifts recorded in the follow-up study are compared with measurements of hand grip and the subjects' reports of numbness. Finally, an attempt is made to predict the future course of tactile acuity in individual hands from the observed thresholds changes and threshold shifts.

II. SUBJECTS, APPARATUS AND METHODS

A. Subjects

A group of manual workers was selected from forest workers employed by the National Board of Forestry in the Suomussalmi region of Finland, all of whom are required to attend an annual medical examination in order to work in the industry. Their participation in the study was voluntary, and subjects gave their informed consent. The study protocol was approved by the ethics committees of the participating organizations. The measurements were conducted at the same time and place as the medical examinations on two occasions separated by five years.

The work involved operation of lightweight chain saws for up to ten months per year. The forest workers felled and de-branched softwood trees (typically 25–35 cm diameter at the base), and were paid by piecework. None of the subjects had operated power tools on the day of the health examination and vibrotactile measurements and, in consequence, the observed perception thresholds are believed to be free of any temporary loss in sensation resulting from acute exposure of the hand to intense vibration (Brammer *et al.*, 2006).

B. Assessment of hand and arm function

Signs, symptoms and a description of hand and arm function were obtained during a physical examination conducted by a physician. A neurological examination of the upper extremities was conducted, including tests for carpal tunnel syndrome: sensorimotor function in the upper and lower extremities was examined to screen for polyneuropathies (Sutinen *et al.*, 2006). A questionnaire was also completed by each subject prior to the examination, at which time the responses were confirmed. The maximum compressive hand grip force was also measured following the method described by Farkkila (Farkkila *et al.* 1980, 1986), using a hand grip dynamometer (Jamar Ltd.).

C. Measurement of vibrotactile perception thresholds

1. Apparatus

Mechanoreceptor-specific vibrotactile thresholds were determined at the fingertips using a tactometer. The device

has been described elsewhere (Brammer and Piercy 1991; Brammer *et al.*, 2007), and consists of: 1) a vibration stimulator suspended from a beam balance, the fulcrum of which is mounted on a vertically adjustable track; 2) an arm rest on which the hand and forearm are placed in supination; 3) a small diameter (3 mm) cylindrical probe through which the stimulus is applied to the skin; 4) an accelerometer and electronics to record the motion at the surface of the skin, and; 5) a computer to administer the stimulus consisting of sinusoidal bursts, and calculate perception thresholds. The bursts were 800 ms in duration at frequencies of 20 Hz, and above, and 1.6 s in duration at frequencies of 6.3 Hz, and below. The quiescent interval was 0.6 s. Successive bursts initially increased in intensity until the subject signaled, by pressing a switch, that the stimulus had been detected. This action defined the first ascending, or “upper” threshold. Successive bursts then decreased in intensity until the subject signaled, by releasing the switch, that the stimulus could no longer be felt, so defining the first descending or “lower” threshold. The burst intensity was then, once again, increased. This cycling of burst intensity was repeated at least four times, and the mean threshold acceleration was calculated from the arithmetic sum of the sequence of ascending and descending thresholds when each were expressed in dB re 10^{-6} m/s² (ISO 13091-1, 2001). The algorithm employed for the follow-up study incorporated a test for the consistency of the subject’s performance during the threshold tracking task (Brammer *et al.*, 2007), and could automatically adjust the step size to reduce the measurement time. The minimum step size was 2 dB.

2. Method

The subject was seated with forearm and hand supported horizontally below shoulder level, and positioned the armrest for maximum comfort. This was usually achieved by adjusting the elbow angle and rotating the wrist. Once comfortably seated, the stimulator was positioned so that the probe could be lowered onto a fingertip and maintained in contact with the skin with a static compressive force of 0.05 N. Vibrotactile stimuli were then applied to the skin at amplitudes close to the threshold of perception, according to the algorithm. The skin-stimulator contact conditions and stimulation frequencies were chosen to elicit responses from the SAI or FAI mechanoreceptor populations (Brammer and Piercy, 2000; ISO 13091-1, 2001). Measurements of vibrotactile thresholds believed mediated by the SAI receptors were conducted at 4 and 6.3 Hz, and by the FAI receptors at 20 and 32 Hz.

The performance of the apparatus was confirmed daily, before commencing measurements. The measurement procedure was first explained to a subject, together with a description of the sensations he was likely to feel. After positioning the stimulator, several practice runs were performed to familiarize the subject with the stimuli and measurement procedure. When the examiner was confident the subject was suitably prepared, thresholds were determined at the fingertips of digits 3 and 5 of both hands.

D. Metric for threshold shifts

When all thresholds are expressed in dB, the summed normalized threshold *shift* at a fingertip, $TS_{Sum(SD)}$, can be written (Brammer *et al.*, 2007):

$$TS_{Sum(SD)} = TS_{SAI(SD)} + TS_{FAI(SD)}. \quad (1)$$

Here the metric has been expressed in terms of the mean normalized threshold shifts recorded in each receptor population, $TS_{SAI(SD)}$ and $TS_{FAI(SD)}$, where

$$TS_{SAI(SD)} = TS_{SAI}/SD_{SAI} \quad (2)$$

and $TS_{FAI(SD)}$ has been equivalently defined. The mean receptor-specific threshold shifts from the values observed in healthy hands, TS_{SAI} and TS_{FAI} , are first calculated for each finger from

$$TS_{SAI} = [TS_4 + TS_{6.3}]/2 \quad (3)$$

and

$$TS_{FAI} = [TS_{20} + TS_{32}]/2, \quad (4)$$

where the threshold shift at a given frequency, TS_4 , at 4 Hz, etc., is the difference, in dB, between the observed threshold and the mean threshold recorded from the hands of healthy persons at that frequency (Brammer *et al.*, 1993). When expressed in dB, the thresholds for a population of healthy persons may be approximated by a Gaussian distribution for each frequency (ISO 13091-2, 2003). The range of thresholds for frequencies mediated by the same receptors are found to be essentially the same and so may be expressed in terms of the standard deviation (SD) of their means (Brammer *et al.*, 2007), which are written SD_{SAI} and SD_{FAI} , for the SAI and FAI distributions, respectively. These SDs are employed to normalize the mean receptor-specific threshold shifts (Eq. (2)). Since the response of the two receptor populations may be correlated, the standard deviation of $TS_{Sum(SD)}$ (Eq. (1)) will involve the covariance of $TS_{SAI(SD)}$ and $TS_{FAI(SD)}$, $Cov(TS_{SAI(SD)}, TS_{FAI(SD)})$. In terms of variances (Meyer, 1975)

