CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 125, No. 1

January 2009

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

PHYSIOLOGICAL ACOUSTICS [64]

PSYCHOLOGICAL ACOUSTICS [66]

SPEECH PRODUCTION [70]

SPEECH PERCEPTION [71]

(Continued)

CONTENTS—*Continued from preceding page*

COMMENTS ON WORK PUBLISHED ELSEWHERE

The *Journal of the Acoustical Society of America* welcomes the submission of letters or papers whose primary content is a commentary on work that has been previously published on acoustics anywhere and under any circumstances. This policy has been tacitly in place ever since the inception of the *Journal* in 1929, and is fully in accord with the stated mission of the Society — to increase and diffuse the knowledge of acoustics and to promote its practical applications.

Possible works that might be commented upon include (i) articles published in other journals, (ii) articles in conference proceedings, (iii) books or sections of books, (iv) widely disseminated or cited reports or thesis, and (v) articles distributed via internet sites.

It is suggested that the title of any such article begin with the words "Comments on" followed by the title of the work being commented upon in the quotation marks, followed by an identification of the source of the publication enclosed in brackets. For an example of an acceptable format, see the article on pp. 1401–1403 of the September 2008 issue of the *Journal*. The portion which contains the title can be abbreviated, and ellipses can be used to keep the overall title of reasonable length. The bracketed identification of the source can also be abbreviated as much as the author deems appropriate. Author names should be omitted. The guiding principle is that there should be enough information in the title so that the commenting article would be easily found in any online search which brought up the original item which is being commented upon.

Any article, if submitted to the *Journal*, would be subjected to the same peer-review processes as are all submitted manuscripts subjected. The article will not be considered unless at least one of the current Associate Editors is willing to handle it, and the criteria for acceptance will include those that the article be "truly acoustics," original, and of substantial significance. Any reviewers consulted will be disinterested third parties. Once an article is accepted, and, if appropriate, the authors of the work being commented upon will be notified and asked if they wish to submit a rebuttal. If they do submit a rebuttal sufficiently promptly, and if it is accepted for publication, the comment article and the rebuttal will be published simultaneously in the *Journal*. The rebuttal article will also be subject to the peer-review process, and any consulted reviewers will be disinterested third parties.

Magneto-mechanical impedance of metallic structures

Andrei N. Zagrai
Department of Mechanical Engineering, New Mexico Institute of Mining and Technology, Socorro, New Mexico 87801
azagrai@nmt.edu

Abstract: Impedance measurements with a magneto-elastic active sensor are explored for inferring the magneto-mechanical impedance (MMI) of a metallic structure. It is shown that the MMI contains electrical response of the sensor and both electrical and mechanical structural responses. An analytical model is suggested that accounts for electrical characteristics of the sensor, sensor/structure electromagnetic interaction, and multimodal structural dynamic behavior. The model is validated with a set of MMI experiments demonstrating feasibility of deducing structural natural frequencies and structural vibration modes.

PACS numbers: 43.40.−r, 43.40.At, 43.40.Yq [JM]
Date Received: June 9, 2008 **Date Accepted:** October 2, 2008

1. Introduction

Impedance measurements are widely used in structural engineering for vibration testing, nondestructive evaluation and structural health monitoring. Application of impedance measurement to vibration diagnosis is discussed in numerous publications and monographs including works of ASME (1958) and Skudrzyk (1968). Lange (1972) provided details on utilizing mechanical impedance measurements for nondestructive inspection. Alternatively, structural impedance at relatively high frequencies can be inferred via the electromechanical impedance (EMI) testing pioneered by Liang *et al.* (1994). The EMI method employs thin piezoelectric wafers permanently bonded to a structural element. The electrical impedance of a piezoelectric wafer sensor is measured and, because of the mechanical coupling at the sensor/structure interface and the electromechanical transformation enabled by the sensor, the resultant electrical impedance reflects structural dynamic response of the host structure. A central point of mechanical and electromechanical impedance measurements is that both measurement methods explore structural dynamic characteristics of test specimens.

In this letter we discuss magneto-mechanical impedance (MMI) measurements. This term is used to reflect the principle of the method in which the mechanical response is obtained through impedance measurements facilitated by the electromagnetic interaction between the magneto-elastic sensor and the metallic structure. It is possible that one of the first measurements of the magneto-mechanical impedance was conducted by Johnson *et al.* (1994). In this initial work involving relatively large set of transducers, multimodal structural response was not considered as well as a continuous structural model. The aim of this letter is to show that MMI measurements can be used for structural dynamic identification involving multiple resonances and to provide an analytical model incorporating contributions of the continuous structure, the electromagnetic coupling, and the sensing element.

2. Analytical model for magneto-mechanical impedance sensing

Magneto-mechanical impedance sensing is based on the effect of electromagnetic generation and reception of elastic waves thoroughly described by Banik and Overhauser (1977). By analogy to the EMI method employing an array of piezoelectric wafer active sensors for local impedance measurements, we intend to utilize a network of magneto-elastic active sensors (MEAS) for MMI measurements at various structural locations (Zagrai and Cakan, 2007). The

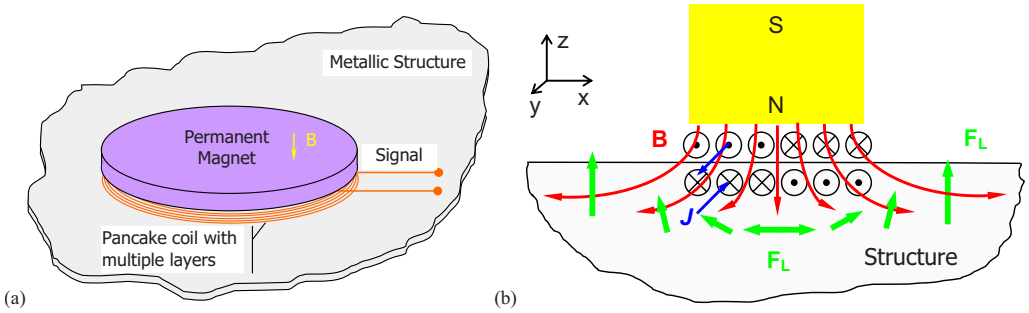


Fig. 1. (Color online) (a) Schematics of a MEAS; (b) electro-magneto-elastic interaction of MEAS and the metallic structure.

sensor consists of an electromagnetic coil and a nickel plated neodymium magnet depicted in Fig. 1. Because of these two elements, the impedance of the sensor is dominated by the frequency dependent inductive component.

For the sensor arrangement depicted in Fig. 1, configuration of magnetic field is complex. As it can be seen in the figure, field lines at the center of the sensor show distinct z component, but field lines further away from the center are primarily oriented along x direction. When the MEAS is positioned on or above the structural surface, these magnetic induction field lines extend into the metallic structure adjacent to the sensor. An electric current flowing in the sensor coil induces eddy currents in the specimen, which flow in the opposite direction. Mutual orientation of the induced eddy currents and a vector of magnetic induction B define the Lorentz force acting on electrons. This force is transferred to lattice ions and is responsible for generation of the elastic wave. In the CW regime, emission and reflection of elastic waves result in standing waves at respective structural natural frequencies. Consequently, the structural dynamic behavior is reflected in the MEAS impedance signature due to a reverse transduction mechanism.

In this letter, a continuous one-dimensional nonferromagnetic metallic structure is considered. The structure is excited by a harmonically varying Lorentz force $F_L(x,t)=(J \times B)\cdot e^{i\omega t}$ defined by a product of the electric current J and magnetic induction B. In the vicinity of the sensor, F_L is not uniform and is comprised of vertical and horizontal components. The former excites flexural vibrations in a one-dimensional structure modeled with the Euler–Bernoulli beam theory

$$\rho A \cdot \frac{\partial^2 w(x,t)}{\partial t^2} + EI \cdot \frac{\partial^4 w(x,t)}{\partial x^4} = F_L(x,t). \tag{1}$$

Equation (1) incorporates the following parameters of the beam: ρ is the density, A is the cross-sectional area, E is the Young's modulus, and I is the moment of inertia. If dimensions of a sensor are much smaller than the major dimension of a structure, spatial distribution of the excitation force may be ignored and the single-point excitation may be assumed.

$$F_L(x,t) = J_y \cdot B_x \cdot b_a \cdot \delta(x - x_a) \cdot e^{i\omega t}, \tag{2}$$

where J_y and B_x are electric current and magnetic induction in the beam, ω is the excitation frequency, b_a is the y-axis dimension of the sensor equal to beam's width, $\delta(x\text{-}x_a)$ is the Dirac delta function, and x_a is the MEAS position.

The solution of Eq. (1) for the displacement of the beam, $w(x,t)$, is sought in terms of modal expansion

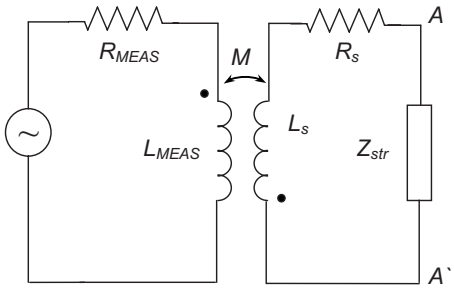


Fig. 2. Equivalent circuit for magneto-elastic active sensing.

$$w(x,t)=\sum_{n=0}^{\infty} W_n(x)\cdot T_n(t), \tag{3}$$

with the temporal component, $T_n(t)=C_n\cdot e^{i\omega t}$, and the spatial mode-shape corresponding to free-free boundary conditions: $W_n(x)=A_n[\cosh\gamma_n x+\cos\gamma_n x-\sigma_n(\sinh\gamma_n x+\sin\gamma_n x)]$. The mode-shape parameters $A_n=1/\sqrt{\int_0^L W_n^2(x)dx}$, γ_n, σ_n can be calculated numerically; e.g., for free-free boundary conditions and for $n>5$, $\gamma_n=\pi(2n+1)/2L$, $\sigma_n=1$. Natural frequencies of the system are obtained as $\omega_n=\gamma_n^2 c_w$, where the sound speed $c_w=\sqrt{EI/\rho A}$.

Substituting the modal expansion (3) into Eq. (1), employing the orthogonality condition, and introducing damping via a damping ratio ζ_n for each vibration mode, the modal participation factor is determined

$$C_n=\frac{1}{\rho A}\sum_{n=0}^{\infty}\int_0^L W_n(x)\cdot F_L(x)\mathrm{d}x/(\omega_n^2+2i\varsigma_n\omega\omega_n-\omega^2). \tag{4}$$

The displacement (3) yields

$$w(x,t)=\sum_{n=0}^{\infty}\frac{W_n(x)\cdot W_n(x_a)\cdot J_y\cdot b_a\cdot B_x}{\rho A\cdot(\omega_n^2+2i\varsigma_n\omega\omega_n-\omega^2)}\cdot e^{i\omega t}. \tag{5}$$

Considering a reciprocal effect of inducing the electromagnetic field due to the propagating elastic wave (Turner *et al.*, 1969), displacement (5) can be used to determine the resulting voltage

$$V=\dot{w}(x_a,t)\cdot B_x\cdot\int_0^{b_a}dy=\dot{w}(x_a,t)\cdot B_x\cdot b_a, \tag{6}$$

where $\dot{w}(x_a,t)$ represents velocity measured at location x_a. The magneto-mechanical impedance is presented in terms of the voltage/current ratio

$$Z_{\mathrm{str}}(\omega)=\frac{V}{J_y}=\sum_{n=0}^{\infty}\frac{i\omega\cdot(W_n(x_a)\cdot b_a\cdot B_x)^2}{\rho A\cdot(\omega_n^2+2i\varsigma_n\omega\omega_n-\omega^2)}. \tag{7}$$

Impedance (7) was obtained under an assumption of perfect electromagnetic coupling between MEAS and the structure. In addition, this expression does not contain the impedance contributed by the sensing element. To incorporate both effects in the MMI model, a transformer and an associated equivalent electrical circuit are introduced. An equivalent electrical circuit, which includes inductance of the sensor L_{MEAS}, resistance of the sensor R_{MEAS}, inductance and resistance of the metallic structure L_S and R_S, and the dynamic impedance $Z_{\mathrm{str}}(\omega)$, is illustrated in Figure 2. In this circuit, generation of the elastic wave and the associated resonance phenom-

enon contribute into the diagram across points A-A'. Analysis of the transformer circuit leads to the following formulation for the cumulative impedance seen by the magneto-elastic sensor:

$$Z(\omega) = R_{\text{MEAS}} + i\omega L_{\text{MEAS}} + \frac{\omega^2 M^2}{i\omega L_S + R_S + Z_{\text{str}}(\omega)}. \tag{8}$$

In Eq. (8) the mutual inductance M accounts for the electromagnetic coupling between the sensor the structure. For practical calculations, it is convenient to consider a coupling coefficient ranging from 0 (no coupling) to 1 (perfect coupling) rather than the mutual inductance. In the theory of the electromagnetic coupling, the electromagnetic coupling coefficient k_C is presented as a ratio of the mutual inductance M to the square root of participating inductances L_{MEAS} and L_S,

$$k_C = M/\sqrt{L_S L_{\text{MEAS}}}. \tag{9}$$

Using the definition of the coupling coefficient (9), the expression for magneto-mechanical impedance (8) can be rewritten as follows:

$$Z(\omega) = R_{\text{MEAS}} + i\omega L_{\text{MEAS}} + \frac{\omega^2 L_{\text{MEAS}} L_S \cdot k_C^2}{i\omega L_S + R_S + Z_{\text{str}}(\omega)}. \tag{10}$$

Equation (10) describes the magneto-mechanical impedance measured at MEAS terminals and incorporates three distinct contributions: mechanical dynamic response via $Z_{\text{str}}(\omega)$, electromagnetic coupling, and sensor characteristics.

3. Experimental testing

Experiments were performed on the aluminum 2024T3 beam specimen with the following dimensions: length—304.8 mm, width—25.4 mm, thickness—1.587 mm. Exact material properties were not reported by manufacturer, therefore, approximate values for modulus of elasticity $E=73.1$ GPa and density $\rho=2780$ kg/m^3 were considered in this study. To imitate free-free boundary conditions, the beam was suspended in air using thin fishing line. MEAS utilized in this study consisted of a multilayer coil with approximately 500 turns of No. 30 AGW wire and a grade N42 nickel neodymium magnet of 3.2 mm thickness. The diameter of a magnet was 19.1 mm, which was slightly smaller than the diameter of the coil. Noticeable manifestation of the structural dynamic response was observed with the single magnet MEAS. However, the amplitude of the MMI peaks could be increases by utilizing more magnets in the sensor. Noncontact measurements with the sensor lift-off of 1 mm were conducted in two locations along the beam: at the midspan (1/2—length) and at 1/4 of beam's length. Terminals of the magneto-elastic active sensor were connected to HP 4192A impedance analyzer. The excitation signal supplied by the analyzer was swept in the lower kilohertz frequency range to obtain the MMI response of the aluminum beam. The raw data obtained in the experiment with the MEAS positioned at the 1/4 of beam's length are presented in Fig. 3 as a solid red line.

The analytical model discussed in the preceding section was utilized to calculate the magneto-mechanical impedance of the magneto-elastic sensor positioned next to the structural surface. The calculations were performed on a personal computer by coding Eqs. (7) and (10) in the mathematical software. Inductance and resistance of MEAS, $L_{\text{MEAS}}=1.9$ mH and $R_{\text{MEAS}}=10\ \Omega$, were determined experimentally from the impedance measurements of the sensor. Additional parameters participating in the model, L_S, R_S, and k_C, were estimated as $L_S=0.1\ \mu$H, $R_S=0.8\ \text{m}\Omega$ and $k_C=0.4$. Results of theoretical calculations are presented in Fig. 3 as a solid blue line. The figure indicates an agreement between experimental and theoretical magneto-mechanical impedances of the metallic beam. The position of impedance peaks match particularly well and the slope of the impedance curve is consistent with the experimental data. Minor discrepancies between theoretical and experimental MMI responses may be attributed to limitations of the considered one-dimensional structural model and idealizations associated with the sensor field distribution and strength of the electromagnetic coupling.

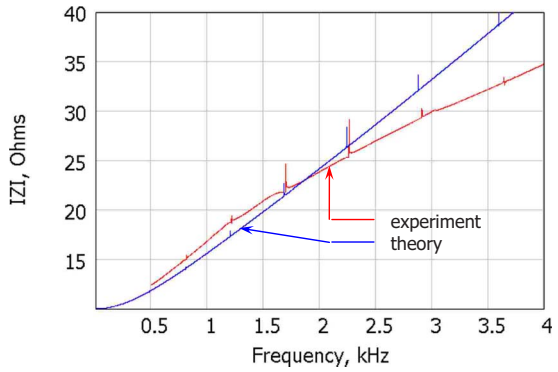


Fig. 3. (Color online) Theoretically calculated and experimentally measured MMI responses of an aluminum beam.

Manifestation of the structural dynamic response in the MMI signature has been validated in two tests with MEAS positioned at different locations underneath the beam. In these noncontact measurements, the gap between the sensor and the beam was kept constant while MEAS was positioned at the middle and at the 1/4 length of the beam. The MMI response of the free-free beam was measured in 0.3–5 kHz frequency range and the acquired data were processed to eliminate the inductive impedance slope. Figure 4 illustrates sensor positions and corresponding MMI responses. Noticeable in Figs. 4(b) and 4(c), distribution of impedance peaks depend on the position of the exciting/receiving transducer. This is consistent with structural dynamics theory suggesting that a vibration mode may not be excited if position of the excitation point coincides with modal nodes. To illustrate that this principle also holds in MMI

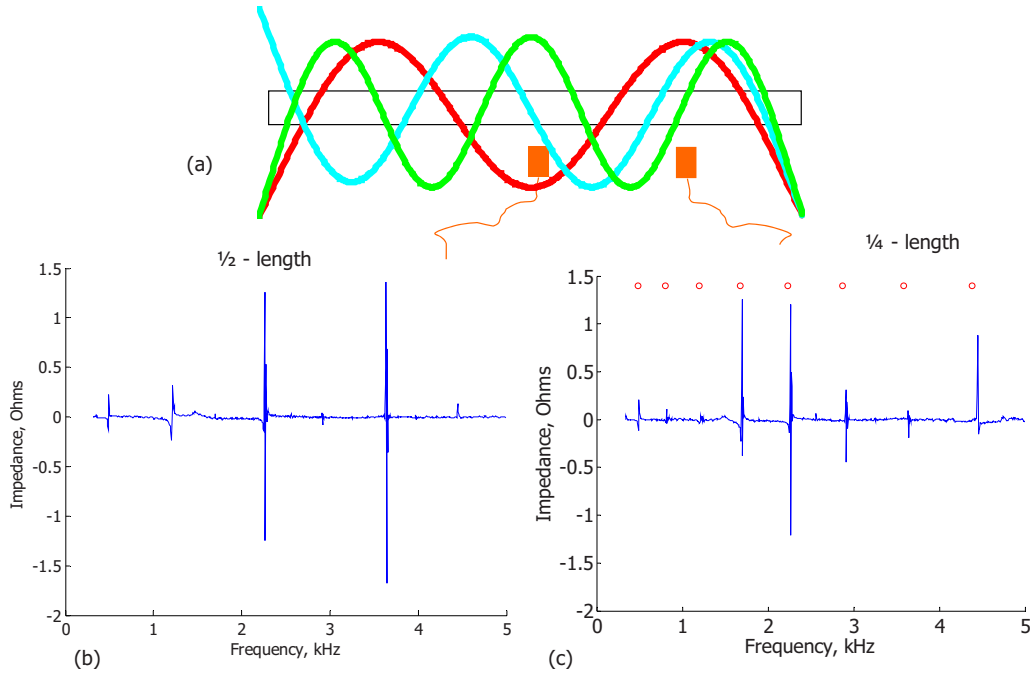


Fig. 4. (Color online) (a) Schematic of an aluminum beam showing several vibration modes and position of magneto-elastic active sensors. Magneto-mechanical impedances of the aluminum beam measured at the 1/2 (b) and at the 1/4 (c) of beam's length.

measurements, consider several vibration modes and their MMI manifestation. Calculated natural frequencies of third, fourth, fifth, and other modes are indicated with red dots in Fig. 4(c). The discrepancy between experimental and calculated values is below 2%. Spatial distributions of these modes and MMI measurement points are presented in Fig. 4(a). When MEAS is positioned at the middle of the beam, mostly odd modes are excited including the third and the fifth. Negligible amplitude of the fourth mode is observed. However, when the magneto-elastic sensor is positioned at the quarter of the beam's length, the third and the fourth modes are noticeable while the fifth mode is diminished. This experimental observation supports the notion of associating magneto-mechanical impedance peaks with structural dynamic response. As it can be seen from formulations (5), (7), and (10), the presented theoretical model accounts for the structural modal behavior and as such opens possibilities for modal analysis via MMI.

4. Conclusions

The presented work indicates that by sweeping excitation frequencies within wide band, a magneto-mechanical impedance signature of the structural element can be obtained, which incorporates electrical responses of both the transducer and the metallic element and the multimodal structural dynamic response. An analytical model of the MMI response has been developed, which explores the Lorentz force actuation mechanism in nonferromagnetic metals. The model accounts for multimodal vibrations of the structural element, electrical characteristics of the measuring sensor and electromagnetic interaction between the structure and the sensor. The model has been validated against experiments with results demonstrating good agreement between theoretical and experimental MMI responses. Theoretical and experimental studies suggest several important aspects of the MMI technology.