$$\begin{aligned} Var(TS_{Sum(SD)}) &= Var(TS_{SAI(SD)}) + Var(TS_{FAI(SD)}) \\ &\quad + 2Cov(TS_{SAI(SD)}, TS_{FAI(SD)}) \end{aligned} \quad (5)$$

or, as $Var(TS_{SAI(SD)}) = Var(TS_{FAI(SD)}) = 1$:

$$Var(TS_{Sum(SD)}) = 2[1 + Cov(TS_{SAI(SD)}, TS_{FAI(SD)})], \quad (6)$$

where the covariance is, in this case, equal to the correlation between $TS_{SAI(SD)}$ and $TS_{FAI(SD)}$.

For the purposes of the present work, it is considered necessary to construct conservative assessment tools, as it is intended that health-related decisions will be based, in part, on their value. Accordingly, a restriction is placed on $TS_{SAI(SD)}$ and $TS_{FAI(SD)}$, which are required, separately, to exceed a given magnitude before a value of $TS_{Sum(SD)}$ is considered not “normal:”

$$TS_{SAI(SD)} \geq c, \quad TS_{FAI(SD)} \geq c, \quad (7)$$

where c is expressed in units of standard deviations. The introduction of this condition reduces the possibility of erro-

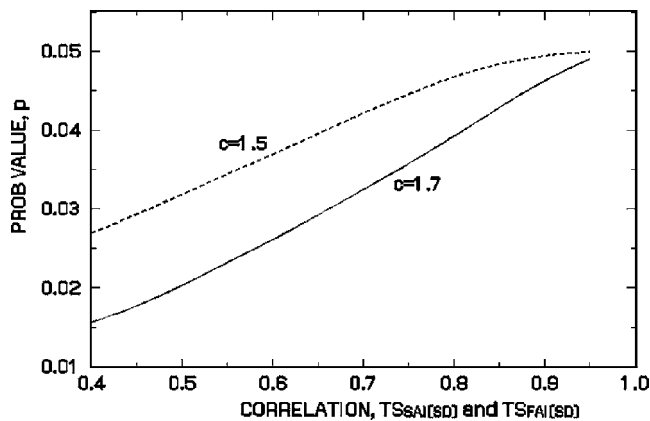


FIG. 1. Effect on probability value of $TS_{Sum(SD)}$ of correlation between $TS_{SAI(SD)}$ and $TS_{FAI(SD)}$, for two values of c .

neous measurements unduly influencing the interpretation of results, for example, $TS_{SAI(SD)}$ being very large when $TS_{FAI(SD)}$ is very small. The effect on the probability value (p value) for $TS_{Sum(SD)}$ is shown in Fig. 1 for two values of c as a function of the correlation between $TS_{SAI(SD)}$ and $TS_{FAI(SD)}$. Inspection of this diagram shows that the influence of the correlation between $TS_{SAI(SD)}$ and $TS_{FAI(SD)}$ on the p value is small when the threshold shifts are highly correlated and, as expected, tends overall to reduce the p value of $TS_{Sum(SD)}$.

The threshold at a fingertip is then considered “normal” (N) if its value of $TS_{Sum(SD)}$ falls within 1.96 standard deviations of the mean value recorded in the hands of healthy persons (i.e., $p < 0.05$). The sensitivity is considered “abnormal” (A) if the value of $TS_{Sum(SD)}$ falls within the parts of the distribution defined by (two-sided) probability values in the range $0.005 < p < 0.05$, and “very abnormal” (AA) when $0.0005 < p < 0.005$. Values of $TS_{Sum(SD)}$ that fall within the tails of the threshold distribution for which $p < 0.0005$ are considered to be “extremely abnormal” (AAA).

E. Metric for threshold changes with time

The threshold *changes* during the five year interval between studies have been estimated, for each finger, from the summed mean threshold change, TC_{Sum} , which becomes when all thresholds are expressed in dB (Brammer *et al.*, 2007):

$$TC_{Sum} = TC_{SAI} + TC_{FAI}. \quad (8)$$

In this expression, TC_{SAI} and TC_{FAI} are the mean threshold changes for thresholds believed mediated by the SAI and FAI receptor populations, respectively, and are constructed from the threshold changes observed at each measurement frequency using expressions equivalent to Eqs. (3) and (4) (e.g., with TC_4 replacing TS_4 , and TC_{SAI} replacing TS_{SAI} , etc.). The threshold change for each finger at each frequency is the difference between the thresholds recorded from the same finger in the two studies. Data were rejected if the difference between threshold changes recorded at frequencies mediated by the same receptor population exceeded a consistency condition with a two-sided probability of occurrence of $p \approx 0.05$ (Brammer *et al.*, 2007).

The analysis of the change in threshold during the five year interval is then based on the statistical significance of the magnitude of TC_{Sum} , as derived from published values for the repeatability of threshold determinations (Brammer *et al.*, 1992). The variance of TC_{Sum} will be related to those of TC_{SAI} and TC_{FAI} by an expression equivalent to Eq. (5). In this case, however, the correlation between TC_{SAI} and TC_{FAI} is related to the covariance by (Meyer, 1975):

$$\rho(TC_{SAI}, TC_{FAI}) = \frac{Cov(TC_{SAI}, TC_{FAI})}{\sqrt{Var(TC_{SAI})Var(TC_{FAI})}} \quad (9)$$

so that as $Var(TC_{SAI}) \cong Var(TC_{FAI})$:

$$Var(TC_{Sum}) \cong 2Var(TC_{SAI})[1 + \rho(TC_{SAI}, TC_{FAI})]. \quad (10)$$

Hence to obtain, again, a conservative assessment tool, the threshold is considered unchanged if the magnitude of TC_{Sum} is less than, or equal to, $\pm 2.81\sqrt{Var(TC_{Sum})}$ ($p=0.005$). Values of TC_{Sum} outside this range are considered to indicate an improvement in sensitivity during the time period when negative (I), and a deterioration in sensitivity when positive (D).

F. Relational tests

Descriptive statistics and Pearson product moment correlations were calculated for the continuous variables. Contingency tables were constructed for symptom reports (categorical variables). The symptoms reported by the forest workers, as confirmed by the examining physician, were taken to be the “gold standard” (i.e., assumed correct), and the ability of the observed threshold shifts, expressed per subject (i.e., four fingers combined), to predict the presence or absence of a symptom has been estimated by calculating the sensitivity, specificity, positive predictive value, and false positive rate (Altman, 1991).

III. RESULTS

A. Subjects, symptoms and signs

Almost 20% of the Suomussalmi forest workers employed by the National Board of Forestry were recruited into the study (23/124 persons). Eighteen of the twenty-three members of this subgroup (78%) attended the compulsory medical examination almost five years later and agreed to continue to participate. Data for these 18 male subjects, who represent almost 15% of the total cohort at inception, are presented here. The reasons for the absence of the other five members of the subgroup from the follow-up examination were pursued. Three of the forest workers were suffering from nonwork related diseases, and one from low back pain: none were still working in the forest industry. One subject could not be traced.

A comparison between the ages, and the prevalences of numbness and vibration-induced white fingers (VWF) for the subgroup and for all the Suomussalmi forest workers is shown in Table I. It is evident that the mean age of the subgroup at inception was somewhat less than that of all the forest workers (39 years, as opposed to 43 years). The age range of the members of the subgroup was also somewhat

TABLE I. Age, prevalence of numbness and white fingers at study inception (all forest workers and subgroup) and at follow-up (subgroup).

	All workers inception	Subgroup	
		Inception	Follow-Up
Mean Age	43	39	44
Age Range	24–60	25–52	30–56
Prevalence of Numbness (%)	28	28	39
Prevalence of VWF (%)	5	6	0

less than that of all the forest workers (25–52 years, as opposed to 24–60 years). More importantly for the purposes of the present work, the prevalences of finger and hand numbness, and finger blanching, reported by the subgroup closely matched the prevalences of this symptom and sign among all the forest workers at inception. In these respects the subgroup is believed to represent the status of the hands of the Suomussalmi forest workers. It should be noted that the prevalence of white fingers is included here as it was the original reason for conducting a prospective study of forest workers in this region of Finland.

A summary of symptoms, hand grip and power tool operating time (TWT) is given for each subgroup member in Table II. Detailed results were unavailable for three subjects (Nos. 4, 12, and 16), who were foremen and did not operate chain saws. In the follow-up study, 39% of the workers reported numbness in the hands, which in two cases (11%) was described as being sufficiently incapacitating to influence the person's ability to work. The maximum hand grip force was measurably less in five subjects (Nos. 10, 11, 13, 14, and 18). The maximum hand grip force was also reduced in the left hand of subject 5. Three subjects had reported vibration-induced white fingers in previous surveys.

TABLE III. Summed normalized shifts in threshold from expected values, at follow-up [expressed in units of SDs, see Eq. (1)].

Subject	LEFT HAND		RIGHT HAND	
	Digit 3	Digit 5	Digit 3	Digit 5
1	7.1	8.0	7.0	7.1
2	3.5	0.0	1.9	-1.8
3	2.9	2.2	4.8	5.2
4	3.8	3.4	2.2	3.2
5	2.9	1.9	1.9	4.3
6	1.3	1.0	1.0	2.4
7	2.3	-0.3	-0.8	-1.8
8	3.4	3.3	0.1	2.7
9	-0.3	-1.1	-0.1	-0.2
10	2.5	1.1	-0.3	-0.4
11	2.6	3.3	2.4	2.0
12	-0.5	-2.3	2.3	3.5
13	6.1	9.6	7.9	7.7
14	4.4	4.8	4.2	4.0
15	3.0	3.9	4.2	4.8
16	1.8	2.1	2.9	3.1
17	5.2	6.5	4.6	4.6
18	4.5	4.6	4.8	5.2

The subjects also estimated the total time during the three complete calendar years between studies that they had worked with chain saws: this is listed as TWT in Table II. Owing to the lack of alternative employment in this region of Finland, the *lifetime* TWT (not estimated) is expected to be related to age in most cases.

B. Status of thresholds relative to those of healthy hands

The summed normalized threshold *shifts* for the left and right hands of each subject at follow-up are shown in Table

TABLE II. Summary of symptoms, hand grip and work history at follow-up for each subject.

Subject	Age (years)	TWT (h) (3 years)	Numbness status	VWF in past	Hand grip force, kg	
					Right	Left
1	43	4620	Yes	No	56.5	50.5
2	46	2900	Yes	No	50.5	46.5
3	49	4620	Yes	No	50.5	46.6
4	37	No
5	46	3000	Yes	No	40.5	33.4
6	38	4320	No	No	54.5	54.5
7	47	2450	No	No	54.5	49.5
8	30	4560	No	No	56.6	59.6
9	30	3000	No	No	48.5	46.5
10	49	2900	No	No	32.3	26.4
11	55	4440	No	Yes	39.4	46.5
12	39	No
13	40	3700	Yes	No	38.3	38.4
14	41	3450	No	No	38.4	36.4
15	53	3720	Yes	Yes	41	50
16	42	No
17	43	3150	No	No	57.6	62.6
18	56	3840	Yes	Yes	32.3	28.3

TABLE IV. Summary of thresholds at inception and follow-up for hands deviating from normal.