(a) It is known that amplitude of elastic waves generated via the Lorentz force mechanism is rather small. Efficiency of this excitation mechanism is at least an order of magnitude lower than excitation achievable with piezoelectric transducers of the same size. Although low efficiency of the magneto-elastic transformation may pose difficulties in applications involving relatively large structural elements, MMI responses with well-pronounced peaks can be obtained for metallic components of smaller size. Measurements may be noticeably improved by utilizing a larger sensor, adding more turns to the sensor coil (increases L_{MEAS}), considering stronger magnets (increases B), and augmenting the sensor/structure electromagnetic coupling.

(b) MMI response depends on position of the measuring transducer. Consistent with structural dynamics theory, the best excitation is achieved for vibration modes with maxima at the sensor location. Insignificant amplitude is observed for vibration modes with nodes coinciding with the sensor location. Therefore, MMI provides opportunities for non-contact structural dynamic identification and modal analysis.

Acknowledgment

The author would like to acknowledge funding through AFOSR Grant No. FA9550-08-1-0128 and support of New Mexico Institute of Mining and Technology in conducting this study.

References and links

ASME Colloquium on Mechanical Impedance Methods for Mechanical Vibrations, edited by R. Plunkett (ASME, New York, **1958**).

Banik, N. C., and Overhauser, A. W. (**1977**). "Electromagnetic generation of ultrasound in metals," Phys. Rev. B **16**(8), 3379–3388.

Johnson, W., Auld, B. A., and Alers, G. A. (**1994**). "Spectroscopy of resonant torsional modes in cylindrical rods using electromagnetic-acoustic transduction," J. Acoust. Soc. Am. **95**(3), 1413–1418.

Lange, Yu. V. (**1972**). "Characteristics of the impedance method of inspection," Soviet Journal of Nondestructive Testing (English translation of Defektoskopiya), **8**, 47–54.

Liang, C., Sun, F. P., and Rogers, C. A. (**1994**). "Coupled electro-mechanical analysis of adaptive material system-determination of the actuator power consumption and system energy transfer," J. Intell. Mater. Syst. Struct. **5**, 12–20.

Skudrzyk, E. (**1968**). *Simple and Complex Vibratory Systems* (University Press, **1968**).
Turner, R., Lyall, K. R., and Cochran, J. F. (**1969**). "Generation and detection of acoustic waves in metals by means of electromagnetic radiation," Can. J. Phys. **47**(21), 2293–2301.
Zagrai, A. N., and Cakan, H. (**2007**). "Magneto-mechanical impedance technique for dynamic identification of metallic structures," Proceedings of 6th International Workshop on Structural Health Monitoring, September 11–13, 2007, Stanford University, CA, pp. 1693–1700.

Modulation frequency features for phoneme recognition in noisy speech

Sriram Ganapathy, Samuel Thomas, and Hynek Hermansky
Idiap Research Institute, Rue Marconi 19, 1920 Martigny, Switzerland and
Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
ganapathy@idiap.ch, tsamuel@idiap.ch, hermansky@ieee.org

Abstract: In this letter, a new feature extraction technique based on modulation spectrum derived from syllable-length segments of subband temporal envelopes is proposed. These subband envelopes are derived from autoregressive modeling of Hilbert envelopes of the signal in critical bands, processed by both a static (logarithmic) and a dynamic (adaptive loops) compression. These features are then used for machine recognition of phonemes in telephone speech. Without degrading the performance in clean conditions, the proposed features show significant improvements compared to other state-of-the-art speech analysis techniques. In addition to the overall phoneme recognition rates, the performance with broad phonetic classes is reported.

PACS numbers: 43.72.Ne, 43.72.Ar [DO]
Date Received: August 21, 2008 **Date Accepted:** October 15, 2008

1. Introduction

Conventional speech analysis techniques start with estimating the spectral content of relatively short (about 10–20 ms) segments of the signal (short-term spectrum). Each estimated vector of spectral energies represents a sample of the underlying dynamic process in production of speech at a given time frame. Stacking such estimates of the short-term spectra in time provides a two-dimensional (time-frequency) representation of speech that represents the basis of most speech features [for example (Hermansky, 1990)]. Alternatively, one can directly estimate trajectories of spectral energies in the individual frequency subbands, each estimated vector then representing the underlying dynamic process in a given subband. Such estimates, stacked in frequency, also forms a two-dimensional representation of speech [for example (Athineos *et al.*, 2004)].

For machine recognition of phonemes in noisy speech, the techniques that are based on deriving long-term modulation frequencies do not preserve fine temporal events like onsets and offsets which are important in separating some phoneme classes. On the other hand, signal adaptive techniques which try to represent local temporal fluctuation, cause strong attenuation of higher modulation frequencies which makes them less effective even in clean speech (Tchorz and Kollmeier, 1999).

In this letter, we propose a feature extraction technique for phoneme recognition that tries to capture fine temporal dynamics along with static modulations using subband temporal envelopes. The input speech signal is decomposed into 17 critical bands (Bark scale decomposition) and long temporal envelopes of subband signals are extracted using the technique of frequency domain linear prediction (FDLP) (Athineos and Ellis, 2007). The subband temporal envelopes of the speech signal are then processed by a static compression stage and a dynamic compression stage. The static compression stage is a logarithmic operation and the adaptive compression stage uses the adaptive compression loops proposed in Dau *et al.* (1996). The compressed subband envelopes are transformed into modulation frequency components and used as features for hybrid hidden Markov model-artificial neural network (HMM-ANN) phoneme recognition system (Bourlard and Morgan, 1994). The proposed technique yields more

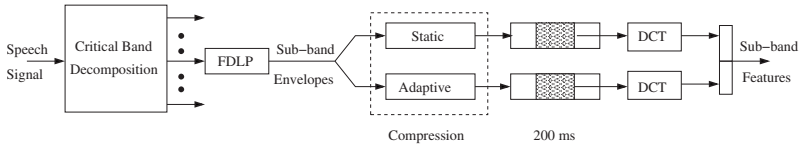


Fig. 1. Block schematic for the subband feature extraction. The steps involved are critical band decomposition, estimation of subband envelopes using FDLP, static and adaptive compression, and conversion to modulation frequency components by the application of cosine transform.

accurate estimates of phonetic values of the speech sounds than several other state-of-the-art speech analysis techniques. Moreover, these estimates are much less influenced by distortions induced by the varying communication channels.

2. Feature extraction

The block schematic for the proposed feature extraction technique is shown in Fig. 1. Long segments of speech signal are analyzed in critical bands using the technique of FDLP (Athineos and Ellis, 2007). FDLP forms an efficient method for obtaining smoothed, minimum phase, parametric models of temporal rather than spectral envelopes. Being an autoregressive modeling technique, FDLP captures the high signal-to-noise ratio peaks in the temporal envelope. The whole set of subband temporal envelopes, which are obtained by the application of FDLP on individual subband signals, forms a two dimensional (time-frequency) representation of the input signal energy.

The subband temporal envelopes are then compressed using a static compression scheme which is a logarithmic function and a dynamic compression scheme (Dau *et al.*, 1996). The use of the logarithm is to model the overall nonlinear compression in the auditory system which covers the huge dynamical range between the hearing threshold and the uncomfortable loudness level. The adaptive compression is realized by an adaptation circuit consisting of five consecutive nonlinear adaptation loops (Dau *et al.*, 1996). Each of these loops consists of a divider and a low-pass filter with time constants ranging from 5 to 500 ms. The input signal is divided by the output signal of the low-pass filter in each adaptation loop. Sudden transitions in the subband envelope that are very fast compared to the time constants of the adaptation loops are amplified linearly at the output due to the slow changes in the low pass filter output, whereas the slowly changing regions of the input signal are compressed. This is illustrated in Fig. 2, which shows (a) a portion of 1000 ms of full-band speech signal, (b) the temporal envelope extracted using the Hilbert transform, (c) the FDLP envelope, which is an all-pole approximation to (b) estimated using FDLP, (d) logarithmic compression of the FDLP envelope, and (e) adaptive compression of the FDLP envelope.

Conventional speech recognizers require speech features sampled at 100 Hz (i.e., one feature vector every 10 ms). For using our speech representation in a conventional recognizer, the compressed temporal envelopes are divided into 200 ms segments with a shift of 10 ms. Discrete cosine transform of both the static and the dynamic segments of temporal envelope yields the static and the dynamic modulation spectrum respectively. We use 14 modulation frequency components from each cosine transform, yielding modulation spectrum in the 0–70 Hz region with a resolution of 5 Hz. This choice is a result of series of optimization experiments (which are not reported here).

3. Experiments and results

The proposed features are used for a phoneme recognition task on the HTIMIT database (Reynolds, 1997). We use a phoneme recognition system based on the HMM-ANN paradigm (Bourlard and Morgan, 1994) trained on clean speech using the TIMIT database downsampled to 8 kHz. The training data consists of 3000 utterances from 375 speakers, cross-validation data set consists of 696 utterances from 87 speakers, and the test data set consists of 1344 utterances from 168 speakers. The TIMIT database, which is hand-labeled using 61 labels is mapped to the

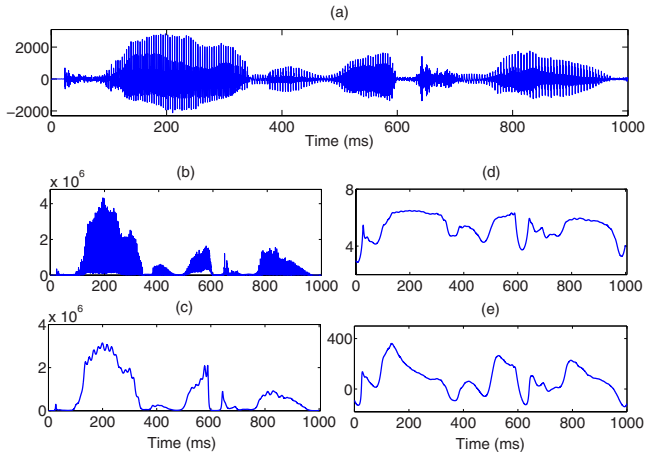


Fig. 2. (Color online) Static and dynamic compression of the temporal envelopes: (a) a portion of 1000 ms of full-band speech signal, (b) the temporal envelope extracted using the Hilbert transform, (c) the FDLP envelope, which is an all-pole approximation to (b) estimated using FDLP, (d) logarithmic compression of the FDLP envelope, and (e) adaptive compression of the FDLP envelope.

standard set of 39 phonemes (Pinto *et al.*, 2007). For phoneme recognition experiments in telephone channel, speech data collected from nine telephone sets in the HTIMIT database are used, which introduce a variety of channel distortions in the test signal. For each of these telephone channels, 842 test utterances, also having clean recordings in the TIMIT test set, are used. The system is trained only on the original TIMIT data, representing clean speech without the distortions introduced by the communication channel but tested on the clean TIMIT test set as well as the HTIMIT degraded speech.

The results for the proposed technique are compared with those obtained for several other robust feature extraction techniques namely RASTA (Hermansky and Morgan, 1994), auditory model based front-end (Old.) (Tchorz and Kollmeier, 1999), Multiresolution RASTA (MRASTA) (Hermansky and Fousek, 2005), and the advanced-ETSI (noise-robust) distributed speech recognition front-end (ETSI, 2002). The results of these experiments on the clean test conditions are shown in the top panel of Table 1. The conventional perceptual linear prediction (PLP) feature extraction used with a context of nine frames (Pinto *et al.*, 2007) is denoted as PLP-9. RASTA-PLP-9 features use a nine frame context of the PLP features extracted after applying the RASTA filtering (Hermansky and Morgan, 1994). Old.-9 refers to the nine frame context of the auditory model based front-end reported in (Tchorz and Kollmeier, 1999). The ETSI-9 corresponds to a nine frame context of the features generated by the ETSI front-end. The FDLP features derived using static, dynamic and combined (static and dynamic) compres-

Table 1. Recognition accuracies (%) of individual phonemes for different feature extraction techniques on clean and telephone speech.

Clean speech							
PLP-9	R-PLP-9	Old.-9	MRASTA	ETSI-9	FDLP-Stat.	FDLP-Dyn.	FDLP-Comb.
64.9	61.2	60.3	63.9	63.1	63.1	59.7	65.4
Telephone speech							
PLP-9	R-PLP-9	Old.-9	MRASTA	ETSI-9	FDLP-Stat.	FDLP-Dyn.	FDLP-Comb.
34.4	46.2	45.3	47.5	47.7	50.8	48.7	52.7

Table 2. Recognition accuracies (%) of broad phonetic classes obtained from confusion matrix analysis on clean and telephone speech.

Clean speech					
Class	PLP-9	MRASTA	FDLP-Stat.	FDLP-Dyn.	FDLP-Comb.
Vowel	83.3	81.9	82.7	81.3	83.8
Diphthong	75.1	73.0	70.7	67.9	74.2
Plosive	81.6	80.5	79.5	78.2	81.6
Affricative	69.1	68.8	64.6	62.5	69.9
Fricative	81.8	80.1	80.0	77.8	81.9
Semivowel	72.2	71.6	70.7	69.5	73.5
Nasal	80.4	79.2	80.8	77.7	82.4
Telephone speech					
Class	PLP-9	MRASTA	FDLP-Stat.	FDLP-Dyn.	FDLP-Comb.
Vowel	61.1	74.2	77.5	77.6	79.8
Diphthong	51.1	68.2	63.4	61.7	67.2
Plosive	46.9	52.5	56.1	59.0	59.0
Affricative	28.0	38.5	35.7	36.9	39.8
Fricative	63.3	70.7	78.5	74.0	79.4
Semivowel	55.8	61.3	60.5	60.7	63.8
Nasal	35.4	57.7	66.6	64.9	68.7

sion are denoted as FDLP-Stat., FDLP-Dyn., and FDLP-Comb. respectively (Sec. 2). The performance on clean conditions for the FDLP-Dyn. and Old.-9 features validates the claim in Tchorz and Kollmeier (1999) regarding the effects of the distortions introduced by adaptive compression model on the higher signal modulations. The experiments on clean conditions also illustrate the gain obtained by the combination of the static and dynamic modulation spectrum for phoneme recognition. The bottom panel of Table 1 shows the average phoneme recognition accuracy [100—PER, where PER is the phoneme error rate (Pinto *et al.*, 2007)] for all the nine telephone channels. The proposed features, on the average, provide a relative error improvement of about 10% over the other feature extraction techniques considered.

4. Discussion

Table 2 shows the recognition accuracies of broad phoneme classes for the proposed feature extraction technique along with a few other speech analysis techniques. For clean conditions, the proposed features (FDLP-Comb.) provide phoneme recognition accuracies that are competent with other feature extraction techniques for all the phoneme classes. In the presence of telephone noise, the FDLP-Stat. features provide significant robustness for fricatives and nasals (which is due to modeling property of the signal peaks in static compression) whereas the FDLP-Dyn. features provide good robustness for plosives and affricates (where the fine temporal fluctuations like onsets and offsets carry the important phoneme classification information). Hence, the combination of these feature streams results in considerable improvement in performance for most of the broad phonetic classes.

5. Summary

We have proposed a feature extraction technique based on the modulation spectrum. Subband temporal envelopes, estimated using FDLP, are processed by both a static and a dynamic compression and are converted to modulation frequency features. These features provide good robustness properties for phoneme recognition tasks in telephone speech.

Acknowledgments

This work was supported by the European Union 6th FWP IST Integrated Project AMIDA and the Swiss National Science Foundation through the Swiss NCCR on IM2. The authors would like to thank the Medical Physics group at the Carl von Ossietzky-Universitat Oldenburg for code fragments implementing adaptive compression loops.

References and links

Athineos, M., and Ellis, D. P. W. (**2007**). "Autoregressive modelling of temporal envelopes," IEEE Trans. Signal Process. **55**(11), 5237–5245.

Athineos, M., Hermansky, H., and Ellis, D. P. W. (**2004**). "LP-TRAPS: Linear predictive temporal patterns," Proceedings of INTERSPEECH, pp. 1154–1157.

Bourlard, H., and Morgan, N. (**1994**). *Connectionist Speech Recognition—A Hybrid approach* (Kluwer Academic, Dordrecht).

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the "effective" signal processing in the auditory system: I. Model structure," J. Acoust. Soc. Am. **99**(6), 3615–3622.

ETSI (**2002**). "ETSI ES 202 050 v1.1.1 STQ; Distributed speech recognition; Advanced front-end feature extraction algorithm; Compression algorithms."

Hermansky, H. (**1990**). "Perceptual linear predictive (PLP) analysis of speech," J. Acoust. Soc. Am. **87**(4), 1738–1752.

Hermansky, H., and Fousek, P. (**2005**). "Multi-resolution RASTA filtering for TANDEM-based ASR," Proceedings of INTERSPEECH, pp. 361–364.

Hermansky, H., and Morgan, N. (**1994**). "RASTA processing of speech," IEEE Trans. Speech Audio Process. **2**, 578–589.

Pinto, J., Yegnanarayana, B., Hermansky, H., and Doss, M. M. (**2007**). "Exploiting contextual information for improved phoneme recognition," Proceedings of INTERSPEECH, pp. 1817–1820.

Reynolds, D. A. (**1997**). "HTIMIT and LLHDB: speech corpora for the study of hand set transducer effects," Proceedings of ICASSP, pp. 1535–1538.

Tchorz, J., and Kollmeier, B. (**1999**). "A model of auditory perception as front end for automatic speech recognition," J. Acoust. Soc. Am. **106**(4), 2040–2050.

Time-forward speech intelligibility in time-reversed rooms

Laricia Longworth-Reed, Eugene Brandewie, and Pavel Zahorik
Department of Psychological and Brain Science, University of Louisville, Louisville, Kentucky 40292
pavel.zahorik@louisville.edu

Abstract: The effects of time-reversed room acoustics on word recognition abilities were examined using virtual auditory space techniques, which allowed for temporal manipulation of the room acoustics independent of the speech source signals. Two acoustical conditions were tested: one in which room acoustics were simulated in a realistic time-forward fashion and one in which the room acoustics were reversed in time, causing reverberation and acoustic reflections to precede the direct-path energy. Significant decreases in speech intelligibility—from 89% on average to less than 25%—were observed between the time-forward and time-reversed rooms. This result is not predictable using standard methods for estimating speech intelligibility based on the modulation transfer function of the room. It may instead be due to increased degradation of onset information in the speech signals when room acoustics are time-reversed.

PACS numbers: 43.66.Lj, 43.71.Gv, 43.55.Hy [Q-JF]
Date Received: August 20, 2008 **Date Accepted:** October 24, 2008

1. Introduction

Classic demonstrations have revealed that backwards playback of speech recordings in reverberant environments results in greatly increased audibility of the acoustic reflections and reverberation (Houtsma *et al.*, 1987). Unfortunately, this time-reversal manipulation also causes the speech source signal to become unintelligible, since it too is reversed in time. Objective measures of speech intelligibility would therefore not be valid measures of the time-reversal effects of room acoustics in such demonstrations. Here, an improved method using virtual auditory space (VAS) techniques to manipulate temporal aspects of (simulated) room acoustics independent of the speech source signals is described and tested. Speech intelligibility was measured for normal time-forward speech signals in two acoustical conditions: one in which room acoustics were simulated in a realistic time-forward fashion; and one in which the room acoustics were reversed in time, causing reverberation and acoustic reflections to precede the direct-path energy. This manipulation is interesting because it represents an extreme listening situation with which listeners have likely had very little experience. It also does not affect the modulation transfer function (MTF) of the room, which describes the way the room modifies the temporal energy distribution of a sound signal and has been shown to be highly predictive of speech intelligibility in rooms (Houtgast and Steeneken, 1973, 1985).

Of additional interest is the extent to which listeners might adapt to realistic "plausible" listening environments (Clifton *et al.*, 1994; Hartmann, 1997) and not to "implausible" time-reversed environments, as well as the impact of realistic binaural input on these effects. Results by Watkins (2005b) indicate that the prior listening exposure in reverberant environments returns performance in a categorical speech perception task to levels consistent with those observed in much less reverberant listening situations, presumably through a process of perceptual adaptation to the listening environment. The adaptation is decreased when the acoustics of the prior listening context are made implausible through time-reversal (Watkins, 2005b, 2005a), and does not appear to depend on binaural input (Watkins, 2005b), although binaural listening is in general known to facilitate speech intelligibility in reverberant environments (Nabelek and Robinson, 1982).