Subject	Age	LEFT HAND		RIGHT HAND	
		Inception	Follow-Up	Inception	Follow-Up
1	43	LH-N	LH-AAA	RH3-N, RH5-A	RH-AAA
3	49	LH-N	LH-N	RH-N	RH-A
4	37	LH-N	LH3-A, LH5-N	RH-N	RH-N
5	46	LH3-A, LH5-N	LH-N	RH-AA	RH3-N, RH5-A
13	40	LH-N	LH3-AA, LH5-AAA	RH3-N, RH5-A	RH-AAA
14	41	LH3-N, LH5-A	LH-A,	RH-N	RH-A
15	53	LH-N	LH3-N, LH5-A	RH-N	RH-A
17	43	LH-N	LH3-A, LH5-AA	RH-N	RH-A
18	56	LH-N	LH-A	RH-N	RH-A
Abnormal Hands (A+AA+AAA)		2 (11.1%)	7 (38.9%)	3 (16.7%)	8 (44.4%)

III expressed in units of SD (Eq. (1)). Inspection of Table III shows that the majority of summed normalized threshold shifts are positive rather than negative, implying that the thresholds are, on average, less sensitive in this group of workers than those to be expected from persons engaged in either professional or manual work. For the subgroup, the mean values of $TS_{Sum(SD)}$ for the digits at follow-up were: 3.14 (left hand digit 3-LH3), 2.89 (LH5), 2.82 (RH3), and 3.08 (RH5). In this paper, however, the interest is the status of *individual* hands. Examination of Table III reveals that large threshold shifts in one finger are commonly observed in the corresponding finger of the same, or other, hand, and so are unlikely to be artifactual in origin. There are, however, different patterns of threshold shifts. For example, several subjects possessed thresholds more sensitive than the mean value found in healthy hands (i.e., negative summed normalized threshold shifts), but remained within the range associated with healthy persons.

The status of each hand at inception and follow-up with sensitivity that deviates from the range associated with the hands of healthy persons is summarized in Table IV. To obtain these results, boundaries between the various classifications of $TS_{Sum(SD)}$ have been calculated using Eq. (6) and the values of the correlation between $TS_{SAI(SD)}$ and $TS_{FAI(SD)}$, which are given in Table V. A value of $c=1.5$ has been employed in evaluating the relative magnitude of $TS_{SAI(SD)}$ and $TS_{FAI(SD)}$ [see Eq. (7) and Fig. 1]. The boundary value in Table V has not been adjusted for the restriction placed on

the values of $TS_{SAI(SD)}$ and $TS_{FAI(SD)}$, as the error introduced for these correlations is less than 0.5%. The perhaps unexpectedly large correlations between $TS_{SAI(SD)}$ and $TS_{FAI(SD)}$ may reflect the existence of pathological changes in the hands of our subjects, by which mechanisms thresholds at a given vibrotactile stimulation frequency may be mediated by more than one receptor population.

It can be seen from Table IV that statistically significant shifts in threshold were observed in the hands of four of the 18 subjects at inception (5/36 hands, or 13.9%), and of nine of the 18 subjects (15/36 hands, or 41.7%) at follow-up. Significantly, only one hand was considered to have possessed very or extremely abnormal sensitivity at the inception of the study, that is, with thresholds rated as AA or AAA (RH of subject 5), while a total of 5/36 hands (13.9%) possessed very or extremely abnormal sensitivity approximately five years later at follow-up. The thresholds of both hands of subjects 1 and 13, with the exception of one finger of the latter, were found to be extremely reduced in sensitivity at follow-up: the probability that this threshold would occur in the hands of a healthy person is $p < 0.0005$.

C. Changes in threshold between studies

The values of the summed mean threshold changes, TC_{Sum} , for each finger are shown in Table VI. Note that application of the consistency condition will only lead to the absence of a result (indicated by “R”) when both TC_{SAI} and

TABLE V. Boundaries for summed normalized threshold shifts and summed threshold changes.

Digit	Summed normalized threshold shift					Summed threshold change between studies		
	Correlation $TS_{SAI(SD)}, TS_{FAI(SD)}$	SD ^a	Boundary (SDs)			Correlation TC_{SAI}, TC_{FAI}	SD ^b	Boundary (dB)
			N-A	A-AA	AA-AAA			
LH3	0.87	1.93	3.78	5.42	6.76	0.57	2.57	7.22
LH5	0.97	1.98	3.88	5.56	6.93	0.89	2.82	7.92
RH3	0.94	1.97	3.86	5.54	6.90	0.81	2.76	7.75
RH5	0.87	1.93	3.78	5.42	6.76	0.78	2.73	7.67

^aSee Eq. (6).

^bSee Eq. (10).

TABLE VI. Summed changes in threshold between studies, in dB [see Eq. (8)]. Legend: R-data rejected; M-data missing.

Subject	LEFT HAND		RIGHT HAND	
	Digit 3	Digit 5	Digit 3	Digit 5
1	23.0	22.2	21.4	21.8
2	9.9	-11.5	2.9	-10.2
3	5.5	-4.6	7.1	12.2
4	13.7	M	7.6	M
5	-9.6	-8.6	-19.3	-9.3
6	4.1	25.6	12.0	9.1
7	6.2	1.0	-10.1	-10.6
8	16.3	13.4	-3.9	6.3
9	10.5	2.8	4.8	6.6
10	6.3	R	6.1	-3.7
11	3.0	8.6	7.4	3.6
12	14.5	M	13.0	M
13	20.9	34.0	35.4	19.1
14	4.9	1.9	6.6	3.8
15	16.5	19.3	18.6	22.8
16	4.2	4.4	4.4	15.7
17	13.0	22.3	18.3	14.3
18	20.8	21.9	10.8	7.0

TC_{FAI} have been rejected. Inspection of this table reveals that large summed mean threshold changes were recorded in all fingers of several subjects (e.g., subjects 1, 13, and 15). These changes imply diminishing sensitivity during the five year period. The magnitude of the changes should be compared with those recorded in subject 14, whose thresholds appear to have changed much less during the five years. Also, clearly negative summed mean threshold changes, that is, improving sensitivity during the five year period, were observed in the hands of subject 5.