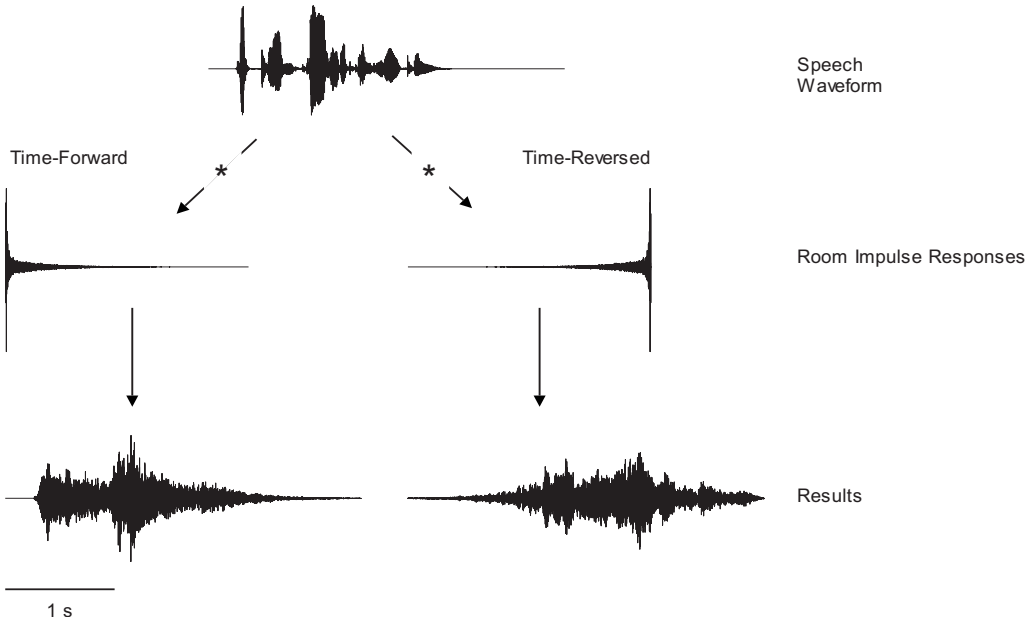


Fig. 1. Stimulus generation scheme for time-forward and time-reversed binaural room impulse responses (left ear shown). Results displayed are from Mm1 and Mm2.

2. Methods

2.1 Listeners

Ten listeners (four female) ages 19–51 years participated in the experiment. All had normal hearing, as verified by standard (ANSI-S3.9, 1989) audiometric screening.

2.2 Stimuli

VAS techniques were used to simulate a reverberant room with dimensions of $5.7 \times 4.3 \times 2.6$ m and broadband reverberation time (T_{60}) of approximately 1.5 s. To implement the simulation, a simple model of a binaural room impulse response (BRIR) was constructed using an image-model (Allen and Berkley, 1979) to simulate early reflections and a statistical model of the late reverberant energy. The direct-path and 500 early reflections were all spatially rendered using head-related transfer functions measured from 613 spatial locations surrounding a single representative listener in an anechoic chamber, using methods fundamentally similar to those described in detail by Wightman and Kistler (1989). This listener did not participate in any subsequent behavioral testing. Each reflection was attenuated based on path-length, an average surface absorption coefficient, and the reflection order. Diffuse late reverberation was simulated in the BRIR using independent Gaussian noise samples for each ear shaped by decay functions derived from the Sabine equation (Sabine, 1922) in each of six octave bands ranging from 125 to 4000 Hz. In general, this simulation technique is similar to those implemented in other studies (Heinz, 1993; Naylor, 1993) and has been shown to produce simulations that are perceptually similar to those derived from measurements in a real room (Zahorik, 2004).

Time-forward and time-reversed BRIRs were convolved (using MATLAB® software) with time-forward sentences (adult talker) from the Hearing in Noise Test (HINT) sentence corpus (Nilsson *et al.*, 1994). The simulated spatial location of the speech source was 1.4 m in front of the listener at ear level, in the approximate center of the virtual room. A graphical description of this stimulus generation procedure is shown in Fig. 1. Multimedia examples of the time-forward and time-reversed stimuli shown in Fig. 1 are available (Mm1 and Mm2),

along with analogous stimuli generated using a different sentence (Mm3 and Mm4).

Mm. 1. Example of a sentence ("The tub faucet is leaking") convolved with a time-forward BRIR. This is a file of type "wav" (856 KB).

Mm. 2. Example of a sentence ("The tub faucet is leaking") convolved with a time-reversed BRIR. This is a file of type "wav" (856 KB).

Mm. 3. Example of a sentence ("The truck drove up the road") convolved with a time-forward BRIR. This is a file of type "wav" (793 KB).

Mm. 4. Example of a sentence ("The truck drove up the road") convolved with a time-reversed BRIR. This is a file of type "wav" (793 KB).

In addition to binaural simulations that would be representative of those received in a real-room listening situation, diotic presentations—in which the simulated results for each ear were summed and delivered to both ears—were tested to determine the extent to which natural binaural information is important for speech recognition abilities in the simulated listening environments.

All stimuli were presented over equalized Beyerdynamic DT-990-Pro headphones using a Digital Audio Labs CardDeluxe for D/A conversion (24-bit, 44.1 kHz) at moderate level (approximately 65 dB SPL) within a double-walled sound isolation chamber (Acoustic Systems).

2.3 Design and procedure

Listeners were tested in both binaural and diotic presentation conditions of the two room acoustic conditions: time-forward and time-reversed. The four resulting conditions were run in blocks, with order counterbalanced. Subjects were presented with each sentence stimulus one time and asked to repeat as many words in the sentence as possible. Forty sentences were presented in two 20-sentence blocks for each condition. Sentence recognition scores (proportion of correct words within a given sentence) were computed for each sentence based on the subject's responses. Subjects also completed an initial baseline condition in which ten sentences were presented diotically in the absence of any room acoustic processing.

3. Results

3.1 MTF analyses

MTFs were computed for both time-forward and time-reverse BRIRs in octave bands with center frequencies ranging from 125 Hz to 8 kHz, following methods described by Steeneken and Houtgast (1980). Figure 2 displays MTF results for time-forward (filled symbols) and time-reversed (open symbols) impulse responses in the 1-octave band centered at 1 kHz. The solid curve represents a theoretical MTF (Houtgast and Steeneken, 1985) for a room with 1.5 s reverberation time (T_{60}), which has a low-pass magnitude characteristic. Close agreement may be observed between most measured and predicted values, indicating that time reversal of the impulse response does not affect the MTF. Similar results were observed in other octave bands (not shown).

The Speech Transmission Index (STI) integrates results from MTFs measured in octave bands from 125 Hz to 8 kHz (Houtgast and Steeneken, 1985) and indicates the amount of temporal modulation reduction imposed by the room (ranging from 0 "complete reduction" to 1 "no reduction"). As a result, high STI values have been shown to correspond to high levels of speech intelligibility (Houtgast and Steeneken, 1973, 1985). Since STI is derived from the MTF, it too should be unaffected by the time-reversal manipulation. STI values were computed for both the time-forward and time-reversed room impulse responses using one of the HINT sentence stimuli. As predicted, the values were very similar: 0.4549 (time-forward) and 0.4325 (time-reversed). Previous results indicate that these STI values would produce approximately 85% correct identification in a CVC-type task (Houtgast and Steeneken, 1973).

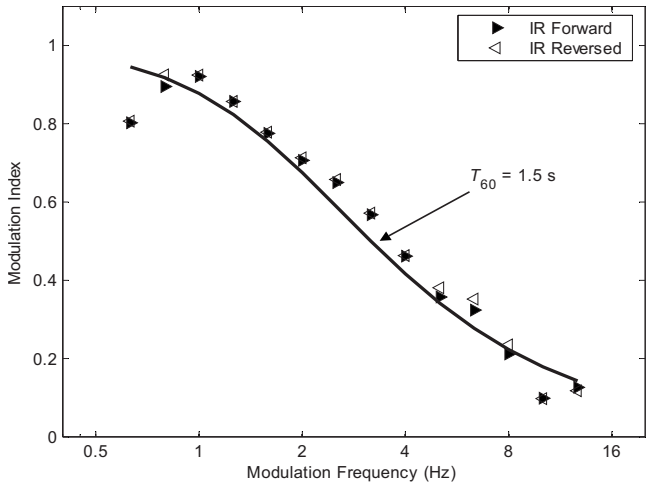


Fig. 2. MTF for the 1-octave band centered at 1 kHz for both time-forward (filled symbols) and time-reversed (open symbols) binaural room impulse responses. Data for only the left ear are shown. The solid curve represents a theoretical MTF (Houtgast and Steeneken, 1985) for a room with 1.5 s reverberation time (T_{60}).

It is important to note that past study of the MTF/STI concept in relation to speech intelligibility and room acoustics has considered only the magnitude characteristics of MTFs (Houtgast and Steeneken, 1973, 1985), yet MTFs are technically complex functions with both magnitude and phase characteristics. Although it has been demonstrated that the time-reversal manipulation in this study does not affect the magnitude characteristic of the MTFs (or the STI), the phase characteristics of the MTFs were dramatically affected by this manipulation.

3.2 Word recognition

Figure 3(a) displays the mean word recognition scores for each of the experimental conditions. All subjects performed predictably well in the baseline condition (no room simulation): correctly identifying all the words in the sentences presented. The addition of normal time-forward acoustics of a reverberant room reduced word recognition scores to approximately 89% on average. This result is roughly consistent with the 85% speech intelligibility prediction based on

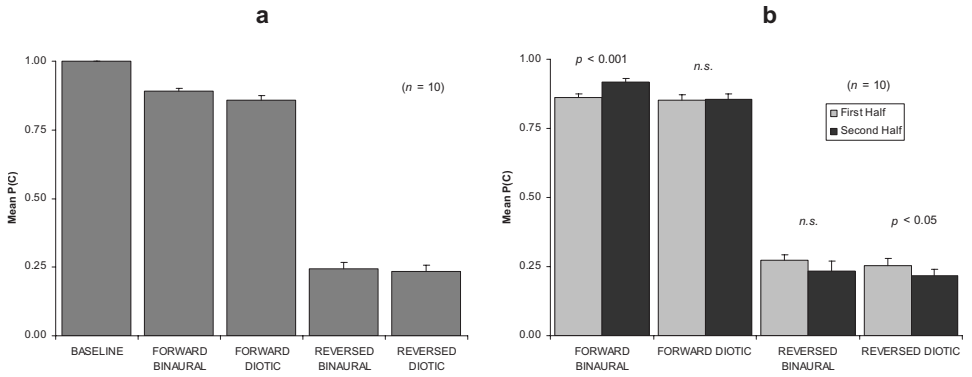


Fig. 3. Mean word recognition scores, $P(C)$, for all experimental conditions. Results are displayed based on all trials within an experimental block (a), and separately based on trails within the first half and second half of a block (b). This latter analysis (b), which allows performance changes within a block of trials to be assessed (results of significance testing shown), was not conducted on the baseline condition (no room simulation) due to the small number of trials collected for this condition. Error bars indicate 1 SEM.

the STI for the simulated room. Sentence recognition scores were dramatically reduced for the time-reversed acoustics (approximately 25% correct on average) relative to the time-forward acoustics. A repeated-measures ANOVA on the arcsine-transformed (Kirk, 1982) word recognition scores confirmed that this difference was highly significant, $F(1,9)=1055$, $p<0.0001$, $\eta^2=0.992$. These effects of time-reversing the room impulse response are not explainable on the basis of MTF-type analyses, because the time-forward and time-reversed conditions yielded nearly identical STI values. The effect is also readily apparent under informal listening situations, such as those demonstrated in Mm1 versus Mm2 and Mm3 versus Mm4.

Slight, but statistically significant, increases in performance were also observed (Fig. 3(a)) for binaural listening conditions relative to diotic listening, $F(1,9)=5.35$, $p<0.05$, $\eta^2=0.373$. This result is consistent with previous literature (Nabelek and Robinson, 1982).

To test for potential room adaptation effects, we compared average performance for the first ten sentences presented in each block of trials versus the last ten sentences on a listener-by-listener basis. If room adaptation takes place over the course of a trial block in which repeated listening from the same acoustic environment takes place, then word recognition performance should also improve during the trial block. This result was observed for the time-forward binaural condition only, where word recognition scores improved by nearly six percentage points (see Fig. 3(b)). This improvement was highly significant, $F(1,9)=31.23$, $p<0.001$ (Bonferonni correction), $\eta^2=0.776$, and is consistent with the notion of an environmental adaptation effect in which familiarity with a natural acoustic environment can aid in speech intelligibility performance (Watkins, 2005b, 2005a). All other conditions show either no change, or else a small decrease, $F(1,9)=5.67$, $p<0.05$ (Bonferonni correction), $\eta^2=0.386$, in performance between first and last ten sentences. These latter results are inconsistent with simple practice effects causing the improvement in performance for the time-forward binaural condition.

3.3 Onset analysis

One potential explanation for the dramatic decrease in intelligibility between time-forward and time-reversed room acoustics is that the latter disrupts onset information in the speech stimulus much more than the former. To examine this possibility, we quantified the onset information present in the various sentence stimuli (full 8 kHz bandwidth) used in this experiment by first computing the instantaneous slope (1st derivative) of the time waveform's amplitude envelope (extracted using a second-order Butterworth low-pass filter with 10 Hz cutoff) and then taking the mean of all positive slope values in each sentence. Larger values of this measure generally indicate more rapid amplitude onsets in the time waveform. Figure 4 displays the distributions of this onset measure for each of 170 sentences with either no processing, time-forward room processing (left ear only), or time-reversed room processing (left ear only). Mean onset measures for the three conditions (matched-observations) were significantly different, $F(2,338)=1676$, $p<0.0001$, $\eta^2=0.908$, and strongly correlated with mean word recognition performance in the three conditions ($r=0.88$).

4. Discussions and conclusions

These results demonstrate that speech intelligibility is dramatically degraded when reverberation and echoes precede direct path energy in simulated room environments. Although the precise cause of the intelligibility degradation is not known, it cannot be explained by MTF-type analyses of the rooms and does not appear to depend critically on binaural input. It may instead be related to the resulting degradation of onset information in the speech signals, which has been argued to be particularly important for speech intelligibility, from both computational (Régnier and Allen, 2008) and neurophysiological perspectives (Heil, 2003). Further study will be needed to more fully test this onset hypothesis, using stimulus conditions more conducive to evaluating speech onset information, such as consonant-vowel pairs.

Additional results from this study are suggestive of adaptive processes that mediate the contributions of room acoustics under plausible time-forward conditions (Clifton *et al.*, 1994;

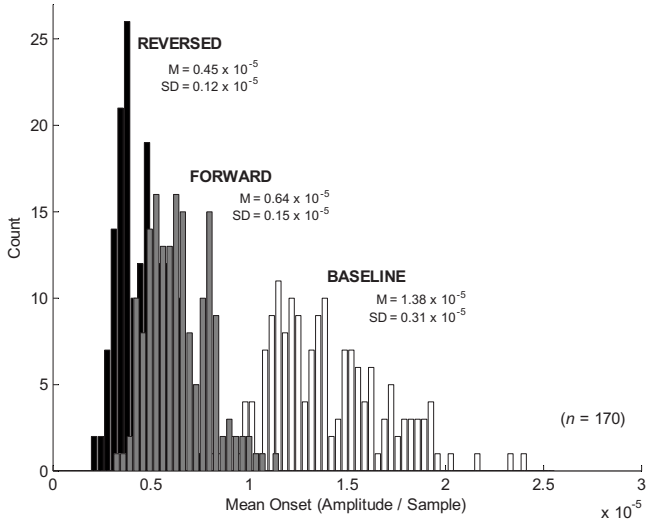


Fig. 4. Distributions of onset measures in the speech stimuli for three different room acoustic manipulations: base line—no room (white), time-forward room (grey), time-reversed room (black). Means (M) and standard deviations (SD) are shown for each distribution.

Hartmann, 1997), but not in implausible time-reversed conditions. Although consistent with results of Watkins (2005b, 2005a), further research will be required to more fully understand the conditions under which the adaptation occurs, its potential mechanisms, and its relationship to other adaptation phenomena in acoustically reflective environments that relate to apparent sound direction (Freyman *et al.*, 1991).

Acknowledgments

Work supported by NIH-NIDCD (R01DC008168).

References and links

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.
ANSI-S3.9 (**1989**). *American National Standard Specification for Audiometers* (American National Standards Institute, New York).
Clifton, R. K., Freyman, R. L., Litovsky, R. Y., and McCall, D. (**1994**). "Listeners' expectations about echoes can raise or lower echo threshold," J. Acoust. Soc. Am. **95**, 1525–1533.
Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (**1991**). "Dynamic processes in the precedence effect," J. Acoust. Soc. Am. **90**, 874–884.
Hartmann, W. M. (**1997**). "Listening in a room and the precedence effect," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), pp. 191–210.
Heil, P. (**2003**). "Coding of temporal onset envelope in the auditory system," Speech Commun. **41**, 123–134.
Heinz, R. (**1993**). "Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail," Appl. Acoust. **28**, 145–159.
Houtgast, T., and Steeneken, H. J. M. (**1973**). "The modulation transfer function in room acoustics as a predictor of speech intelligibility," Acustica **28**, 66–73.
Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.
Houtsma, A. J. M., Rossing, T. D., and Wagenaars, W. M. (**1987**). "Effects of echoes," in *Auditory Demonstrations* (Institute for Perception Research, Eindhoven, The Netherlands).
Kirk, R. E. (**1982**). *Experimental Design: Procedures for the Behavioral Sciences* (Brooks/Cole, Monterey, CA).
Nabelek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.
Naylor, G. M. (**1993**). "ODEON—Another hybrid room acoustic model," Appl. Acoust. **38**, 131–143.
Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the Hearing in Noise Test for the measurement

of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.
Régnier, M. S., and Allen, J. B. (**2008**). "A method to identify noise-robust perceptual features: application for consonant/t," J. Acoust. Soc. Am. **123**, 2801–2814.
Sabine, W. C. (**1922**). "Reverberation," in *Collected Papers on Acoustics* (Harvard University Press, Cambridge, MA).
Steeneken, H. J., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326.
Watkins, A. J. (**2005a**). "Perceptual compensation for effects of echo and of reverberation on speech identification," Acust. Acta Acust. **91**, 892–901.
Watkins, A. J. (**2005b**). "Perceptual compensation for effects of reverberation in speech identification," J. Acoust. Soc. Am. **118**, 249–262.
Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening: I. Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.
Zahorik, P. (**2004**). "Perceptual scaling of room reverberation," J. Acoust. Soc. Am. **115**, 2598.

Sequential effects on the detectability of a tone added to a multitone masker

Xiang Cao
Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104
xiangcao@seas.upenn.edu

Rong Huang and Virginia M. Richards
Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104
rongh@sas.upenn.edu; richards@psych.upenn.edu

Abstract: The detectability of a tone added to a masker is superior when the detection trial is preceded by the masker than the signal-plus-masker. This auditory enhancement can withstand long temporal gaps between the precursor and the trial, suggesting that for yes/no trials sensitivity may depend on the stimulus presented in the prior trial. The results from an experiment examining the detectability of a 1000-Hz tone added to 6-tone maskers confirmed sequential effects on sensitivity. The values of d' were higher when the prior trial was a no-signal (masker alone) trial compared to a signal (signal-plus-masker) trial.

PACS numbers: 43.66.Dc, 43.66.Mk [Q-JF]
Date Received: September 18, 2008 **Date Accepted:** October 24, 2008

1. Introduction

The detectability of a tone added to a masker can be substantially improved if the signal's onset is delayed relative to the masker's onset (e.g., Green, 1969). For relatively long-duration signal tones (hundreds of ms) delaying the signal onset relative to the masker onset can lead to threshold reductions of 10–20 dB if the masker and signal are not spectrally overlapping (e.g., Viemeister, 1980; McFadden and Wright, 1990; Wright *et al.*, 1993; Richards *et al.*, 2004; see also McFadden and Wright, 1992, for an example of reduced thresholds when there is masker energy at the signal frequency). Here, this phenomenon is described using a single term, enhancement, even though experiments in which this phenomenon has been demonstrated include experiments that appear to depend relatively more or relatively less on peripheral versus central limitations/processes (e.g., multiple narrowband noise maskers, McFadden and Wright, 1990; Wright *et al.*, 1993; noise maskers, Viemeister, 1980; multitone maskers, Viemeister, 1980; versus maskers composed of harmonics of a common fundamental frequency, Viemeister, 1980; multitone informational maskers, Richards and Neff, 2004; Richards *et al.*, 2004).

Enhancement occurs not only when the signal onset is delayed relative to the masker onset, but also when there is a temporal gap between the early masker and the onset of the simultaneously presented masker and signal (when presented) (e.g., Viemeister, 1980). Here, we refer to the early masker as the precursor. For a masker and signal that shared a common fundamental frequency, Viemeister (1980) found that enhancement withstood precursor-trial delays as long as 6 s. Indeed, owing to the persistence of the enhancement effect, Viemeister (1980) pointed out that the phenomenon does not lend itself to study using traditional psychophysical methods such as the two-interval, two-alternative forced-choice procedure. In a similar vein, in an informational masking task, Richards *et al.* (2004) decomposed same-different trials separated by several hundred ms and discovered that the resulting "virtual" single-interval trials preceded by a masker yielded higher values of d' than trials preceded by a signal-plus-masker. Also with regard to informational maskers, Lutfi and Alexander (2005) observed context effects such that the detectability of a tone in quiet could decrease when tone-in-quiet trials were intermixed with informational masking trials. This result may ultimately be found to reflect sequen-

tial trial-by-trial interactions similar to enhancement, although that hypothesis was not examined in their paper. With regard to experiments likely to reflect relatively more peripheral limitations, and again for relatively long signal durations (60 ms), McFadden and Wright (1990) obtained, for some subjects, enhancement that withstood at least a 350 ms delay between the precursor and the stimulus interval. In their experiment the signal to be detected was a narrowband of noise and the masker was composed of multiple narrow bands of noise.