An assessment of changes in the summed mean threshold from inception to follow-up is shown for all subjects in Table VII (left hand—column 3; and right hand—column 5). To obtain these results, the boundary between unchanged and changed sensitivity has been established from the SD of TC_{Sum} . The magnitude of this metric has been calculated using Eq. (10) and the measured values of the correlation between TC_{SAI} and TC_{FAI} (see Table V). Changes in sensitivity during the time interval between studies were recorded in 16 of the 18 subjects (25/36 hands, or 69.4%) ($p < 0.005$). Of these, 20 hands were assessed to be deteriorating (D) in either one or both fingers, and four improving (I). The threshold changes recorded in the fingers of one hand gave conflicting results. Both hands were found to be improving in only one subject (5.6%), while both hands of two subjects were considered not to have changed during the five years (11.1%).

D. Relationships between threshold metrics, age, grip force and work history

In order to explore the relationships between a subject's age, hand grip force, work history and either threshold shifts, or changes, it is necessary to construct metrics for each hand, and for each person, rather than for each finger. For each hand, the larger of the threshold changes, or shifts, observed for digit 3 or 5 has been taken as the appropriate metric. Similarly, the larger of the threshold change, or shift, constructed for each hand has been taken as the appropriate single metric for the subject.

The correlations between the threshold metrics, age, grip force and work history are shown in Table VIII. The thresholds shifts and maximum grip forces are those recorded at follow-up. Inspection of the correlations in Table VIII re-

TABLE VII. Changes in threshold sensitivity over five years, and prediction of thresholds ten years after study inception, for all subjects.

Subject	Age (years)	LEFT HAND		RIGHT HAND	
		Change in sensitivity	Threshold predicted in 10 years	Change in sensitivity	Threshold predicted in 10 years
1	43	LH-D	LH-AAA	RH-D	RH-AAA
2	46	LH3-D, LH5-1	LH3-A, LH5-N	RH3-NC, RH5-1	RH-N
3	49	LH-NC	LH3-A, LH5-N	RH3-NC, RH5-D	RH3-AA, RH5-AAA
4	37	LH-D	LH-AA	RH-NC	RH-N
5	46	LH-1	LH-N	RH-1	RH-N
6	38	LH3-NC, LH5-D	LH3-N, LH5-AA	RH-D	RH3-N, RH5-A
7	47	LH-NC	LH3-A, LH5-N	RH-1	RH3-N, RH5-A
8	30	LH-D	LH-AA	RH-NC	RH3-N, RH5-A
9	30	LH3-D, LH5-NC	LH-N	RH-NC	RH-N
10	49	LH-NC	LH-N	RH-NC	RH-N
11	55	LH3-NC, LH5-D	LH3-N, LH5-A	RH-NC	RH-N
12	39	LH-D	LH-N	RH-D	RH-A
13	40	LH-D	LH-AAA	RH-D	RH-AAA
14	41	LH-NC	LH-A	RH-NC	RH-A
15	53	LH-D	LH3-AA, LH5-AAA	RH-D	RH-AAA
16	42	LH-NC	LH-N	RH3-NC, RH5-D	RH3-N, RH5-AA
17	43	LH-D	LH3-AA, LH5-AAA	RH-D	RH-AAA
18	56	LH-D	LH-AAA	RH3-D, RH5-NC	RH-AAA
Abnormal Hands (A+AA+AAA)			13 (72.2%)		12 (66.7%)

TABLE VIII. Correlations between threshold metrics, age, grip force and TWT.

Correlation between:	By hand	By subject
$TS_{Sum(SD)}$ of digits 3 and 5 (RH)	0.87 ^b	
$TS_{Sum(SD)}$ of digits 3 and 5 (LH)	0.90 ^b	
TC_{Sum} of digits 3 and 5 (RH)	0.79 ^b	
TC_{Sum} of digits 3 and 5 (LH)	0.67 ^a	
$TS_{Sum(SD)}$ and Grip Force (RH)	-0.10	
$TS_{Sum(SD)}$ and Grip Force (LH)	0.00	
$TS_{Sum(SD)}$ and age		0.20
TC_{Sum} and age		-0.12
TC_{Sum} and TWT		0.43
Age and Grip Force (RH)	-0.51 ^a	
Age and Grip Force (LH)	-0.41	

^a $p < 0.05$
^b $p < 0.0005$

veals that the summed normalized threshold shifts for fingers within the right, or left, hand are closely related. The Pearson correlation coefficients were 0.87 and 0.90 for the right and left hands, respectively ($p < 0.0005$). Clearly, on a group basis, whatever is causing the threshold shifts appears to be affecting similarly digits 3 and 5 of the same hand. The largest threshold shifts occurring in each hand are not, however, correlated with the maximum grip force. The measured maximum grip force is negatively correlated with age in both hands, as might be expected, but the correlation coefficients were only statistically significant for the right hand. The summed normalized threshold shift constructed for each subject is not related to age. Nevertheless, there is a tendency for the threshold shift to increase slightly with the age of the subject, as can be seen from Fig. 2. There is no physical basis for attempting to relate threshold shifts to data dependent on the time interval between studies, such as the TWT, which may, however, be related to threshold changes.

The Pearson correlation coefficient for the summed threshold changes mediated by receptors in digits 3 and 5 were 0.79 and 0.67 for the right and left hands, respectively, which are smaller than the corresponding correlations between the summed normalized threshold shifts for receptor populations in different fingers, but both were statistically significant. The summed threshold change constructed for

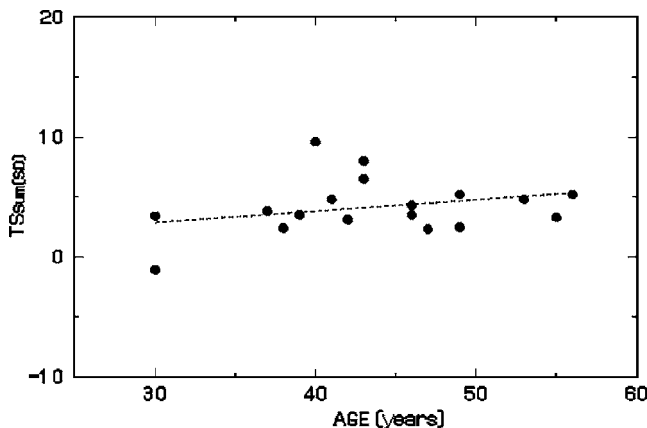


FIG. 2. Relationship between $TS_{Sum(SD)}$ and age.