These examples of relatively long-lived persistence of auditory enhancement suggest that auditory enhancement might give rise to sequential effects in yes/no experiments in a variety of psychophysical tasks. The hypothesis tested here is that sensitivity in a masked detection task changes depending on whether the prior trial was a no-signal or a signal trial. It should be noted that this report was based on a retrospective analysis of an existing data set, a data set for which the masking is presumed to be relatively peripheral (energetic) rather than relatively central. Due to the retrospective aspect of this report, the choice of the stimuli does not reflect an effort to maximize enhancement effects. Instead, the experiment and results reflect stimuli and stimulus presentations that are typically used in psychoacoustics experiments (e.g., the time between trials was self-paced, etc.).

In the experiment analyzed, 12 distinct maskers were tested in separate blocks. Each masker was composed of six tones with different frequencies for the different maskers. The signal to be detected was a 1000-Hz tone. A yes/no procedure was used. In order to test for sequential effects on sensitivity, detection trials were separated into groups based on whether the prior trial was a no-signal trial or a signal trial. In parallel with typical auditory enhancement experiments, for the former group of trials the prior trial would encourage enhancement because there was no energy at the signal frequency and for the latter group of trials the prior trial would not encourage enhancement because there was energy at the signal frequency. Thus, one would expect values of d' to depend on whether the current trial was preceded by a signal or a no-signal trial.

2. Methods

2.1 Subjects

Eight normal-hearing subjects (S1–S8) ranging in age from 20 to 32 years participated in the experiments. All subjects had thresholds in quiet of 15 dB HL or lower for octave frequencies from 250 to 8000 Hz. Four of the subjects (S1–S4) had previously participated in psychoacoustics experiments. Subjects were compensated for their participation, except S5, the second author. Subjects were tested individually in double-walled sound-attenuated booths.

2.2 Stimuli

The subjects' task was to detect a tonal signal added to a six-tone equal-amplitude "frozen" masker. Twelve maskers were tested and Table 1 shows the frequencies of the components that comprised each masker. The level of each component was 50 dB SPL and the phases of the masker components were originally chosen at random, but fixed across trials. The signal to be detected was a 1000-Hz sinusoid whose phase was chosen at random on each presentation. The signal, when present, was synchronous with the masker. The duration of the stimuli was 102 ms including 5-ms cosine-squared onset and offset ramps.

The digitally generated stimuli were presented using two channels of a 16-bit digital-to-analog converter at a sample rate of 20 kHz. The stimuli were low-pass filtered at 7 kHz using matched filters (Stewart VBF 10M), separately attenuated, summed and presented diotically through Sennheiser HD410 SL headphones.

2.3 Procedure

A single-interval yes/no procedure was used. Subjects indicated the presence/absence of the signal by pressing the appropriate response key. As a result, the time between trials was self-

Table 1. The frequencies for each masker component are indicated using dashes.

	Frequency (Hz)											
Masker	252	317	399	502	631	795	1259	1584	1993	2509	2157	3973
1	—	—	—							—	—	—
2				—	—	—	—	—	—			
3	—	—	—				—	—	—			
4				—	—	—				—	—	—
5	—		—		—			—		—		—
6		—		—		—	—		—		—	
7	—			—		—	—		—			—
8	—	—			—			—			—	—
9	—		—			—	—			—		—
10		—	—	—					—	—	—	
11		—			—	—	—	—			—	
12			—	—	—			—	—	—		

paced by the subjects. On average, the time that elapsed between trials was approximately 2.1 s, with a range of 1.5–2.8 s across subjects and maskers. Visual feedback as to the correctness of each response followed each trial.

Experienced subjects ran ten sets of 60 trials, with frequent breaks. Three signal levels were tested, those values having been chosen based on approximately 2 h of practice prior to data collection for each masker such that the values of d' were approximately 1, 1.6, and 2.2. Accordingly, different signal levels were tested for different subjects and different maskers. Within each 60-trial set, 30 trials were no-signal trials. Of the remaining 30 trials the signal levels were intermixed such that each signal level was presented ten times in random order. Data collection was blocked by masker, and the order in which the different maskers were tested was different for the different subjects.

2.4 Analysis

In order to examine the effect of the prior stimulus on signal detection performance, trials were separated into two groups: trials preceded by a signal trial and trials preceded by a no-signal trial. This grouping was done separately for each subject and each masker tested. Values of d' and the criterion, c [defined as z_{hit}-z_{FA} and $-0.5(z_{\text{hit}}+z_{\text{FA}})$, respectively; Macmillan and Creelman, 2005], were then estimated for both groups. However, various constraints were placed on the trial selection, as described below.

The first constraint reflects the fact that in past informational masking experiments (e.g., Richards *et al.*, 2004) when trials were preceded by a "signal-plus-masker" cue, the signal level was the same for both the cue and the signal. In the current experiment, when a trial was preceded by a signal trial, the signal might have one of three levels. In an effort to reduce the potential confound of differences in signal levels presented in the current and previous trial, only trials for which the signal levels were the same for both were considered. Additional work will be required to determine whether this constraint is necessary, or whether the current sorting procedure could be applied to data from experiments in which signal levels are adjusted adaptively. Second, for the lowest signal level the resulting values of d' were sometimes near zero and so the results from the lowest signal level were not considered. Third, for the highest signal levels, the resulting values of d' sometimes approached infinity (a hit rate of 1), and so those trials were also removed from the analysis. Thus, in the final evaluation, only midlevel signal trials were evaluated.

Table 2 shows the summary information for the trials ultimately studied. The intensities of the midlevel signals and the corresponding global values of d' are shown for each masker

Table 2. The signal level (dB SPL) and d' are listed for each masker and each subject.

	Subject															
	S1		S2		S3		S4		S5		S6		S7		S8	
Masker	*L*	*d'*	*L*	*d'*	*L*	*d'*	*L*	*d'*	*L*	*d'*	*L*	*d'*	*L*	*d'*	*L*	*d'*
1	18	*1.7*	22	*1.7*	16	*1.1*	27	*0.9*	26	*1.8*	33	*1.4*	36	*2.2*	35	*1.7*
2	40	*1.5*	39	*2.0*	35	*1.6*	40	*1.4*	37	*2.6*	49	*1.9*	43	*2.1*	48	*1.7*
3	31	*1.1*	30	*1.1*	33	*2.1*	41	*1.1*	36	*2.6*	46	*1.7*	44	*1.5*	36	*0.9*
4	32	*1.7*	35	*1.1*	26	*1.2*	38	*1.7*	31	*2.2*	41	*1.6*	40	*2.5*	41	*2.0*
5	27	*1.7*	41	*2.8*	24	*1.2*	39	*0.9*	27	*1.8*	44	*1.4*	48	*2.3*	39	*1.6*
6	37	*1.4*	38	*1.6*	35	*1.0*	42	*1.4*	34	*0.9*	49	*1.4*	51	*2.0*	45	*1.4*
7	41	*1.6*	47	*2.1*	37	*1.1*	42	*1.7*	39	*1.5*	53	*2.0*	46	*1.1*	48	*1.5*
8	32	*1.2*	35	*1.5*	29	*0.7*	45	*0.8*	37	*1.7*	44	*0.9*	44	*2.0*	36	*1.6*
9	39	*2.2*	36	*1.6*	38	*1.9*	48	*1.3*	37	*1.4*	47	*1.3*	48	*2.2*	48	*1.3*
10	26	*0.9*	33	*1.0*	29	*1.5*	39	*1.6*	31	*1.1*	41	*1.5*	45	*2.0*	43	*1.9*
11	34	*2.1*	32	*0.8*	25	*1.0*	36	*0.1*	35	*1.5*	44	*2.3*	46	*2.7*	45	*1.4*
12	25	*1.4*	32	*1.6*	24	*1.6*	33	*0.9*	28	*0.3*	40	*1.2*	39	*1.7*	46	*2.1*

and each subject. The values of d', while somewhat variable, are near the target value of 1.6. The variation in signal level across maskers makes clear that not all maskers were equally effective. For example, due to the absence of masker energy in the region of the 1000-Hz signal (see Table 1), masker No. 1 provided less masking than the others.

For the primary analysis, the values of d' and c were separately estimated depending on whether the prior trial was a signal or no-signal trial, and this analysis was applied separately for each subject and for each of the 12 maskers tested. As a result, relatively few trials contributed to some of the estimates. Given the total of 600 trials for each masker, half of which were signal trials, and the fact that three signal levels were tested, a midlevel signal trial was preceded, on average, by a midlevel signal trial only 17 times and by a no-signal trial 50 times. For no-signal trials the expected numbers are somewhat improved: a no-signal trial was preceded by a midlevel signal trial 50 times and preceded by a no-signal trial 150 times. The expectation was that even though the number of trials contributing to each value of d' and c may be modest, the large number of maskers (12) and subjects (8) would provide sufficient statistical power to test the hypothesis that there is a change in sensitivity conditional upon whether the prior trial was a signal or a no-signal trial.

3. Results and discussion

Table 3 tabulates the estimates of d' for trials preceded by no-signal (d'_N) and signal (d'_S) trials for each masker and each listener. Figure 1 plots these results as a scatterplot. Results for different subjects are plotted using different symbols and for each subject the results for 12 maskers are plotted separately using only one symbol. The diagonal indicates points of equal d', i.e., no effect of the type of the prior trial. The data tend to fall below the diagonal, suggesting superior sensitivity when a trial was preceded by a no-signal than a signal trial. On average the values of d' were 0.3 larger when the trial was preceded by a no-signal trial than when it was preceded by a signal trial (see Table 3).

These results were analyzed using two repeated-measures ANOVAs. For the first ANOVA the random variable was subject identity and the fixed variables coded the preceding trial type and masker identity. For the second ANOVA, the random variable was masker identity and the fixed variables coded the preceding trial type and subject identity. Note that main effects of either subject or masker identity are not meaningful because the signal levels were chosen for different subjects and different maskers—a main effect would simply point to a failure to achieve equal values of d' across subjects and maskers. The interaction terms are, however, of

Table 3. Values of d' for trials preceded by no-signal trials (d'_N) and signal trials (d'_S) are listed for each masker and each subject.

	Subject																	
	S1		S2		S3		S4		S5		S6		S7		S8		AVG	
Masker	d'_N	d'_S	d'_N	d'_S	d'_N	d'_S	d'_N	d'_S	d'_N	d'_S	d'_N	d'_S	d'_N	d'_S	d'_N	d'_S	d'_N	d'_S
1	1.6	2.3	2.4	1.1	1.5	1.0	1.1	0.6	1.8	2.4	1.6	0.9	2.2	2.2	1.8	1.8	1.7	1.5
2	1.6	1.2	2.7	1.5	2.0	1.5	1.7	1.2	2.4	2.7	2.3	2.0	2.3	2.5	2.2	2.1	2.2	1.8
3	1.0	0.9	0.9	1.3	1.9	2.1	1.5	0.7	2.6	2.8	2.1	1.6	1.4	2.0	1.2	0.6	1.6	1.5
4	1.7	1.6	1.6	0.7	1.7	1.0	2.2	2.2	2.2	2.4	1.8	1.0	2.9	2.3	2.9	3.4	2.1	1.8
5	1.9	1.4	3.5	2.7	1.4	1.3	1.0	1.4	2.1	1.5	1.4	1.4	2.5	2.6	2.1	1.8	2.0	1.7
6	1.6	1.4	2.1	1.5	1.0	1.5	1.8	0.9	0.7	1.2	1.7	1.2	1.9	2.4	2.0	1.8	1.6	1.5
7	1.9	1.6	2.5	1.8	1.4	1.0	1.8	2.1	1.4	1.7	2.1	2.5	1.0	1.0	2.0	1.2	1.8	1.6
8	1.4	1.0	1.6	1.4	0.8	0.8	1.0	0.6	2.1	1.9	0.5	1.7	2.4	1.9	1.8	1.0	1.5	1.3
9	2.3	1.6	1.6	1.5	1.9	2.4	1.7	1.1	1.1	1.9	1.0	1.7	2.5	1.3	1.8	0.9	1.8	1.5
10	1.2	1.0	1.3	1.2	1.6	2.1	1.9	1.7	1.2	0.8	1.6	1.2	2.2	2.8	1.9	2.4	1.6	1.6
11	2.6	1.8	1.3	0.5	1.5	0.3	0.2	0.0	1.8	1.6	2.7	1.7	2.9	2.4	1.7	1.1	1.8	1.2
12	1.7	1.3	1.9	2.0	2.0	1.6	0.8	1.2	0.6	0.0	1.5	0.6	2.1	0.6	2.7	1.4	1.7	1.1
																Grand Average	1.8	1.5

interest in order to evaluate whether the change in sensitivity associated with the type of prior trial (if a change exists) depends on individual differences or masker spectra, respectively.

The former ANOVA indicated a significant effect of the preceding trial type ($F(1,7) = 18.7, p < 0.005$), but the effect of masker identity and the interaction term were not significant ($p > 0.25$). The latter ANOVA indicated a significant effect of the preceding trial type ($F(1,11) = 19.1,\ p < 0.005$). Additionally there was a significant effect of subject identity ($F(7,77) = 2.6,\ p < 0.05$) but the interaction term was not significant ($p > 0.3$). Thus, the 0.3 shift in d' associated with the properties of the prior trial is statistically significant, and is consistent with expectations associated with auditory enhancement. As a counter point to this finding, Purks *et al.* (1980) failed to reveal a significant effect of prior trials on sensitivity in an identification task.

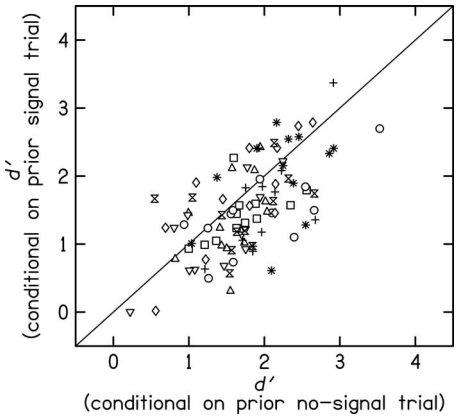


Fig. 1. Values of d' for trials preceded by signal trials (ordinate) and trials preceded by no-signal trials (abscissa) are plotted in a scatter plot. The different symbols are for different subjects. For each subject, the results for 12 maskers are plotted separately.

Next, consider the criteria. Past research has demonstrated both positive and negative sequential effects in detection experiments (e.g., Carterette *et al.*, 1966; Sandusky and Ahumada, 1971). Such phenomena would typically be revealed as a reduction, or increase, in the criterion value relative to neutral, respectively. In the current experiment, however, the value of d' depended on the prior trial. Thus, shifts in "signal" and "no signal" response rates cannot be simply described as changes in criteria because it is unknown how the assumed underlying distributions changed (e.g., do both the signal and no-signal distribution shift depending on the prior trial type?). Nonetheless, we were interested in evaluating whether subjects' criteria placement changed depending on the preceding trial type. A repeated-measures ANOVA was run for the estimated criterion c using subject identity as the random variable and the preceding trial type and masker identity as the fixed variables. Neither the main effects, nor the interaction term, were significant ($p > 0.2$). Thus, listeners' predilection to vote signal did not depend on the prior trial type, although listeners' sensitivity did depend on the prior signal type.

To summarize, the data indicate that for the detection of a tone added to a sparse multitone masker, sensitivity in yes/no trials depends on the prior trial. The effect size was small, a change of 0.3 in d'. Keeping in mind that the maskers tested here were not chosen to maximize the magnitude of the enhancement, the current finding is of particular interest because many types of stimuli tested in psychophysical experiments might, when examined, be found to contain sequential effects.

4. Summary and conclusions

Past work (e.g., Viemeister, 1980) has indicated an enhancement effect in which the detection of a tone added to a masker is improved when a masker is continuous compared to when the masker is pulsed with the signal. Viemeister's results (see also Richards and Neff, 2004) indicate enhancement effects can be quite long-lived, on the order of seconds. This suggests that sensitivity in a yes/no masked detection task might depend on the characteristics of the stimulus in the prior trial. An evaluation of an existing data set thought to reflect energetic, or peripheral, masking confirmed sequential effects of prior trials on sensitivity in a simple yes/no masked detection task. Regardless of whether the masking effects are thought to reflect relatively central or relatively peripheral limitations/processing, there may be value in testing for sequential effects in masked detection experiments when yes/no trials are used.

Acknowledgments

This work was supported by Grant No. RO1 DC02012 from the National Institutes of Health. We thank Dr. Daniel E. Shub, Dr. Frederick L. Wightman, and two anonymous reviewers for their helpful comments on an early version of this manuscript.

References and links

Carterette, E. C., Friedman, M. P., and Wyman, M. J. (**1966**). "Feedback and psychophysical variables in signal detection," J. Acoust. Soc. Am. **39**, 1051–1055.
Green, D. M. (**1969**). "Masking with continuous and pulsed sinusoids," J. Acoust. Soc. Am. **46**, 939–946.
Lutfi, R. A., and Alexander, J. M. (**2005**). "Effects of informational maskers within and across trials," J. Acoust. Soc. Am. **118**, 322–324.
Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory; A Users Guide*, 2nd ed. (Lawrence Erlbaum Associates, Mahwah, NJ).
McFadden, D., and Wright, B. A. (**1990**). "Temporal decline of masking and comodulation detection differences," J. Acoust. Soc. Am. **88**, 711–724.
McFadden, D., and Wright, B. A. (**1992**). "Temporal decline of masking and comodulation masking release," J. Acoust. Soc. Am. **92**, 144–156.
Purks, S. R., Callahan, D. J., Braida, L. D., and Durlach, N. I. (**1980**). "Intensity perception. X. Effect of preceding stimulus on identification performance," J. Acoust. Soc. Am. **67**, 634–637.
Richards, V. M., and Neff, D. L. (**2004**). "Cuing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.
Richards, V. M., Huang, R., and Kidd, G., Jr. (**2004**). "Masker-first advantage for cues in informational masking," J. Acoust. Soc. Am. **116**, 2278–2288.

Sandusky, A., and Ahumada, A. (**1971**). "Contrast in detection with gated noise," J. Acoust. Soc. Am. **48**, 1790–1794.
Viemeister, N. F. (**1980**). "Adaptation of masking," in *Psychophysical, Physiological, and Behavioral Studies in Hearing*, edited by G. van den Brink and F. S. Bilsen (Delft University Press, Delft, The Netherlands), pp. 190–198.
Wright, B. A., McFadden, D., and Champlin, C. A. (**1993**). "Adaptation of suppression as an explanation of enhancement effects," J. Acoust. Soc. Am. **94**, 72–82.

Speaking up: Killer whales (*Orcinus orca*) increase their call amplitude in response to vessel noise

Marla M. Holt and Dawn P. Noren
Marine Mammal Program, National Oceanic and Atmospheric Administration (NOAA), National Marine Fisheries Service (NMFS), Northwest Fisheries Science Center, 2725 Montlake Boulevard East, Seattle, Washington 98112
marla.holt@noaa.gov, dawn.noren@noaa.gov

Val Veirs
Department of Physics, Colorado College, Colorado Springs, Colorado 80903
vveirs@coloradocollege.edu

Candice K. Emmons
Marine Mammal Program, National Oceanic and Atmospheric Administration (NOAA), National Marine Fisheries Service (NMFS), Northwest Fisheries Science Center, 2725 Montlake Boulevard East, Seattle, Washington 98112
candice.emmons@noaa.gov

Scott Veirs
Beam Reach Marine Science and Sustainability School, 7044 17th Avenue NE, Seattle, Washington 98112
scott@beamreach.org

Abstract: This study investigated the effects of anthropogenic sound exposure on the vocal behavior of free-ranging killer whales. Endangered Southern Resident killer whales inhabit areas including the urban coastal waters of Puget Sound near Seattle, WA, where anthropogenic sounds are ubiquitous, particularly those from motorized vessels. A calibrated recording system was used to measure killer whale call source levels and background noise levels (1–40 kHz). Results show that whales increased their call amplitude by 1 dB for every 1 dB increase in background noise levels. Furthermore, nearby vessel counts were positively correlated with these observed background noise levels.

PACS numbers: 43.80.Ka, 43.80.Nd [CM]
Date Received: August 12, 2008 **Date Accepted:** October 31, 2008

1. Introduction

Marine mammals use sound for activities essential to survival and reproduction (NRC, 2003). They are often faced with the challenge of hearing these sounds in environments with noise from both natural and anthropogenic sources. Anthropogenic sound exposure in marine mammals has caused much concern, especially in cases that have extreme outcomes such as beaked whale mass strandings coinciding with naval midfrequency sonar exercises (NRC, 2003). While such exposure might lead to death, the occurrence of these sounds is relatively rare. Other sources of anthropogenic sounds in the ocean, such as motorized vessel traffic in coastal areas, are far more ubiquitous. Long-term exposure to prevalent anthropogenic noise may have deleterious effects on marine mammal populations, particularly those that frequent coastal regions where vessel traffic is concentrated. Although recent work in the laboratory has helped to define what sound pressure levels or sound exposure levels would likely result in auditory effects such as temporary threshold shift or masking in individuals, determining such effects in free-ranging marine mammals is challenging. Furthermore, information is lacking on what

sound levels from specific anthropogenic sources result in behavioral responses in marine mammals and how these might have negative impacts on individuals and populations. This information is particularly crucial for informing policy designed to recover endangered populations.

Fishing-eating "resident" killer whales live in large and stable matrilineal groups, forage cooperatively (Ford and Ellis, 2006), and produce a variety of calls that have been described as discrete, variable, and aberrant (Ford, 1989). Discrete calls of killer whales are stereotyped and distinctive in structure as well as population-specific (Ford, 1987, 1989). Calls are thought to play key roles in social bonds among kin, mates, and other conspecifics. For example, call production and exchange is believed to function to maintain cohesion and coordination among group members when individuals are dispersed and foraging (Ford, 1989). Southern Resident killer whales (SRKWs), currently consisting of approximately 85 individuals among three (J, K, and L) pods, are listed as endangered (NMFS, 2005) and are found in coastal waters of Washington state and British Columbia. These inlands waters are important foraging areas for this population, particularly in the summer and fall where vessels, including ships, ferries, whale-watching boats, and private boats, are also prevalent. For example, the average number of vessels surrounding this population of killer whales in the summer is approximately 20 (Koski *et al.*, 2006) and it is not uncommon to find at least 50 vessels surrounding these whales during summer weekends and holidays. The frequency range of noise emitted from close vessels overlaps with the frequency range of killer whale calls (Ford, 1987; Erbe, 2002). Thus, this study aimed to address the impacts of motorized vessel noise on SRKW vocal behavior which is likely integral to their survival and reproductive success.

Individuals may compensate for background noise by changing their signal's amplitude, duration, repetition rate, and/or frequency. For example, SRKWs produced longer calls in the presence of vessel noise following a large increase in whale-watching boats (Foote *et al.*, 2004). Such vocal compensation is often interpreted as an antimasking strategy for high background noise levels. Our goal was to investigate call amplitude compensation (the Lombard effect), including measuring background noise levels and the number of nearby vessels of all types associated with these noise levels that were undocumented in previous studies. This approach is critical to elucidate how vessel noise affects the behavior of endangered killer whales and how they might compensate for changing levels of background noise to overcome masking by specific anthropogenic sources in their environment.

2. Method

We collected data in waters surrounding the San Juan Islands, WA off of a 8-m Pacific aluminum skiff, RV Noctiluca, over several days from August 23–September 4, 2007 as weather conditions and whale presence allowed it. Given the focus of this study on vessel noise effects, all measurements were made in sea states ranging from $\frac{1}{2}$ to 2.

When whales were sighted, the research vessel was positioned ahead and in the path of the whales (approximately 500–1000 m) to obtain on-axis recordings, the motor was shut down, and the recording equipment was deployed. We collected call and background noise data continuously while recording latitude and longitude, total motorized vessel numbers within 1000 m (measured by a laser range finder, Yardage Pro, Bushnell), and pod and identification (ID) of individual whales every 10 min. Individual whale IDs, distance from the research vessel (estimated with the range finder) and direction (visually estimated in 30 deg increments) of surfacing individuals relative to the research vessel were also taken opportunistically while recordings were made. Water temperature and salinity were measured at 5 m increments down to 30 m using a conductivity and temperature probe (YSI 30-M) at each location that acoustic data were collected.

Call source levels and background noise levels were measured from a calibrated omnidirectional hydrophone (Reson TC-4033) connected to a low-noise bandpass preamplifier (Reson VP2000, 1–100 kHz). A four-element hydrophone array (LabCore Systems) was used for localization that had a 20-m aperture and consisted of 5-m spacing between hydrophones 1–2 and 2–3 and 10-m spacing between hydrophones 3–4. The array was deployed vertically with hydrophone 1 at 5 m depth from the research vessel with a buoy and 10 kg weight. Signals

were digitized using a MOTU Traveler (eight-channels, sampling rate 192 kHz), recorded using a customized version of Ishmael 1.0 (Mellinger, 2002), and stored as five-channel wave files on a PC laptop for analysis.

The range of a call was determined in Ishmael using time-of-arrival differences between hydrophone pairs in the array relative to the Reson hydrophone (hyperbolic localization). A sound speed of 1485 m/s was assumed based on average temperature and salinity profiles (MacKenzie, 1981). We determined the accuracy of Ishmael's range estimates and transmission loss in situ at two locations representative of where data were collected when no killer whales were sighted in the area over several days. Previously recorded S1 calls (see Ford, 1987) were projected using a sound source (Lubell LL 9816, 9 m depth) deployed from a dinghy at known horizontal distances. Received levels indicated that spherical spreading loss was an appropriate assumption. The largest resulting errors in source level occurred when the estimated range was relatively close (<40 m) or far (>400 m). This was expected because hyperbolic curves intersect at large enough angles to fix a location in the region that is neither too far from the linear array nor near its axis. At close ranges, range errors also results in larger source level errors given the logarithmic nature of transmission loss as defined in Eq. (2). Thus, we only included calls that were localized within an estimated range of 40–400 m. The subsequent range errors resulted in an average calculated source level error of 2.6 dB.

Call source level and background noise level measurements in dB_{rms} re 1 μPa were made over a 250 ms duration and a bandwidth of 1–40 kHz using SpectraPLUS v5.0 (Pioneer Hill) that was calibrated with 1 kHz pure tone projected in the water at a known received sound pressure level. The bandwidth was chosen based on both the mean hearing curve of captive killer whales (Szymanski *et al.*, 1999) and the observed frequency range of SRKW discrete calls when recorded on-axis. Background noise levels were taken just prior to a call (within 9 s) unless other overlapping whale sounds were present. In those cases, background levels measured just after the call (within 9 s) were used instead. Only source levels for stereotyped, discrete SRKW calls (Ford, 1987) are reported in the current study. The highest amplitude in dB_{rms} over a 250 ms duration within each call was chosen for calculating source levels. Call received levels (RL) were calculated by subtracting background noise levels from the signal logarithmically as follows:

$$RL = 10 \log[10^{(dBsignal/10)} - 10^{(dBnoise/10)}]. \quad (1)$$

This was necessary given that the received signal level of the call was not always well above the corresponding background level. Call source levels were then calculated (in dB rms re 1 μPa at 1 m) as

$$SL = RL + 20 \log R, \quad (2)$$

where R was the range of the call estimated by Ishmael. Similar to Foote *et al.* (2004), call duration (to the nearest 0.01 s) was determined by using the cursor function of the sound analysis software program (SpectraPLUS v5.0). Linear regression analysis was used to determine relationships between call source levels and background noise levels, between call duration and background noise levels, and between background noise levels and $\log_{10}$ vessel counts. A t-test was also used to compare differences in call duration between low (<110 dB re 1 μPa) and high (>110 dB dB re 1 μPa) noise conditions.

3. Results and discussion

Call source levels and background noise levels were determined from recordings collected on four days (8/28/07, 8/29/07, 9/1/07, and 9/2/07). Only members of *J* pod were present on three of these four days, while all three pods (J, K, and L) were present on the other day. Call source levels in the 1–40 kHz band ranged from 133 to 174 dB re 1 μPa at 1 m with a mean of 155.3 dB re 1 μPa at 1 m (±7.4 SD). This mean was within 3 dB of stereotyped call source levels reported by Miller (2006) in Northern Resident killer whales despite differences in bandwidth, duration, and transmission loss assumptions used in making these measurements. Back-

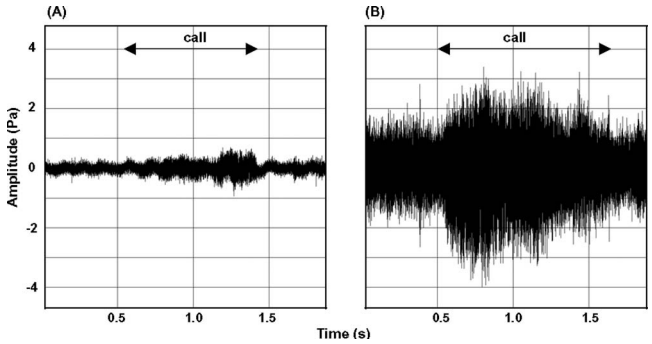


Fig. 1. Examples of killer whale calls recorded from an estimated range of 70 m in different noise levels based on dB_{rms} measurements between 1 and 40 kHz (a) S1 call in noise level of 100 dB re 1 μPa and (b) S1 call in noise level of 115 dB re 1 μPa. Arrows indicate the call in the time series.

ground noise levels in the 1–40 kHz band ranged from 98 to 123 dB re 1 μPa with a mean of 110.1 dB re 1 μPa (±4.1 SD). Nearby vessel counts ranged from 1 to 46.

We found a significant positive correlation between call source level and background noise level across all call types ($p<0.001$, $R^2_{adj}=0.25$, $n=274$). Since source levels and duration vary by call type (Ford, 1987; Miller, 2006) the subsequent analyses were restricted to one call type (S1) with the largest sample size. Examples of the S1 call recorded at similar ranges in relatively quiet and noisy conditions are shown in Fig. 1. Background noise levels explained approximately 50% of the variation in S1 call source levels ($p<0.001$, $R^2_{adj}=0.48$, $n=104$; Fig. 2(a)). Sample sizes were too small for similar regression analyses of other call types. The slope of the fitted regression line indicated that S1 call source level increased by approximately 1 dB for every 1 dB increase in background noise level (Fig. 2(a)). Furthermore, vessel traffic was clearly correlated with background noise levels ($p<0.001$, $R^2_{adj}=0.45$, $n=274$; Fig. 2(b)). Although it appears that S1 calls were produced at more or less equal source levels for corresponding background levels below 105 dB re: 1 μPa (Fig. 2(a)), the sample size below this background noise level was insufficient to precisely determine a threshold effect, and this observation warrants further investigation. Durations of these S1 calls ranged from 0.49 to 1.58 s with a mean of 0.95 s (±0.24 SD). Although the example in Fig. 1 shows a longer S1 call in the noisier condition, we found no significant slope or difference in call duration with changes in background noise level.

To our knowledge, these are the first data describing the Lombard effect in killer whales. Whales increased their call source level by 1 dB as background noise levels increased by 1 dB, at least over the range of background noise level measurements observed in this study. Schiefele *et al.* (2005) also reported a similar rate of vocalization level increase in response to a passing vessel in St. Lawrence River beluga. The upper range of background noise levels reported in the current study corresponded to approximately 45 nearby vessels. Such vocal compensation behavior by calling whales is presumably an effort to maintain adequate signal to noise ratios relative to listening whales.

Killer whales did not adjust their call duration over the range of background noise levels measured in this study. In contrast, Foote *et al.* (2004) reported a significant increase in killer whale call duration in the presence of vessel noise compared to in the absence of vessel noise. Differences in the results might be related to methodological differences between studies such as how vessel noise was assessed and/or the time span of data collection. Given that the current study did not include data from "no vessel noise" conditions, it is also possible that killer whales adjust call duration as a step response while they adjust call amplitude as a graded response to high background noise levels.

The results presented here show that as background noise from vessel traffic increases, killer whales adjust their vocal behavior by increasing call amplitude. The potential costs of

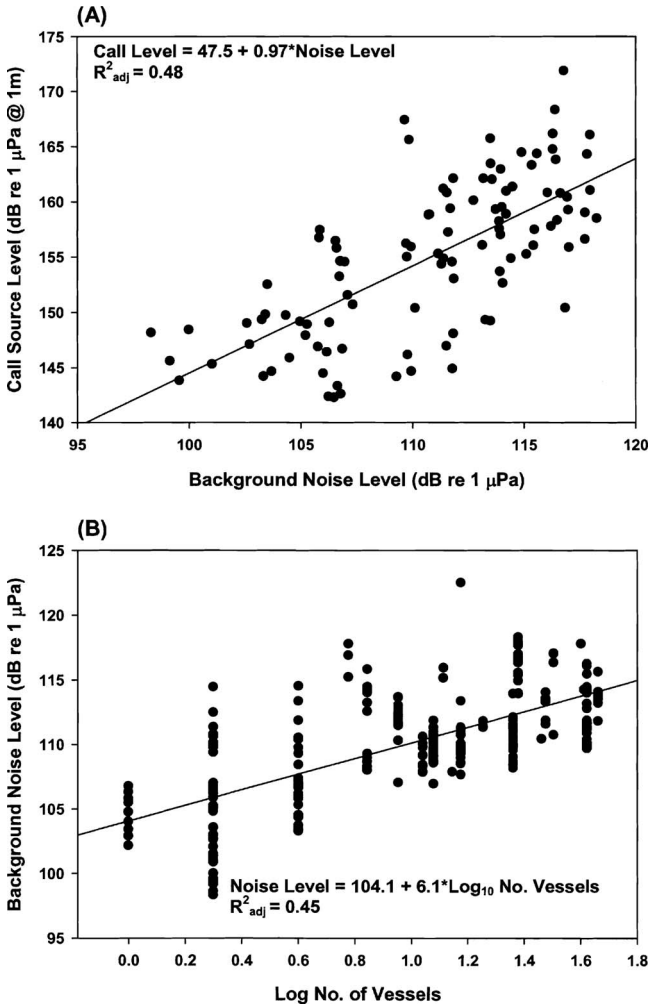


Fig. 2. (a) Killer whale call source level (S1 call type) as a function of background noise level, both based on dB_{rms} measurements between 1 and 40 kHz. (b) Background noise level as a function of the $\log_{10}$ number of vessels within 1000 m from the hydrophone.

such vocal compensation are important to consider. For example, increasing vocal output to compensate for noise might have energetic costs (Oberweger and Goller, 1991), lead to increased stress levels, or degrade communication among individuals which could affect their activity budget. At some level, background noise could also completely impede the use of calls by killer whales for communicative functions.

Acknowledgments

We thank Mike Ford, Brad Hanson, and two anonymous reviewers for helpful comments on the manuscript. We also thank Teresa Mongillo, Amy Johnson, and students of beamreach.org for field assistance. Work was funded by the National Research Council Postdoctoral Research Associateship Program and NOAA Northwest Fisheries Science Center. Data were collected under Research Permit No. 781-1824-00 from the U.S. NMFS Office of Protected Resources and Species at Risk Act/Marine Mammal License No. MML 2007-08/SARA-68A from the Department of Fisheries and Ocean, Canada.

References and links

Erbe, C. (**2002**). "Underwater noise of whale-watching boats and potential effects on killer whales (Orcinus orca), based on acoustic impact model," Marine Mammal Sci. **18**, 394–418.

Foote, A. D., Osborne, R. W., and Hoelzel, A. R. (**2004**). "Whale-call response to masking boat noise," Nature (London) **428**, 910.

Ford, J. K. B. (**1987**). "A catalogue of underwater calls produced by killer whales (Orcinus orca) in British Columbia," Can. Data Rep. Fish. Aquat. Sci., p. 1-165.

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., and Ellis, G. M. (**2006**). "Selective foraging by fish-eating killer whales," Mar. Ecol.: Prog. Ser. **316**, 185–199.

Koski, K., Osborne, R., and Tallmon, R. (**2006**). "Soundwatch public outreach/boater education project 2004–2005 final program report," NMFS Contract No. AB133F-04-SE-0835 (Report No. NFFP 5000-4-00026), p. 1-25.

Mackenzie, K. V. (**1981**). "Discussion of seawater sound-speed determination," J. Acoust. Soc. Am. **70**, 807–812.

Mellinger, D. K. (**2001**). "ISHMAEL: Integrated System for Holistic Multi-channel Acoustic Exploration and Localization," U.S. Dept. Commer., NOAA Tech. Memo. OAR-PMEL-120, 26 p.

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.

NMFS (National Marine Fisheries Services) (**2005**). "Endangered and threatened wildlife and plants: Endangered status for Southern Resident killer whales," Fed. Regist. [Docket No. 041213348-5285-02, 18 November 2005] **70**(217), 69903–69912.

NRC (National Research Council) (**2003**). *Ocean Noise and Marine Mammals* (National Academies Press, Washington, DC).

Oberweger, K., and Goller, F. (**2001**). "The metabolic cost of birdsong production," J. Exp. Biol. **204**, 3379–3388.

Scheifele, P. M., Andrew, S., Cooper, R. A., Darre, S., 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.

Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., and Henry, K. R., (**1999**). Killer whale (Orcinus orca) hearing: Auditory brainstem response and behavioral audiograms. J. Acoust. Soc. Am. **106**, 1134–1141.

 [DOI: 10.1121/1.3040031] Published Online 22 December 2008

Threshold fine structure affects amplitude modulation perception

Stephan J. Heise, Manfred Mauermann, and Jesko L. Verhey
Institut für Physik, Universität Oldenburg, D-26111 Oldenburg, Germany
stephan.heise@uni-oldenburg.de, manfred.mauermann@uni-oldenburg.de, jesko.verhey@uni-oldenburg.de

Abstract: Modulation detection thresholds of a sinusoidally amplitude-modulated tone were measured for two different positions of the low-level carrier relative to the fine structure of the threshold in quiet. Modulation detection thresholds were higher for a carrier at a fine-structure minimum than for a carrier at a fine-structure maximum, regardless of whether the carriers had the same sound pressure level or the same sensation level. This indicates that even for small variations of the carrier frequency, the sensitivity to amplitude modulation can vary substantially due to the frequency characteristics of the threshold in quiet.

PACS numbers: 43.66.Cb, 43.66.Mk [BLM]
Date Received: October 6, 2008 **Date Accepted:** November 6, 2008

1. Introduction

The common finding of quasiperiodic fluctuations in the threshold in quiet is known as threshold fine structure (or microstructure). In fine-structure regions the threshold may change by as much as 15 dB within only 1/10 octave, which also affects equal-loudness contours at low levels (Mauermann *et al.*, 2004). Fine-structure minima, i.e., very sensitive regions of hearing, are often associated with spontaneous otoacoustic emissions (SOAEs), and both phenomena are thought to share similar cochlear origins (for an overview, see Heise *et al.*, 2008).

This letter investigates the influence of threshold fine structure on the perception of sinusoidal amplitude modulation (AM) of pure tones at low levels. The spectrum of such an amplitude modulated stimulus consists of three components, the carrier frequency f_{car} and two sidebands at $f_{car} \pm f_{mod}$ which are separated by the modulation frequency f_{mod} from the carrier. It has been shown that for sinusoidal carriers the detectability of an amplitude modulation is independent of the carrier frequency, at least for modulation frequencies that are smaller than about 100 Hz (Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001). The first question of this letter is whether this still holds in regions with threshold fine structure. That is, are modulation detection thresholds (MDTs) the same, regardless of whether the carrier frequency falls in a fine-structure minimum or maximum? Previous studies by Zwicker (1986) and by Long (1993) indicate that differences may be expected. The second question addressed in this study quite naturally arises when trying to answer the first: Which carrier level should be used when comparing MDTs at levels near the threshold in quiet? Should one compare MDTs of carriers that have the same absolute level (sound pressure level (SPL)) or the same level relative to the individual threshold in quiet (sensation level (SL))? As mentioned above, the difference between these two measures of level can amount to 10–15 dB, if the carrier frequencies are that of a fine-structure minimum and a fine-structure maximum. Besides, this second question is not only restricted to studies dealing with threshold fine structure. When setting up an experiment to compare MDTs at low levels across subjects, the question whether to use equal carrier SPLs or SLs needs to be decided upon. In the literature, however, it is only indirectly present and has been answered in different ways. For example, Kohlrausch *et al.* (2000) measured MDTs at equal carrier SPLs, but gave the subjects' thresholds in quiet at the carrier frequencies. Moore and Glasberg (2001) used a carrier level of 30 dB SPL whereas Millman and Bacon (2008) used 30 dB SL.

The current study compares MDTs for carriers in fine-structure minima and maxima with equal carrier SPLs and with equal carrier SLs within the same subjects. In contrast to previous studies (Zwicker, 1986; Long, 1993), we chose the modulation frequency in such a way that the sidebands always fell in fine-structure extreme values adjacent to that at the carrier frequency. That is, in a condition where the carrier frequency was that of a fine-structure maximum the sidebands fell into the adjacent fine-structure minima, and vice versa.