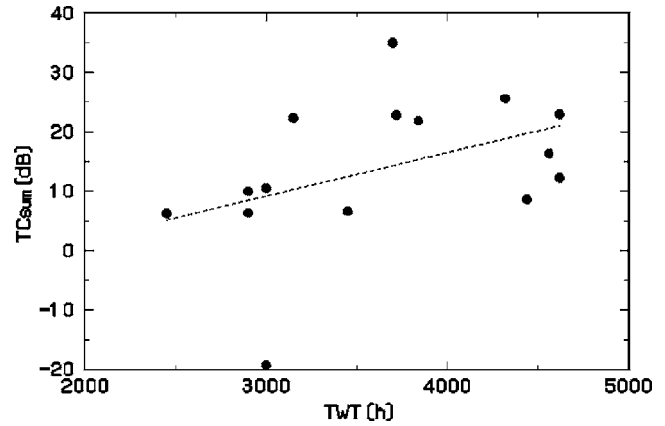


FIG. 3. Relationship between TC_{Sum} and TWT.

each subject is also not related to age. The summed threshold change for each subject is, however, somewhat correlated with the estimated TWT, but the association does not reach statistical significance ($p = 0.10$). The tendency for the summed threshold changes to increase with the magnitude of the TWT can be seen from Fig. 3. In this diagram a linear regression for these data with the form $y = mx + c$, shown by the dashed line, possesses a gradient, m , of 7.3 dB/1000 h. If the summed threshold change is normalized by the corresponding TWT for each subject, to remove a relationship between threshold change and TWT, there is a tendency for the threshold change to decrease with age (Fig. 4), but the effect does not reach statistical significance.

E. Prediction of numbness from summed normalized threshold shift

The ability of the measured threshold shifts to predict the presence or absence of numbness in the hands was examined by forming contingency tables. The threshold metric was again the summed normalized threshold shift constructed for each subject. The positive predictive value of the metric was 75% and the negative predictive value was 85.7%, when the presence or absence of the symptom was defined by the boundary between “normal” and “abnormal” (symptom prevalence—38.9%). These values were associated with a sensitivity of 85.7% and specificity of 75%.

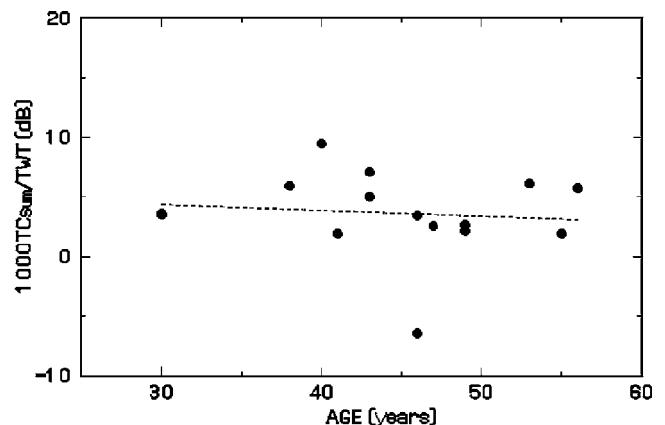


FIG. 4. Relationship between TC_{Sum} and age. TC_{Sum} has been normalized by the TWT.

IV. DISCUSSION

A. Interpretation of threshold shifts and threshold changes

In essence, the threshold *shifts* provide a measure of the status of each finger/nerve system at inception, and follow-up, for an individual relative to the mean thresholds of healthy persons, while the threshold *changes* provide a measure of what has happened to each finger/nerve system over the five-year period.

Inspection of Table III reveals that the summed normalized threshold shifts tend to follow one of several patterns. Overall, there is a general elevation in threshold (i.e., positive threshold shift) corresponding, implicitly, to a reduction in acuity in the hands of this group of forest workers. Note that care must be taken not to attach significance to small differences between summed normalized threshold shifts (e.g., less than 2 SDs) when searching for patterns.

In one pattern, the elevations in threshold of digits 3 and 5 of the same hand tend to be similar in magnitude. The pattern is often common to both hands. This was more evident in the data obtained at follow-up (shown in Table III). The pattern is displayed by a majority of the forest workers in this study, though the magnitude of the summed threshold shifts, and therefore their interpretation, differs between individuals (e.g., see Table IV). Thus, for subject 1, the summed normalized threshold shifts observed in the left hand (Table III) are: 7.1 (digit 3), and 8.0 (digit 5); while those in the right hand are: 7.0 (digit 3), and 7.1 (digit 5). For this subject, the magnitude of the threshold shifts would rank them as “extremely abnormal.” The results suggest that there are neurological changes occurring similarly in the nerve fibers and/or end organs mediating tactile perception at the fingertips of digits 3 and 5 in both hands of this subject. It may therefore be inferred that both median and ulnar sensory nerve fibers and/or end organs are affected similarly, though the site(s) of any pathophysiological changes cannot, of course, be established from these results.

A second pattern, which is uncommon in this group of subjects, is a variation on the first in which the summed normalized threshold shifts of digit 3 and 5 of the same hand tend to be similar in magnitude, but the magnitude of the shift differs between the left and right hands of a subject. An example is provided by the results for subject 3 (Table III).

A third pattern, which is also uncommon in this group of subjects, involves larger threshold shifts in digit 3 than in digit 5 of the same hand. Examples are the threshold shifts from subject 2 (left, and right), and 7 (left). In this case the inference would be that the median sensory nerve fibers and/or end organs are more affected than the ulnar sensory nerve fibers and/or end organs of these hands. A corresponding fourth pattern involves larger threshold shifts in digit 5 than in digit 3 of the same hand. Examples are the threshold shifts for the hands of subject 5 (right) and 13 (left), for which the interpretation would be the opposite of the third pattern.

Overall, the trend from study inception to follow-up is one of increasing summed normalized threshold shift (Table IV). There was one exception. The thresholds of subject 5

show a marked reduction (i.e., improvement in sensitivity), which was coincident with a reduction in alcohol consumption during this period. A small fiber neuropathy had been detected by electroneuromyography in this subject.

A clearer picture of the progression in individual hands is obtained from the summed threshold changes over the five-year period (Table VI). In this table the improvement in sensitivity in the hands of subject 5 is reflected in summed threshold changes of: -9.6 dB (digit 3), and -8.6 dB (digit 5) in the left hand, and; -19.3 dB (digit 3), and -9.3 dB (digit 5) in the right hand. Unfortunately, such improvements were infrequently observed in these forest workers. The common trend was one of deteriorating sensitivity, with summed mean threshold changes of up to 35.4 dB recorded. Large threshold changes when recorded from all fingers of a subject (e.g., subject 1), are suggestive of substantial neurological changes occurring similarly in both hands. The status of this subject was classified as deteriorating.

Positive summed mean threshold changes (i.e., reductions in sensitivity) in excess of 20 dB were observed in the fingers of, in total, eight hands (subjects 1, 6, 13, 15, 17, and 18). Studies of the rate of threshold change with age in healthy persons (including professional and manual workers) would yield a mean summed threshold change of $TC_{Sum} \cong 0.7$ dB due to aging during the five year period (Brammer *et al.*, 1993), much less than the magnitude of TC_{Sum} observed in these subjects. If the thresholds recorded from the three foremen included in this study (who did not work with power tools) are considered to provide a more appropriate measure of threshold change over the five year period for the lifestyle and environment of the Suomussalmi region, then the summed threshold change for persons not using power tools at work would be $TC_{Sum} = 9.7 \pm 5.0$ (SD) dB, leading to a 95% confidence interval of $0 < TC_{Sum} < 19.5$. While the large confidence interval and small number of workers from which it was derived may be questioned in view of the small threshold change anticipated solely from aging, the changes in subjects 1 and 13 (both hands), and subjects 6, 15, 17, and 18 (one hand) still remain statistically significant. Thus, arguably 40% of the remaining subjects are experiencing *work-related* threshold changes in their hands.

The statistical associations between threshold shifts and threshold changes in digits 3 and 5 of the right, or left, hand (Table VIII) suggest that it is reasonable to consider the vibrotactile effects by hand rather than by finger. Thus there is justification for attempting to associate the detailed psychophysical measurements with symptoms or gross measurements of hand function in these manual workers. The correlations between threshold shifts tend to be greater than those between threshold changes. There could be several reasons for this observation. Two psychophysical measurements are involved in the measurement of threshold change, introducing the variability associated with two test/retest errors. An indication of the confidence interval resulting from the use of this metric in this study has been calculated from the results obtained from the foremen, which also introduces environmental and lifestyle related factors. As already noted, the confidence interval obtained in this way is much larger than that employed generally for the analysis. Second, the psy-

chophysical threshold determinations at inception employed a comparatively primitive measurement algorithm, which was incapable of flagging potential inconsistencies in a subject's responses. This would be expected to lead to larger variability in the thresholds recorded at inception, which are required to calculate the threshold changes, than those at follow-up, which have been used to calculate the threshold shifts analyzed here.

While there was a tendency for the threshold metrics to be influenced by the age of the subject (Figs. 2 and 4), the effects never reached statistical significance. Of the possible relationships between the threshold metrics and symptoms, hand function and work history, that between threshold change and TWT (the "conditioning" vibration exposure) came closest to statistical significance, suggesting that the observed thresholds may well be related to the vibration exposure (as opposed to other aspects of the work). The ability of the threshold shifts to predict the numbness reported by the forest workers shows that this metric is closely associated with the tactile sensory changes occurring in the hands of these subjects.

B. Prediction of future thresholds

In view of the significant deterioration in threshold during the five-year interval between studies in the majority of hands (Table VII), together with the increase in the number of hands with thresholds deviating from the values expected in healthy hands from inception to follow-up (Table IV), it appeared appropriate to predict what the thresholds of each subject would be five years after the follow-up study was performed, assuming the work patterns remained the same. This has been done by using the magnitude of the threshold changes at each measurement frequency to adjust the thresholds observed at the follow-up study.

The results are shown for each subject in Table VII, and provide the predicted status of each hand ten years after the commencement of the study (columns 4 and 6). It can be seen that the sensitivity of 13 left and 12 right hands within the group (69.4%) is predicted to be abnormal five years after the most recent measurements. For the affected workers, five are expected to possess hands with extremely abnormal sensitivity (1, 13, 15, 17, and 18), and one additional subject is predicted to possess one hand with extremely abnormal sensitivity (3). Since the sensitivity of the hands of the workers with both hands affected is demonstrably deteriorating, their prognosis is disturbing if the present pattern of work is continued in the future. Also of concern is the predicted deterioration in the sensitivity of the left hand of subject 8 in view of his age (25 years at inception).

V. CONCLUSIONS

Two tools for assessing tactile sensory disturbances in the hands have been constructed from mechanoreceptor-specific vibrotactile threshold shifts, and thresholds changes with time, and applied to a small group of forest workers (N=18). Statistically significant positive threshold shifts (i.e., reductions in sensitivity compared to the hands of healthy persons) were found in five hands at study inception

(13.9%), and 15 hands at follow-up (41.7%). One hand was judged to have possessed very abnormal sensitivity at the inception of the study, while a total of 5/36 hands possessed very or extremely abnormal sensitivity approximately five years later at follow-up. Four patterns of threshold shift could be identified, involving either one or both hands, and either the median and/or ulnar sensory nerve pathways and/or end organs. The most common pattern in this group of subjects involved both hands and both nerves. Statistically significant positive threshold changes (i.e., reductions in sensitivity) were recorded in 69.4% of the hands over a five-year period ($p < 0.005$), even though a majority of workers remained symptom free. If the thresholds recorded from subjects not working with power tools are considered to define the limits of threshold change applicable to aging, lifestyle and environmental factors during the five year period, then 40% of the remaining subjects are found to be experiencing work-related threshold changes in their hands. Five subjects (27.8%) are predicted to possess hands with extremely abnormal sensitivity, and one subject (5.6%) one hand with extremely abnormal sensitivity, in five years' time if the present work practices are continued. Of the possible relationships between the threshold metrics and symptoms, hand function and work, the ability of the threshold shifts to predict the numbness reported by the forest workers shows that this metric is closely associated with the tactile sensory changes occurring in the hands of these subjects.