2. Methods

Nine subjects aged between 17 and 30 years took part in this study. In a first step the subjects' thresholds in quiet were screened for fine structure by means of a tracking procedure (see Heise *et al.*, 2007, 2008). In regions that showed threshold fine structure, a threshold minimum and an adjacent maximum were chosen, whose frequencies $f_{\min}$ and $f_{\max}$ served as carrier frequencies in the modulation detection experiment. The frequency distance between these extreme values defined the modulation frequency: $f_{\mathrm{mod}} := |f_{\min} - f_{\max}|$. An adaptive three-alternative forced-choice (3-AFC) procedure was used to remeasure thresholds in quiet at those frequencies that occured in the modulation detection experiment, i.e., $f_{\min}$, $f_{\min} \pm f_{\mathrm{mod}}$, $f_{\max}$, and $f_{\max} \pm f_{\mathrm{mod}}$. Note that due to the manner in which f_{mod} was defined this is a total of four different frequencies for each subject. The stimuli had a duration of 250 ms (including 2×25 ms $\cos^2$ ramps) and were separated by 500 ms of silence. The stimulus level started at 15 dB SPL and initially changed in steps of 6 dB. After two reversals the step size was reduced to 3 dB and after four reversals it was changed to 1 dB where it remained for another eight reversals, whose average gave the threshold estimate. Two to three threshold estimates were obtained for every frequency. The thresholds in quiet obtained by the two methods are given in Fig. 1 together with the individual values for $f_{\min}$ and $f_{\max}$. For two subjects (HK and SD) the experiment was done twice in different frequency regions. Note that in order to facilitate the comparison between the data from the two methods the tracking thresholds were shifted vertically so as to minimize their squared distance to the AFC data. For although the shape of threshold fine structure is measured very accurately by the tracking procedure, the absolute threshold values may be biased by the subject's internal threshold criterion (Heise *et al.*, 2008).

MDTs for sinusoidally amplitude-modulated sinusoids were measured in the following three conditions (cf. sketch in Fig. 2):

- Max—The carrier frequency was that of a fine-structure maximum, so that the sidebands fell close to adjacent fine-structure minima. The carrier level was 15 dB SL.
- MinSPL—The carrier frequency was that of a fine-structure minimum, so that the sidebands fell close to adjacent fine-structure maxima. The carrier had the same SPL as in the Max condition.
- MinSL—The carrier frequency was that of a fine-structure minimum, so that the sidebands fell close to adjacent fine-structure maxima. The carrier level was 15 dB SL.

Again, an adaptive 3-AFC procedure with a 1-up/2-down step rule was used. The stimuli were 500 ms long (including 2×20 ms $\cos^2$ ramps) and were separated by 300 ms of silence. As a tracking variable the modulation depth in dB ($20 \log_{10} m$) was used, starting at -4 dB. The step size was 4 dB initially, then changed to 2 dB after two reversals, and finally to 1 dB after another two reversals. A further eight reversals were recorded and averaged to give the threshold estimate. For each condition three threshold estimates were obtained.

The stimuli were generated digitally, amplified by a Tucker Davis HB7 amplifier and played back via Sennheiser HDA200 headphones. Subjects were seated in a double-walled sound-attenuating booth. The threshold measurements with the 3-AFC procedure were done immediately before the modulation detection experiments in order to minimise errors in the carrier sensation levels of the AM stimuli.

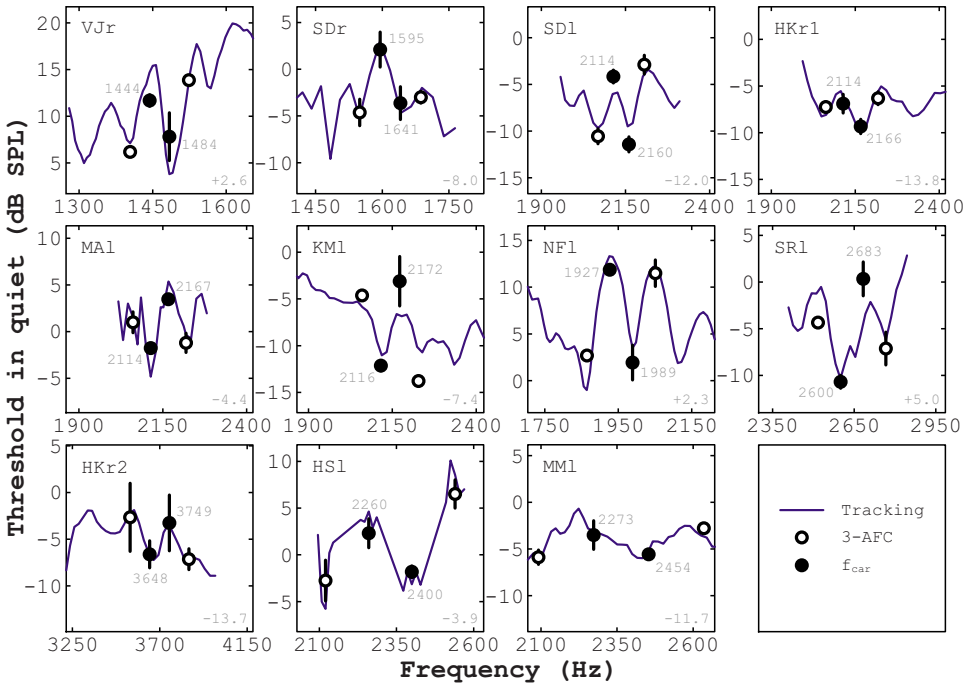


Fig. 1. (Color online) Individual thresholds in quiet obtained by two methods: a tracking procedure (lines) and a 3-AFC procedure (circles). The tracking thresholds were shifted vertically in order to facilitate the comparison to the 3-AFC thresholds. The amount of the shift in decibels is given in the bottom right corner of each panel. Error bars indicate standard deviations of two (subjects VJr, NFl) or three (other subjects) threshold estimates. Frequencies that were used as carrier frequencies in the modulation detection experiment are marked by filled circles and their values in hertz are given.

3. Results and discussion

The individual and the mean MDTs in the three experimental conditions Max, MinSPL, and MinSL are shown in Fig. 2. Although the absolute MDTs show large interindividual differences, for all subjects the MDTs in the Max condition are lower than in the MinSPL and the MinSL conditions. This difference in the MDTs is highly significant when tested with Fisher's permutation test ($p<0.01$). That is, modulation detection is better when the carrier falls in a fine-structure maximum than when it falls in a minimum, regardless of whether the carriers have an equal SPL or SL. In other words, regions that are more sensitive to tones are less sensitive to amplitude modulations. Further, for the conditions where the carrier was in a fine-structure minimum, the MDTs in the MinSPL condition are lower than in the MinSL condition, which reflects the well-known improvement of MDTs with level (Millman and Bacon, 2008; for a review see Kohlrausch, 1993). Again, this difference is highly significant ($p<0.01$).

The difference between the thresholds in quiet at the minimum and maximum that were used as carrier frequencies is given as a bar in Fig. 2 for each subject. This estimate of the local amount of threshold fine structure is correlated with the difference between the MDTs in conditions Max and MinSPL ($r=0.78$), as well as with the difference between the MDTs in conditions Max and MinSL ($r=0.81$).

Our results are in qualitative agreement with previous results from the literature (Zwicker, 1986; Long, 1993). However, the effect of threshold fine structure on MDTs is larger by 2–5 dB in the present study, even though the local amount of fine structure is on average about 3.5 dB less. This may be due to the choice of the modulation frequency, which was adapted to the individual fine structure in the current study. From a spectral point of view one

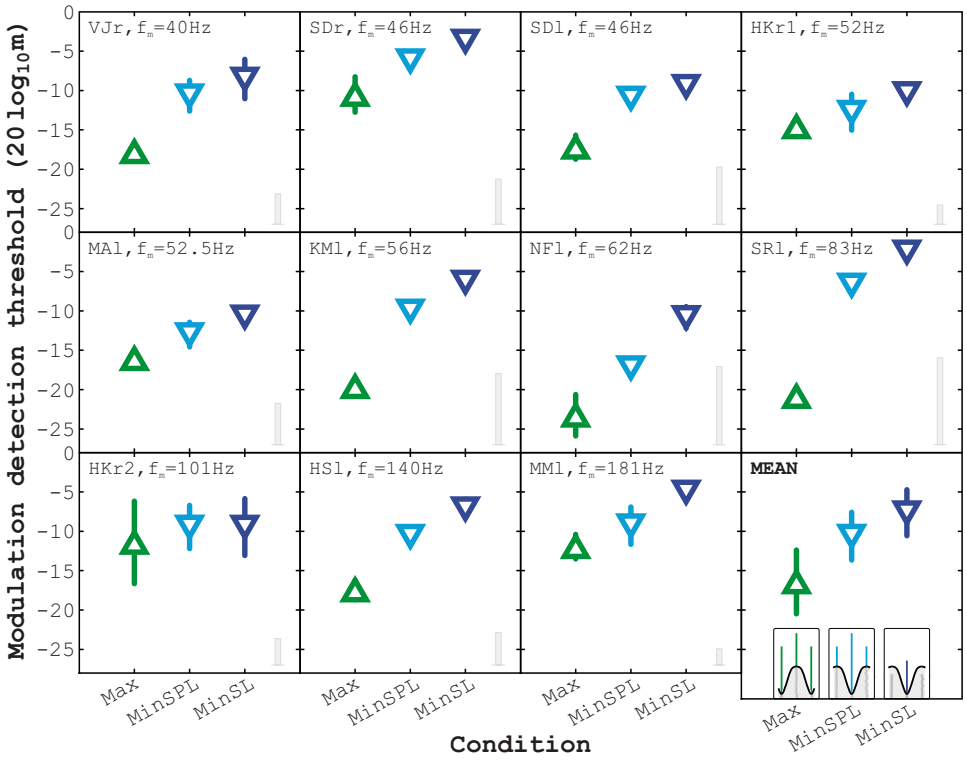


Fig. 2. (Color online) Individual and mean modulation detection thresholds in the three conditions. Error bars indicate standard deviations. The panels are sorted by modulation frequency. The bottom right corner of each panel provides the difference between the thresholds in quiet at the minimum and maximum that served as carrier frequencies. The bottom right panel also contains sketches of the three conditions in the frequency domain.

may expect the effect to be largest when the sidebands fall in fine-structure extreme values adjacent to that at the carrier frequency (this is discussed in more detail below). Previous research focused only on the position of the carrier relative to the fine structure of the threshold in quiet.

For high levels it is known that sinusoidal AM of tones is almost independent of modulation frequency for $f_{\mathrm{mod}} \lesssim 100$ Hz. It is assumed that in this regime of low modulation frequencies the detection of the AM is based on temporal fluctuations in the envelope of the signal (e.g., Yost and Sheft, 1997). For high modulation frequencies that are larger than the width of the auditory filter at the carrier frequency, MDTs improve significantly. This is ascribed to the sidebands becoming resolvable and thus useable as a detection cue. At low levels this cue may not be available since the level of the side bands may fall below the threshold in quiet. For intermediate modulation frequencies MDTs deteriorate, possibly because of some internal noise. In the current study the modulation frequencies were chosen individually. In eight subjects modulation frequencies were less than 100 Hz, i.e., these would be expected to have based their detection on temporal envelope fluctuations. In three subjects the modulation frequency was above 100 Hz. These subjects may already have been in the regime of intermediate modulation frequencies and therefore their data may not be comparable to that of the first group. However, from their MDTs these two groups are not distinguishable—all subjects qualitatively show the same effect.

One possible explanation for the different MDTs in fine-structure minima and maxima assumes that the variation of the threshold in quiet with frequency on a fine scale implies that the gain applied by the active processes in the cochlea also varies on a fine scale; the gain is

maximal when the threshold is at a minimum. Such a frequency-specific gain function can be implemented as a filter whose shape is the inverse of the fine structure. Different gain factors in fine-structure minima and maxima would modify the ratio between carrier and sideband levels, thus altering the effective modulation depth. For example, in the Max condition the gain applied to the carrier would be less than the gain applied to the sidebands. This would increase the effective modulation depth, leading to relatively low MDTs. This hypothesis would also predict larger effects for modulation frequencies that match the individual minimum-maximum distance than for those used in the previous studies by Zwicker (1986) and Long (1993). However, if the difference in AM detectability in fine-structure minima and maxima were determined by a spectral filtering alone, then the effective modulation depth at the MDT after filtering should be the same for all conditions, at least within one subject. This is not the case in the current data set. The effective modulation depth was determined from the spectral components of the filtered stimulus at MDT, i.e., from the SL of the spectral components. It fluctuates across conditions within a subject by about 6 dB on average. This is true whether the calculation is based on the mean SL of the two sidebands of a stimulus at MDT or whether it is based on the SL of one of the two sidebands. Thus, the data are not in line with a spectral filtering as the only underlying mechanism. The data also argue against a perception of modulation based on the detection of the spectral sidebands, which could be motivated by assuming narrower auditory filters at levels near the threshold in quiet (Glasberg and Moore, 2000). For in 76% of the cases the levels of the sidebands at the MDT were below threshold (the mean sideband sensation level at MDT was −2.5 dB).

Another explanation for the observed effect could be an interaction of the stimuli with SOAEs that occur at fine-structure minima. If the carrier of the stimulus falls in a fine-structure minimum, it is likely to synchronise an associated SOAE. Then the SOAE would add to the stimulus in phase and thereby reduce the effective modulation depth, making the modulation harder to detect (Long, 1993).

Should the SOAE not be synchronised by the stimulus—for instance, if the fine-structure minimum is not measured exactly—then beating between the stimulus and the SOAE can occur and may to some extent mask the test modulation. However, beating between tones and SOAEs (monaural diplacusis, e.g., Long (1993)) is a very unstable phenomenon, even though it is often reported by subjects. Therefore, an explanation based on modulation masking by beating needs to be further tested in future studies.

In summary, MDTs of sinusoidally amplitude modulated tones near the threshold in quiet are higher when the carrier falls in a fine-structure minimum than when it falls in a fine-structure maximum, regardless of whether the carriers have the same absolute level or the same sensation level. That is, modulation sensitivity near the threshold in quiet is not determined by the level of the carrier above threshold alone, but it depends on the location of the spectral components of the stimulus relative to the fine structure of the threshold in quiet. This implies that neither using equal carrier SPLs nor using equal carrier SLs may lead to comparable MDTs across subjects if the carriers fall in regions with threshold fine structure. A possible explanation for this finding is the interaction of a SOAE with the stimulus.

Acknowledgments

Parts of this study were presented at the "9. Jahrestagung der Deutschen Gesellschaft für Audiologie" in Cologne, Germany. This work was supported by the Deutsche Forschungsgemeinschaft (Grant Nos. GRK591/3, KO942/18-1&2, and SFB/TRR31).

References and links

Glasberg, B. R., and Moore, B. C. J. (**2000**). "Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise," J. Acoust. Soc. Am. **108**, 2318–2328.
Heise, S. J., Verhey, J. L., and Mauermann, M. (**2007**). "Ein effizientes Verfahren zur Bestimmung der Feinstruktur der Hörschwelle [An efficient measurement technique for the detection of hearing threshold fine structure]," Z. Fuer Audiologie, Audiological Acoust. **46**, 126–136.
Heise, S. J., Verhey, J. L., and Mauermann, M. (**2008**). "Automatic screening and detection of threshold fine structure," Int. J. Audiol. **47**, 520–532.
Kohlrausch, A. (**1993**). "Comment on 'Temporal modulation transfer functions in patients with cochlear

implants' [J. Acoust. Soc. Am. **91**, 2156–2164 (1992)]," J. Acoust. Soc. Am. **93**, 1649–1652.
Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.
Long, G. R. (**1993**). "Perceptual consequences of otoacoustic emissions," in *Contributions to Psychological Acoustics: Results of the 6th Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick (University of Oldenburg Press, Oldenburg, Germany).
Mauermann, M., Long, G. R., and Kollmeier, B. (**2004**). "Fine structure of hearing threshold and loudness perception," J. Acoust. Soc. Am. **116**, 1066–1080.
Millman, R. E., and Bacon, S. P. (**2008**). "The influence of spread of excitation on the detection of amplitude modulation imposed on sinusoidal carriers at high levels," J. Acoust. Soc. Am. **123**, 1008–1016.
Moore, B. C. J., and Glasberg, B. R. (**2001**). "Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 1067–1073.
Yost, W. A., and Sheft, S. (**1997**). "Temporal modulation transfer functions for tonal stimuli: Gated versus continuous conditions," Aud. Neurosci. **3**, 401–414.
Zwicker, E. (**1986**). "Spontaneous oto-acoustic emissions, threshold in quiet, and just noticeable amplitude modulation at low levels," in *Auditory Frequency Selectivity*, edited by B. C. J. Moore, and R. D. Patterson (Plenum, New York).

Monitoring progressive damage in polymer-based composite using nonlinear dynamics and acoustic emission

Mourad Bentahar
Laboratoire d'Acoustique de l'Université du Maine (LAUM), UMR CNRS 6613. Université du Maine. Av. Olivier Messiaen, 72085, Le Mans Cedex 9, France and MATEIS INSA de Lyon, 7 Av. Jean Capelle, Bât. Blaise Pascal, 69621 Villeurbanne, France
mourad.bentahar@univ-lemans.fr

Rachid El Guerjouma
Laboratoire d'Acoustique de l'Université du Maine (LAUM), UMR CNRS 6613. Université du Maine. Av. Olivier Messiaen, 72085, Le Mans Cedex 9, France
rachid.el_guerjouma@univ-lemans.fr

Abstract: In this work quantitative results of applying nonlinear acoustic dynamics to study progressive damage in a polymer-based composite SMC (sheet molding compound) are presented. Via carefully controlled resonant plate experiment, nonlinear slow dynamics (SNLD) response of SMC in terms of relaxation time and frequency shift has been shown to be very sensitive to gradual damage induced using three-point bending tests. Besides, acoustic emission monitoring is used to characterize damage through the elastic energy released by SMC at every damage step. Interesting logarithm-like changes of the SNLD parameters as a function of the acoustic emission cumulated energy are found.

PACS numbers: 43.25.Dc, 43.25.Ba, 43.25.Gf, 43.40.Le [MH]
Date Received: April 18, 2008 **Date Accepted:** August 17, 2008

1. Introduction

Hysteretic nonlinear behavior has been shown in the last years to be features characterizing the elastic behavior of a wide range of materials. This behavior is generally related to inhomogeneities within materials such as grain boundaries, fiber matrix interfaces, microcracks, etc. When nonlinear hysteretic materials are subject to a strong harmonic perturbation (conditioning), they show a frequency shift and attenuation increase, both associated to continuous changes in their viscoelastic properties. These properties do not draw back instantaneously to their initial value when perturbation is removed, but follow a very slow recovery. The return to the equilibrium state, which could be observed through resonance frequency change in time, is called slow nonlinear dynamics (SNLD).[1–3] As SNLD was observed on a wide range of materials[4] (cracked glass, rocks, damaged metals, composites, concrete, etc.), it appears difficult to explain it through a unified mechanism due to the diversity of the materials constituents and structures. It logically follows that a better understanding of hysteretic nonlinear mechanisms which happen during and after the dynamic perturbation will make the formulation of predictive models easier for large-scale structural materials. In the open literature, SNLD was used to compare the behavior of two concrete samples (damaged and undamaged) but no definitive conclusion was drawn from these experiments since samples were not the same, and hence no reference parameters were available.[1] More recently, SNLD experiments, made on intact and damaged concrete samples manufactured simultaneously with the same constituents, showed clear evolutions of frequency drop and relaxation time with damage.[3] However, these results must be confirmed on the same sample taken from the intact state and gradually damaged. In this case, damage should be controlled and carefully quantified for a good evaluation of SNLD variations. The aim of this letter is to study the evolution of the dynamical behavior of a gradually damaged polymer-based

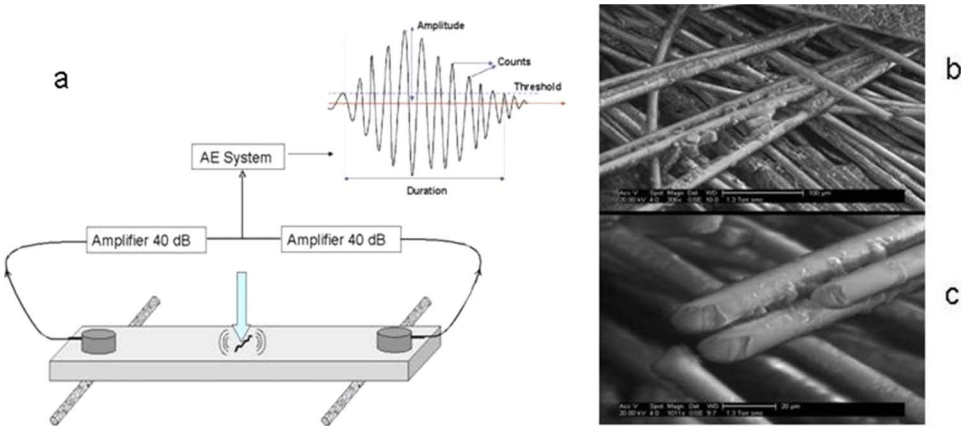


Fig. 1. (Color online) (a) Damage of SMC plate using three-point bending test: the use of acoustic emission system (MISTRAS-2001) allowed real time registration of AE hits (shown inset), (b) fiber/matrix debonding and matrix micro-cracking, and (c) fiber cracking.

composite material through its SNLD response. Gradually induced damage is quantified using acoustic emission (AE). SNLD evolution is then clearly related to AE parameters, particularly to the cumulated energy freed during damage creation and propagation.