ACKNOWLEDGMENTS

The authors wish to acknowledge the provision of financial assistance by the Finnish National Board of Forestry and the Finnish Forestry Fund. They also wish to acknowledge the contribution of the late Dr. K. Koskimies to the physical examinations of the forest workers, and of S. Das, Department of Statistics, University of Connecticut, to the statistical analysis. The authors greatly appreciated the cheerful support of other members of the research team, and the staff and interpreters of the Regional Health Clinic at Ämmänsaari, Finland. The authors remain indebted to the forest workers who generously and patiently provided their time.

- Altman, D. G. (1991). *Practical Statistics for Medical Research* (Chapman & Hall/CRC, Boca Raton, FL).
- Bolanowski, S. J., Gescheider, G. A., Verrillo, R. T., and Checkosky, C. M. (1988). "Four channels mediate the mechanical aspects of touch," *J. Acoust. Soc. Am.* **84**, 1680–1694.
- Bovenzi, M., Giannini, F., and Rossi, S. (2000). "Vibration-induced multifocal neuropathy: Electrophysiological findings in relation to vibration exposure and finger circulation," *Int. Arch. Occup. Environ. Health* **73**, 519–527.
- Brammer, A. J., Peterson, D. R., Cherniack, M. G., and Diva, U. A. (2006). "Temporary changes in mechanoreceptor-specific vibrotactile perception to stimuli simulating impact power tools," in *Proc. of Inter-Noise 2006* (Noise Control Foundation, New York).
- Brammer, A. J., and Piercy, J. E. (1991). "Measuring vibrotactile perception thresholds at the fingertips of power-tool operators," in *Proc. of the U.K. Informal Group on Human Response to Vibration* (Buxton, United Kingdom), pp. 1–7.
- Brammer, A. J., and Piercy, J. E. (2000). "Rationale for measuring vibrotactile perception at the fingertips as proposed for standardization in ISO 13091-1," *Arbetslivsrapport* **4**, 125–132 (ISSN 1400-8211).
- Brammer, A. J., Piercy, J. E., Nohara, S., and Nakamura, S. (1992). "Repeatability of threshold measurements used to assess sensory changes in

- the hands," in *Proc. of Inter-Noise 92*, edited by G. A. Daigle and M. R. Stinson (Noise Control Foundation, New York), pp. 1127–1130.
- Brammer, A. J., Piercy, J. E., Nohara, S., Nakamura, H., and Auger, P. L. (1993). "Age-related changes in mechanoreceptor-specific vibrotactile thresholds for normal hands," *J. Acoust. Soc. Am.* **93**, 2361 (A).
- Brammer, A. J., Piercy, J. E., Pyykkö, I., Toppila, E., and Starck, J. (2007). "Method for detecting small changes in vibrotactile perception threshold related to tactile acuity," *J. Acoust. Soc. Am.* **121**, 1238–1247.
- Coutu-Wakulczyk, G., Brammer, A. J., and Piercy, J. E. (1997). "Association between a quantitative measure of tactile acuity and hand symptoms reported by operators of power tools," *J. Hand Surg. [Am]* **22A**, 873–881.
- Farkkila, M., Pyykkö, I., Korhonen, O., and Starck, J. (1980). "Vibration induced decrease in muscle force in lumberjacks," *Eur. J. Appl. Physiol.* **43**, 1–9.
- Farkkila, M., Aatola, S., Starck, J., Korhonen, O., and Pyykkö, I. (1986). "Hand grip force in lumberjacks: Two year follow-up," *Int. Arch. Occup. Environ. Health* **58**, 203–208.
- Farkkila, M., Pyykkö, I., Jantti, V., Aatola, S., Starck, J., and Korhonen, O. (1988). "Forestry workers exposed to vibration: A neurological study," *Br. J. Ind. Med.* **45**, 188–192.
- Gescheider, G. A., Bolanowski, S. J., Pope, J. V., and Verrillo, R. T. (2002). "A four channel analysis of the tactile sensitivity of the fingertip: Frequency selectivity, spatial summation, and temporal summation," *Somatosens Mot Res.* **19**, 114–124.
- Giannini, F., Rossi, S., Passero, S., Bovenzi, M., Caannava, G., Mancini, R., Cioni, R., and Battistini, N. (1999). "Multifocal neural conduction impairment in forestry workers exposed and not exposed to vibration," *Clin. Neurophysiol.* **110**, 1276–1283.
- ISO 13091-1. (2001). Mechanical vibration: Vibrotactile perception thresholds for the assessment of nerve dysfunction—Part 1: Methods of measurement at the fingertips (International Organization for Standardization, Geneva).
- ISO 13091-2. (2003). Mechanical vibration: Vibrotactile perception thresholds for the assessment of nerve dysfunction—Part 2: Analysis and interpretation of measurements at the fingertips (International Organization for Standardization, Geneva).
- Johansson, R. S., and Cole, K. J. (1992). "Sensory-motor coordination during grasping and manipulative actions," *Curr. Opin. Neurobiol.* **2**, 815–823.
- Johansson, R. S., and Westling, G. (1984). "Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects," *Exp. Brain Res.* **56**, 550–564.
- Johnson, K. O. (2001). "The roles and functions of cutaneous mechanoreceptors," *Curr. Opin. Neurobiol.* **11**, 455–461.
- Meyer, S. L. (1975). *Data Analysis for Scientists and Engineers* (Wiley, New York), p. 144.
- Sutinen, P., Toppila, E., Starck, J., Brammer, A. J., Zou, J., and Pyykkö, I. (2006). "Hand-arm vibration syndrome with use of anti-vibration chain saws: 19-year follow-up of forestry workers," *Int. Arch. Occup. Environ. Health* **79**, 665–671.
- Vallbo, Å. B., and Johansson, R. S. (1984). "Properties of cutaneous mechanoreceptors in the human hand related to touch sensation," *Hum. Neurobiol.* **3**, 3–14.