2. Material flexural damage process and acoustic emission characterization

The considered material in this study is a glass/fiber polymer composite called SMC (sheet molding compound) of a commercial quality. The relatively low glass fiber volume fraction (30%), the uncontrolled filler, and reinforcement distributions during processing lead to a highly heterogeneous structure with several kinds of stress concentrators, which could create various types of damage at early loading stage.[5] Specimens ($100\times20\times3$ mm^3) were tested with three-point bending (Fig. 1). Gradual damage tests were carried out with a servo-hydraulic Instron universal testing machine with a 5-kN capacity and 0.04 mm/min crosshead velocity. Four damage steps were induced within the same sample at step-off levels from 0 to 1.6 mm in 0.4-mm increments. Simultaneously, information obtained from bending tests (force and displacement) were coupled with acoustic emission (AE) records. MISTRAS-2001 AE system allowed real time registration of the elastic waves emitted by the material during micro-cracks creation and propagation. Numerous bending tests allowed the best calibration of the AE system to individualize AE hits for the best detection of their duration and amplitude using a moving time window technique. To eliminate AE generated from friction, we included an amplitude threshold corresponding to 30 dB, where 0 dB corresponds to 1 μV. This result is supported by previous experimental works which show that, in SMC composites, the main damage mechanisms are matrix cracking, fiber/matrix decohesion and fiber fracture (Fig. 1). For these mechanisms, the AE hits amplitudes ranging from 30 to 100 dB.[6,7]

Parameters of AE hits (duration, amplitude, envelope, rise-time, etc.) could all be used for a damage monitoring strategy. In this work, we were mainly interested in the simultaneous change of AE hits $S(t)$ duration and amplitude through the so-called absolute energy calculated as

$$E=\int_{t_{\text{initial}}}^{t_{\text{final}}} S^2(t)\,dt. \tag{1}$$

Figure 2 presents the information obtained from bending tests coupled with AE records along every damage step. Primary crack initiation could be observed at the first damage step (F_1 $=0.15$ kN) where few AE hits with mainly low and middle amplitudes are detected. Crack mul-

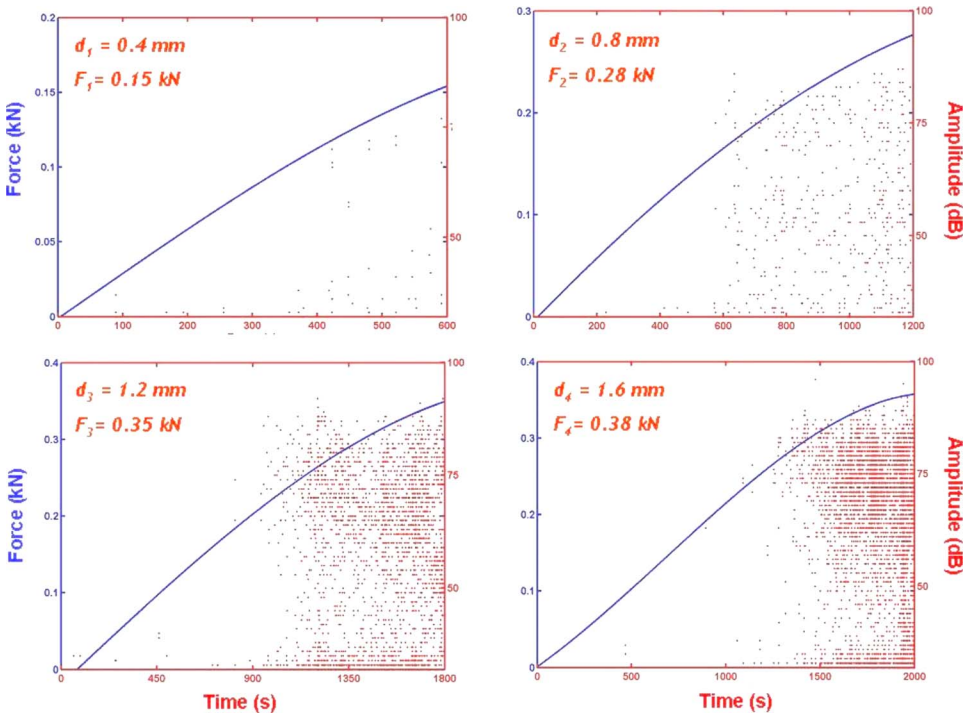


Fig. 2. (Color online) Evolution of the AE hits amplitudes as function of the applied force along every damage step. F_i and d_i are the maximum force and displacement reached at every damage step. The quasi-absence of AE activity could be noticed at every damage step n (n=2,3,4) before the force level F_{n-1} is reached.

tiplicity and growth during the three other tests are characterized by the appearance of new AE hits with low, middle, and high amplitudes with a clear acceleration of the high amplitude events during the fourth test (F_4=0.38 kN). Furthermore, for every damage step, results show the existence of a threshold below which no or little AE is registered. Indeed, in damage step 2, we have no (or little) AE before the maximum applied force in damage step 1 is reached (F_1 =0.15 kN). Equivalent observations could be made using F_2=0.28 kN and F_3=0.35 kN for the third and fourth damage steps, respectively. These interesting observations seem to indicate the existence of a stress memory-damage in accordance with the definition of the so-called *Kaiser effect* (KE) which says that if KE exists, little or no acoustic emission will be recorded before any previous maximum stress level is achieved.[8] Irreversibility of the phenomena at the origin of AE hits in SMC is then experimentally verified with the indication that no additional damage is created between two consecutive tests. Therefore, the energy content of AE hits is a good representation of the elastic energy of the different damage mechanisms that occur in SMC throughout the different loading steps, making the SNLD/AE correlation possible. Results related to this correlation are described in the next section.

3. SNLD of damaged SMC correlated to acoustic emission

For conditioning and SNLD experiments, we have excited the SMC sample continuously around one of its flexural resonance modes, at the initial state and after every damage step. Amplitudes are controlled with a gain/phase analyzer and amplified at 52 dB. Linearity of the experimental device was verified using a high power ultra-broadband coupler and a reference material of atomic nonlinearity (intact aluminum sample).[3] Electronic nonlinear threshold starts at a strain level corresponding to $\varepsilon = 10^{-5}$. For our SNLD experiments this limit is reasonable since hysteretic nonlinear behavior of materials harder than SMC (rocks, concrete, tales,

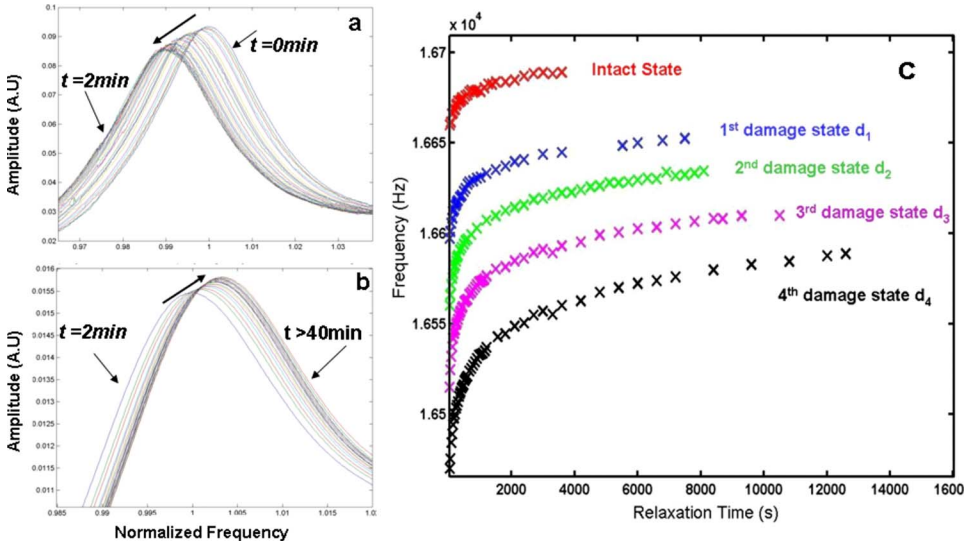


Fig. 3. (Color online) Typical resonance curves obtained for conditioning and relaxation of damaged SMC plates: (a) shift of the resonance frequency at the damage state d_2 during conditioning normalized with respect to the preconditioning resonance $f_{R2}=16\,634$ Hz (note the drop of the amplitude), (b) recovery of the resonance frequency during the slow dynamics of damaged SMC normalized with respect to the first measurable frequency after conditioning is removed $f=16\,557$ Hz, and (c) effect of the induced gradual damage on the SMC log-time relaxation: note the increase of the frequency drop with damage under the effect of the same conditioning.

etc.) is probed for strain rates well bellow this threshold.[3,4,8] In view of the instrumented sample's geometry and the amplifier lower frequency limitation (10 kHz), which prevents from exciting fundamental modes, we developed a laser imaging system to identify the excited modes and determine their sensitivities at different damage steps. After a broadband spectrum sensitivity study, the sixth flexural mode ($f_{R0}=16.685$ KHz) revealed to be the most sensitive to early damage steps. It shall consequently be used for the nonlinear characterization of SMC.

At initial state ($i=0$) and every damage state (d_i, $i=1,2,3,4$), SMC is first excited continuously around the corresponding sixth flexural mode f_{Ri} ($i=0,1,2,3,4$) with a very weak drive level to check if it is completely relaxed (resonance frequency does not change in time). The resonance frequencies f_{Ri} are determined using a third order polynomial interpolation around the maximum of the resonance curves with an accuracy less than 1 Hz. The so found f_{Ri} are registered and used as a reference. For the initial state and each damage state d_i, the SMC sample is then conditioned with a high-level driving tone during 2 min around f_{Ri}. Typical resonance curves obtained for conditioning are presented in Fig. 3(a), where the "fast" frequency shift due to the material softening could be observed. Right after every conditioning, SMC relaxation is probed with the same weak drive level. Figure 3(b) shows the SMC "slow" frequency recovery (relaxation) when conditioning is removed. In these measurements, relaxation time, which is a typical feature of SNLD behavior, is defined as the time spent by SMC to recover its preconditioning reference frequency f_{Ri}. Figure 3(c) shows the conditioning effect on the frequency shift and the relaxation time for the initial state and the four increasing damage states. Qualitatively, SMC relaxation behaves the same way for undamaged and damaged states. Furthermore, the recovery is logarithmic with time, as found in SNLD experiments on several other materials.[4] In addition to the nonlinear behavior of SMC at the intact state, experiments highlight the sensitivity of hysteretic nonlinearity to the presence as well as the evolution of damage. Indeed, the effect of the same conditioning level increases with damage and makes the frequency shift more important at each damage step. Consequently, SMC needs more time to recover its original preconditioning state. Quantitatively, relative variations of the frequency shift as well as relaxation time are from 25% to 170% and from 108% to 250%, respectively, as

Table 1. Quantitative changes in relaxation time and frequency shift during conditioning. Relative variations are computed with respect to the intact state data.

	Frequency shift		Relaxation time	
Damage state	Hz	%	s	%
Intact	44	0	3600	0
D1	55	25	7500	108
D2	74	68	8100	125
D3	95	116	10 500	192
D4	119	170	12 600	250

shown in Table 1. However, as the recovery process is more time consuming to measure, the use of the resonance frequency drop during the 2 min of conditioning is recommended for fast and reliable future structural health monitoring applications.

Despite the sensitivity of SNLD parameters (relaxation time and frequency drop) to damage, a quantification of damage has still to be made. In that sense, we propose the correlation between relaxation time and AE energy presented in Fig. 4. The latter seems to confirm the following intuitive and basic statement: *An intact material, which released a weak amount of elastic energy, would have a quasi-instantaneous (fast) relaxation* $(t \rightarrow 0)$ *in response to a given conditioning. However, if the same conditioning is applied to a highly damaged material, which has lost an important amount of its elastic energy, the relaxation time would be very important* $(t \rightarrow \infty)$. Furthermore, the remarkable result revealed by SNLD/AE correlation is the logarithm-like dependence of the relaxation time evolution as a function of the released energy during the damage process. To our knowledge, such a logarithm correlation is shown for the first time.

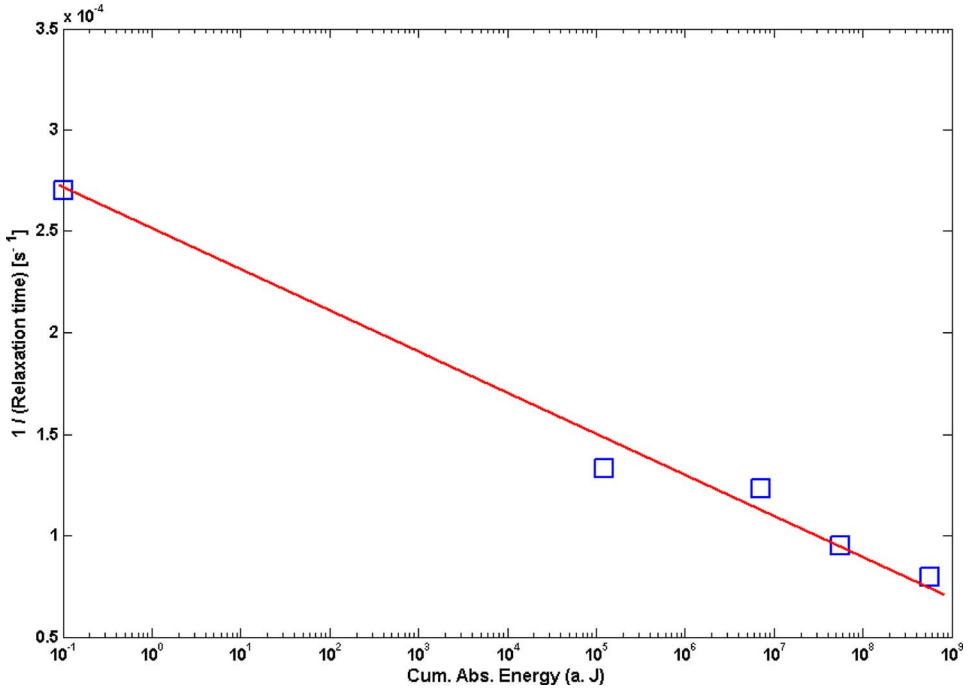


Fig. 4. (Color online) SMC relaxation time correlated to the cumulated elastic energy released during damage creation and propagation: the solid line represents the best fit of the experimental data including the intact state.

However, we still do not know much about the physical origin of this result and/or its universality if compared with the universal SNLD. In particular, we need to know if other materials show the same behavior as SMC and how do phenomena at the origin of their SNLD mechanisms (thermoelastic, hysteretic, or else[1,2,9–11]) influence the correlation with the released elastic energy.

4. Conclusions and prospects

The significant sensitivity of SNLD to progressive damage and its relation to an original damage parameter through the energy of acoustic emission is demonstrated. Despite the valuable knowledge of the SMC hysteretic nonlinear response and its subsequent evolution with a global damage parameter defined through acoustic emission, we still need to understand the time-to failure critical mechanisms involved during damage creation and propagation. The discrimination of the different damage mechanics generated during the different loading steps could be obtained through an adequate classification of AE signals. In view of this, a cluster analysis of AE data has been recently developed in order to correlate clusters to damage mechanisms.[11,12] This promising approach, succeeded in showing the time evolution of the different damage mechanisms occurring in SMC. Therefore, a great opportunity is offered to study in a better way the contribution of each damage mechanisms to the evolution of the hysteretic nonlinear behavior of damaged materials. A work in progress…

Acknowledgments

We wish to thank Dr. V. Tournat, CNRS and Université du Maine, for useful discussions and Dr. P. A. Johnson, Los Alamos National Laboratory, for his encouragements and suggestions.

References and links

[1] J. A. TenCate, E. Smith, and R. Guyer, "Universal slow dynamic in granular solids," Phys. Rev. Lett. **85**, 1020–1023 (2000).
[2] V. Zaitsev, V. Gusev, and B. Castagnède, "Thermoelastic mechanism for logarithmic slow dynamics and memory in elastic wave interactions with individual cracks," Phys. Rev. Lett. **90**(7), 075501 (2000).
[3] M. Bentahar, H. Elaqra, M. Griffa, R. El Guerjouma, and M. Scalerandi, "Hysteretic elasticity in damaged concrete: quantitative analysis of slow and fast dynamics," Phys. Rev. B **73**, 014116 (2006).
[4] P. A. Johnson, and A. Sutin, "Slow dynamics and anomalous nonlinear fast dynamics in diverse solids," J. Acoust. Soc. Am. **117**, 124–130 (2005).
[5] H. G. Kia, in *Sheet Molding Compounds—Science and Technology* (Hanser Publisher, Germany, 1993).
[6] G. Kotsikos, J. T. Evans, A. G. Gibson, and J. M. Hale, "Environmentally enhanced fatigue damage in glass fiber reinforced composites characterized by acoustic emission," Composites, Part A **31**, 969–977 (2000).
[7] H. Nechad, A. Helmstetter, R. El Guerjouma and D. Sornette, "Creep Ruptures in Heterogeneous Materials," Phys. Rev. Lett. **94**, 045501 (2005).
[8] K. E.-A. Van Den Abeele, J. Carmeliet, J. TenCate, and P. A. Johnson, "Nonlinear Elastic Wave Spectroscopy (NEWS) techniques to discern material damage. Part II: Single Mode Nonlinear Resonance Acoustic Spectroscopy," Res. Nondestruct. Eval. **12**, 31–43 (2000).
[9] P. P. Delsanto, and M. Scalerandi, "Modeling nonclassical nonlinearity, conditioning, and slow dynamics effects in mesoscopic elastic materials," Phys. Rev. B **68**, 064107 (2003).
[10] J. G. Williams and A. Pavan, *Fracture of Polymers, Composites and Adhesives* (Elsevier Science, Amsterdam, 2000).
[11] O. O. Vakhnenko, V. O. Vakhnenko, T. J. Shankland, and J. A. TenCate, "Strain-induced kinetics of intergrain defects as the mechanism of slow dynamics in the nonlinear resonant response of humid sandstone bars," Phys. Rev. E **70**, 015602 (2004).
[12] A. Marec, J.-H. Thomas, and R. El Guerjouma, "Damage characterization of polymer-based composite materials: multivariable analysis and wavelet transform for clustering acoustic emission data," Mech. Syst. Signal Process. **22**, 1441–1464 (2008).

Observations of low-frequency temporal and spatial coherence in shallow water

Harry A. DeFerrari
Division of Applied Marine Physics, RSMAS—University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149
hdeferrari@rsmas.miami.edu

Abstract: Measurements of temporal and spatial coherence are most always confounded by the presence of multipath interference. Here, two sets of data are presented that allow separation of arrivals from individual propagating modes and then unambiguous computations of temporal and spatial coherence free of multipath effects. A consistent finding for surface reflected bottom reflected (SRBR) modes is that lower order mode are more coherent in both time and space than higher order (steeper) modes. Also, SRBR paths are much more coherent than refracted bottom reflected mode groups that cannot be separated in time and are affected by multi-mode interference.

PACS numbers: 43.30.Re [JL]
Date Received: January 8, 2008 **Date Accepted:** July 18, 2008

1. Introduction

The data presented here are a small subset from a shallow-water acoustic propagation experiments conducted off the coast of south Florida near Ft. Lauderdale and a preliminary calibration experiment at the same range site. The Florida Straits Propagation Experiments, FSPE, transmitted from a moored source (Nguyen, 1996; DeFerrari, 2003) to a vertical array at a range of 10 km. The *m*-sequence broadband transmissions were continuous for a 1-month period but switching between each of six center frequencies, 100, 200, 400, 800, 1600, and 3200 Hz at 1-h intervals. These data have negligible source motion and are well suited for temporal coherence calculations.

The calibration experiments (CE) of interest consisted of a 20-min transmission of a 250-Hz *m*-sequence from a shallow source suspended from a surface ship that held station by dynamic positioning. The signals were received by a long horizontal array at near broadside at a 10-km range. These data are used for spatial coherence calculations.

2. Modes and arrivals

Broadband parabolic equation (PE) and normal mode propagation models predict pulse arrival times that agree well with observations (DeFerrari, 2003) for the FSPE as shown in Fig. 1. The PE calculation shows the progression of the pulse response down the propagation channel out to 20 km. Classic mode dispersion curves are apparent for the surface reflected bottom reflected (SRBR) modes. The refracted bottom reflected (RBR) modes are known to all have nearly identical group velocities owing to the approximate hyperbolic cosine shape of the sound-speed profile. As a result, all RBR modes have about the same travel times and overlap as a single intense unresolved group. The observed pulse responses generally agree well with the predicted arrival pattern at the 10-km range for the 800-Hz signal (as shown) and for the other frequencies (not shown). The number of arrivals, N, depends on frequency and has been observed to be roughly $N=f/100$. There are usually N resolved SRBR arrivals and N unresolved RBR arrivals.

The deep source of the FSPE ensonifies both SRBR and RBR modes/paths, whereas the shallow source of the CE, in the surface layer, only ensonifies SRBR modes/paths. The comparison and computations that follow focus on the resolvable SRBR modes/paths for both experiments.

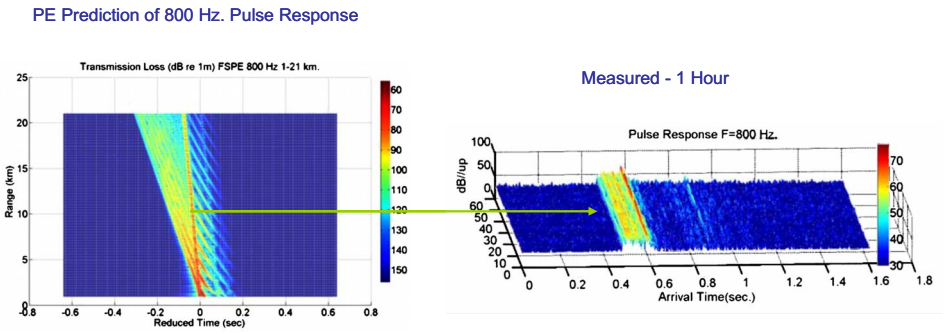


Fig. 1. (Color online) PE model predictions of the channel pulse response compared with observations for the 10-km range of the FSPE.

3. Temporal coherence

Coherence is a statistical measure of the similarity of a waveform either at different time intervals (temporal coherence) or at different distance intervals along an array (spatial coherence). It is a complex quantity that depends on both the amplitude and phase of the waveform.

The coherency formula of Eq. (1) is a time-lagged, normalized covariance calculation,

$$\mathrm{COH}(t,\tau) = \frac{\langle (p(t)^* p(t+\tau))^2 \rangle_{\Delta t,\Delta T}}{\langle p(t)^2 \rangle_{\Delta t,\Delta T} \langle p(t+\tau)^2 \rangle_{\Delta t,\Delta T}}. \tag{1}$$

Here, it is applied to an array of collected data like that shown in Eq. (2). Each row is a channel pulse response, $p(t)$, that is, the complex pulse amplitude as a function of arrival time obtained after matched filter pulse compression of a received m-sequence transmission. Usually, several sequential periods of m-sequences are coherently averaged for a single pulse response $p(t)$ so that the rows sample the pulse response in experimental time of interval equal to the coherent averaging time—usually about a minute.

$$P(T,t) = \begin{bmatrix} \cdots p_{T=1}(t) \cdots \\ \tau \quad t \rightarrow \cdots p_{T=2}(t) \cdots \\ \downarrow \cdots p_{T=3}(t) \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots p_{T=N}(t) \cdots \end{bmatrix} \tag{2}$$

The lagged cross product of Eq. (1) is calculated on the time history of sections of $p(t)$, i.e., the columns in Eq. (2). One could select a section of $p(t)$, DT wide, spanning an arrival of interest by time gating data containing the pulse arrival and then performing the lagged product with the column of time-gated data. Instead, the calculation is repeated for all columns, noise, and signal alike. The result is a coherency for every pulse period of the received signal as shown in Fig. 2. Each estimate of COH has 30 degrees of freedom.

With reference to Fig. 2 top, the first distinct arrival is the lowest-order SRBR arrival, one that has made seven bottom interactions. It is the most coherent arrival, Fig. 2 bottom), with a coherency of 0.95 after 30 min. The second and third arrivals are less coherent (20 and 10 min, respectively) and the most intense later arrival is the unresolved RBR group that decorrelates in 2.5 min. A group of still later arrivals also have long coherence times. These are

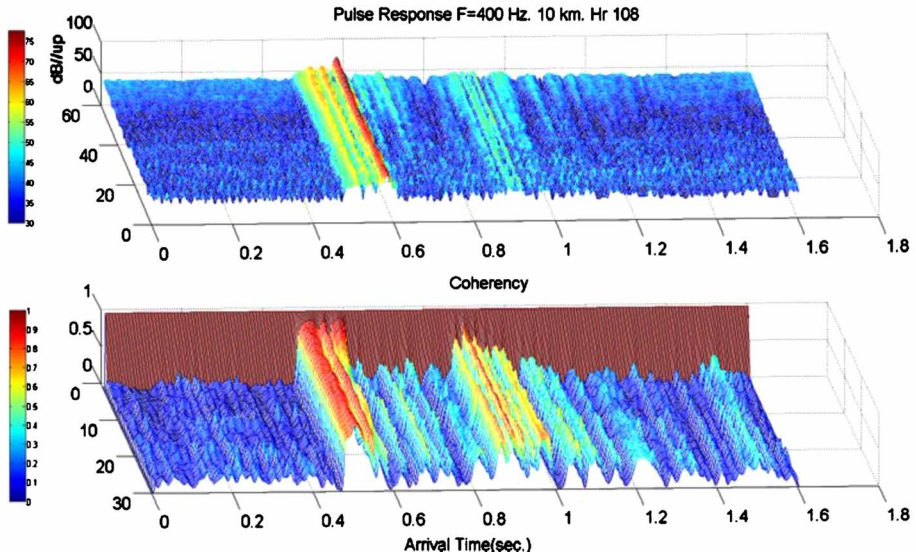


Fig. 2. (Color online) Top: Received pulse response (intensity (dB) vs arrival time (t) for a 1-h time history. Bottom: Coherence COH (tau) for the signal arrivals (and noise) of the top pulse response.

thought to be out-of-plane arrivals that travel up, and then back along the in-shore slope. As such, they are true Fermat paths and even though low in amplitude they are stable and time coherent.

The above observation of higher coherence for lower-order modes is consistent for all frequencies but only on the average. There are large differences from one hour-long sample to another owing to noise and local environmental conditions. Table 1 summarizes the average coherent times for the first arriving SRBR mode and the RBR group arrival. Fourteen days of data were used for the averages. The result required identifying and tracking the arrival by hand, as automated tracking is less reliable.

4. Spatial coherence

The data for the spatial coherence calculation were collected during a continuous 20-min long transmission of a 4.5-s period m-sequence. The carrier frequency was 250 Hz and each digit was four cycles in duration. Shown in Fig. 3 left) are the observed pulse responses for a single hydrophone during the first 5 min. The source turned on after 1.5 min and the time history of individual 4.5-s pulse responses are plotted in sequence. The spatial coherence calculation that follows uses only one of the pulse responses of the time series selected at a time when the source motion was minimal. The display shows the shape of the wavefront for each of the arriving SRBR modes. The first is clearly the most intense, stable, and well-aligned with the array, even though slightly tilted.

The spatial coherence calculation is almost exactly the same as the temporal coherence calculation of Eq. (1). The only difference is that the hydrophone spacing,χ, is substituted for

Table 1. Observed decorrelation times to a value of COH=0.75.

Frequency (Hz)	SRBR mean/max	RBR mean/max
200	15/$\gg$30 (min.)	8.3/$>$30
400	4.2/$>$30	3.4/12
800	2.8/6.5	2.1/4.5

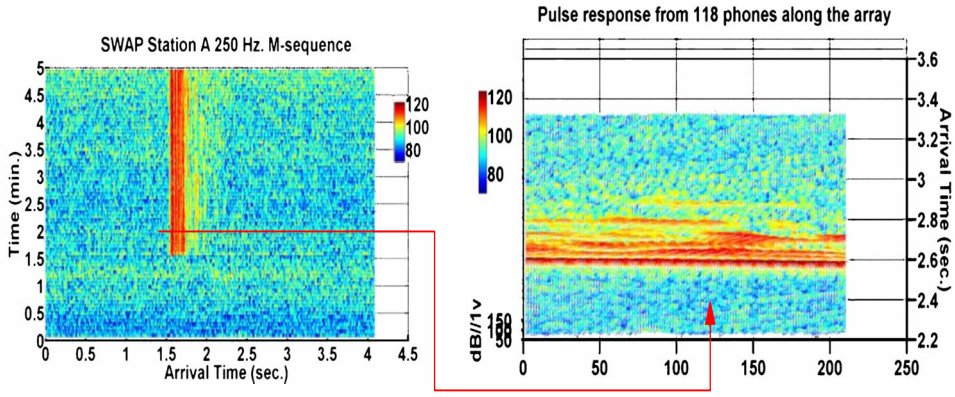


Fig. 3. (Color online) Left: Time history of pulse responses from 250-Hz *m*-sequence transmission. Right: A single pulse response at each of 118 elements across an array at broadside.

the lag time, τ. The spatial coherency calculation is shown in Fig. 4 left. The coherency function is seen to oscillate through four cycles while the magnitude remains high. This is likely a result of misalignment of the wavefront and the array, producing a phase shift proportional to distance along the array. The coherence calculation will oscillate owing to interference at regular intervals.

In order to remove the phase roll, the array can be steered in very fine increments by the following time-shifting theorem calculations at each array element:

$$P(\omega) = F(p(t)),$$

$$P_\tau(\omega) = P(\omega)e^{i\omega\tau},$$

$$p(t-\tau) = F^{-1}(P_\tau(\omega)), \tag{3}$$

where F is the Fourier transform.

The calculation produces an exact time shift by τ without distortion of the waveform. The array is steered by time shifting in proportion to the distance along the array.

Figure 4 left shows the coherency before the array is steered. Figure 4 right shows the best alignment of array angle and wavefront. The coherency has no oscillations and the first arrival is seen to be coherent (unchanging) to well past 100 m, while later arrivals, associated

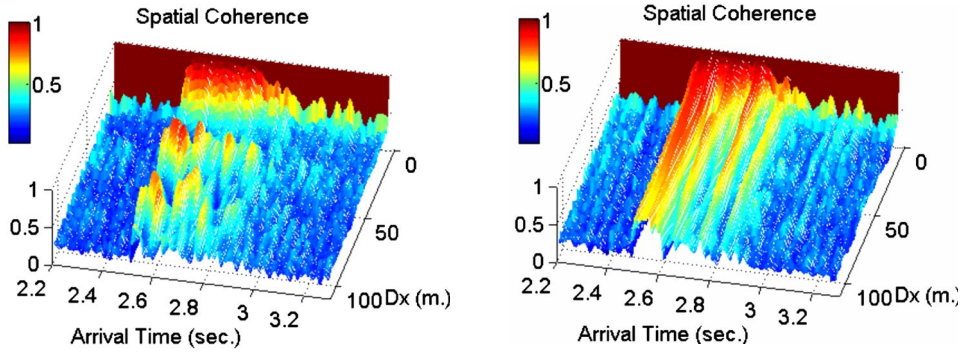


Fig. 4. (Color online) Left: Spatial coherence calculation before the array is steered. Right: Optimal spatial coherence with best alignment of wavefront and array.

with higher modes, are only coherent over shorter distances, a result similar to the temporal coherence measurement.

5. Discussion

Coherence is a statistical measure of the differences in waveform after a time or space interval. Generally, this means phase coherence since the first fluctuations to occur are phase shifts in time or along an array. But, this may not necessarily be a true distortion of the waveform but rather a deterministic effect. Flatte (1979) discusses "phase wrapping" as a special case when the phase changes slowly without randomizing the signal. Likewise, a misalignment of an array with an arriving wavefront will cause a similar deterministic effect. One can generally align the array or phase track slow changes. Both are recoverable effects in practical application. True randomizing effects are unrecoverable. The latter effect distorts the waveform of the individual mode or ray arrival.

The data reported here and the *m*-sequence signal processing avoid multipath interference in the coherence calculations. On average, multipath reduces decorrelation times by a factor of 2 to 3 (Table 1) and in some instances by as much as a factor of 10 (Fig. 2), as compared to the decorrelaton time of a single mode arrival. For all data at all frequencies, observed pulse arrivals were distinct and separable in time and each could be associated with a particular mode of propagation with little ambiguity. Steeper higher-order modes have reduced coherence in both space and time. These consistent finding suggest that the same mechanisms are at play. The angular dependence of mode's bottom interaction angle seem the most likely cause of wavefront distortion as opposed to volume scattering from irregularities in the propagation medium.

Acknowledgments

Work was sponsored by ONR.

References and links

DeFerrari, H. A., Williams, N., and Nguyen, B. H. (**2003**). "Focused arrivals in shallow water propagation in the Straits of Florida" ARLO **4**, 106-111.

Flatté, S. M. Dashen, R., Munk, W. M., Watson, K. M., and Zachariasen, F. (**1979**). *Sound Transmission Through a Fluctuating Ocean*, Mechanics and Applied Mathematics (Cambridge University Press, Cambridge).

Nguyen, B. H., DeFerrari, H. A., and Williams, N. (**1996**). "General purpose autonomous transmitter and receiver system for underwater acoustic experiments," IEEE J. Ocean. Eng. **21**(1), 85–93.

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.

Sound scattering from two concentric fluid spheres (L)

Jared McNew,[a] Roberto Lavarello, and William D. O'Brien, Jr.
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 405 North Mathews, Urbana, Illinois 61801

(Received 29 April 2008; revised 29 October 2008; accepted 4 November 2008)

The solution to the problem of plane wave and point source scattering by two concentric fluid spheres is derived. The effect of differences in sound speed, density, and absorption coefficient is taken into account. The scattered field is then found in the limit as the outer sphere becomes an infinitely thin shell and compared to the solution for a single fluid sphere for verification. A simulation is then performed using the concentric fluid sphere solution as an approximation to the human head and compared to the solution of a single fluid sphere with the properties of either bone or water. The solutions were found to be similar outside of the spheres but differ significantly inside the spheres. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3035901]

PACS number(s): 43.20.Fn [TDM] Pages: 1–4

I. INTRODUCTION

The concentric sphere geometry can be used to approximate many applications. For example, scattering from single cells could be modeled with the concentric sphere model with the inner sphere having the properties of the nucleus and the outer sphere the properties of the cytoplasm.[1] Ultrasound contrast agents are composed of microbubbles enclosed in a polymer, protein, or lipid shell which could also be analyzed using the concentric sphere model.[2] The finite element method is also currently being applied to model sound wave propagation into the human head.[3] It is important to validate the model with geometries that have analytical solutions to test the accuracy of the model. One such geometry is that of two concentric fluid spheres, where the outer sphere has the bulk fluid properties of bone (neglecting the presence of shear waves) and the inner sphere has the properties of water.

Past publications have dealt with situations similar to this but are limited in their application. For example, Goodman and Stern[4] derived the solution to plane wave scattering from an elastic shell in a fluid medium but the medium and inner sphere were assumed to have identical properties. Kakogiannos' and Roumeliotis'[5] solution is limited to spheres whose radii are small relative to a wavelength. The present work combines Sinai and Waag's solution to plane wave scattering from concentric fluid cylinders[6] and Anderson's solution to plane wave scattering from a single fluid sphere[7] to derive the solution to plane wave scattering from two concentric fluid spheres. These solutions are also extended to include attenuation and point sources.

Other publications have solved more general problems involving concentric spheres. For example, Gerard and co-workers[8,9] used resonant scattering theory as a framework to derive solutions to scattering by spherical elastic layers. Martin[10] derived the solution to concentric fluid spheres when the properties of the outer sphere are specific functions of the distance from the center of the sphere. In both cases, the solution to scattering from two concentric fluid spheres can be synthesized but require significant manipulation of the provided results. The contribution of the present work is to provide simple expressions that can be readily used to calculate fields scattered by concentric fluid spheres when shear waves can be neglected.

II. ABBREVIATIONS

The following abbreviations are used in this paper.
P_m = Legendre polynomial
j_m = spherical Bessel function
$h_m^{(k)}$ = spherical Hankel function

III. THEORY

For computational simplicity, the spheres are placed at the origin of a spherical coordinate system (r, θ, ϕ), as shown in Fig. 1. The source is either a plane wave propagated in the $-z$ direction or a point source located on the positive z-axis at a distance R from the origin which eliminates any dependence on ϕ.

[a]Electronic mail: jmcnew2@brl.uiuc.edu

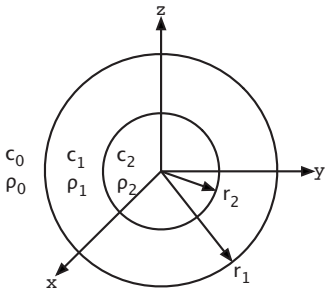


FIG. 1. Physical properties. Infinite medium with density ρ_0, sound speed c_0, and absorption coefficient α_0. The outer sphere has density ρ_1, sound speed c_1, absorption coefficient α_1, and radius r_1. The inner sphere has density ρ_2, sound speed c_2, absorption coefficient α_2, and radius r_2. The spheres are centered at the origin of a spherical coordinate system (r,θ,ϕ), where r is the radial coordinate, θ is the azimuthal coordinate, and ϕ is the polar coordinate.

The pressure in the infinite medium p_0 is the sum of the incident pressure p_{0i} and the scattered pressure p_{0r}.[11]

$$p_0 = p_{0i} + p_{0r} \tag{1}$$

$$p_{0i} = \mathcal{P}_0 \sum_{m=0}^{\infty} (2m+1)\mathcal{L}_m P_m(\mu) j_m((k_0 + i\alpha_0)r) e^{-i\omega t}, \tag{2}$$

$$p_{0r} = \sum_{m=0}^{\infty} A_m P_m(\mu) h_m^{(1)}((k_0 + i\alpha_0)r) e^{-i\omega t}, \tag{3}$$

where $\mu=\cos(\theta)$ and $\mathcal{L}_m$ is given by[12]

$$\mathcal{L}_m = \begin{cases} (-i)^m & \text{plane wave} \\ h_m^{(1)}((k_0 + i\alpha_0)R) & \text{monopole.} \end{cases} \tag{4}$$

The pressure in the outer sphere p_1 is the sum of a standing wave, p_{1r}, and a traveling wave, p_{1i}.

$$p_1 = p_{1i} + p_{1r}, \tag{5}$$

$$p_{1i} = \sum_{m=0}^{\infty} B_m P_m(\mu) h_m^{(1)}((k_1 + i\alpha_1)r) e^{-i\omega t}, \tag{6}$$

$$p_{1r} = \sum_{m=0}^{\infty} C_m P_m(\mu) j_m((k_1 + i\alpha_1)r) e^{-i\omega t}. \tag{7}$$

The pressure in the inner sphere can be written as

$$p_{2i} = \sum_{m=0}^{\infty} D_m P_m(\mu) j_m((k_2 + i\alpha_2)r) e^{-i\omega t}. \tag{8}$$

Applying the boundary conditions of continuity of pressure and radial velocity at the two interfaces, a system of four equations with four unknowns results. This system can be written in the matrix form:

$$\begin{pmatrix} \frac{h_m^{(1)\prime}(\tilde{k}_0 r_1)}{Z_0} & \frac{h_m^{(1)\prime}(\tilde{k}_1 r_1)}{-Z_1} & \frac{j_m'(\tilde{k}_1 r_1)}{-Z_1} & 0 \\ -h_m^{(1)}(\tilde{k}_0 r_1) & h_m^{(1)}(\tilde{k}_1 r_1) & j_m(\tilde{k}_1 r_1) & 0 \\ 0 & \frac{h_m^{(1)\prime}(\tilde{k}_1 r_2)}{Z_1} & \frac{j_m'(\tilde{k}_1 r_2)}{Z_1} & \frac{j_m'(\tilde{k}_2 r_2)}{-Z_2} \\ 0 & h_m^{(1)}(\tilde{k}_1 r_2) & j_m(\tilde{k}_1 r_2) & -j_m(\tilde{k}_2 r_2) \end{pmatrix} \times \begin{pmatrix} A_m \\ B_m \\ C_m \\ D_m \end{pmatrix} = \begin{pmatrix} \frac{-\mathcal{P}_0}{Z_0}(2m+1)\mathcal{L}_m j_m'(\tilde{k}_0 r_1) \\ \mathcal{P}_0(2m+1)\mathcal{L}_m j_m(\tilde{k}_0 r_1) \\ 0 \\ 0 \end{pmatrix}. \tag{9}$$

where $\tilde{k}_n$ is the complex wave number, $\tilde{k}_n = k_n + i\alpha_n$, and

$$Z_n = \frac{\rho_n c_n}{1 + i\frac{\alpha c_n}{\omega}}. \tag{10}$$

The coefficients can then be solved for analytically using Cramer's rule or numerically using LU decomposition.

Often, only the scattered pressure needs to be computed so the values of B_m, C_m, and D_m do not need to be calculated. One can show that by direct manipulation of Eq. (9), the value of A_m is

$$\begin{aligned} A_m = {} & \mathcal{P}_0(2m+1)\mathcal{L}_m \\ & \{[Z_{r2} j_m(\tilde{k}_1 r_2) j_m'(\tilde{k}_2 r_2) - Z_{r1} j_m'(\tilde{k}_1 r_2) j_m(\tilde{k}_2 r_2)] \\ & \times [Z_{r1} h_m'^{(1)}(\tilde{k}_1 r_1) j_m(\tilde{k}_0 r_1) - h_m^{(1)}(\tilde{k}_1 r_1) j_m'(\tilde{k}_0 r_1)] \\ & - [Z_{r1} h_m'^{(1)}(\tilde{k}_1 r_2) j_m(\tilde{k}_2 r_2) - Z_{r2} h_m^{(1)}(\tilde{k}_1 r_2) j_m'(\tilde{k}_2 r_2)] \\ & \times [j_m(\tilde{k}_1 r_1) j_m'(\tilde{k}_0 r_1) - Z_{r1} j_m'(\tilde{k}_1 r_1) j_m(\tilde{k}_0 r_1)]\} \\ & \div \{[Z_{r1} h_m'^{(1)}(\tilde{k}_1 r_2) j_m(\tilde{k}_2 r_2) - Z_{r2} h_m^{(1)}(\tilde{k}_1 r_2) j_m'(\tilde{k}_2 r_2)] \\ & \times [h_m'^{(1)}(\tilde{k}_0 r_1) j_m(\tilde{k}_1 r_1) - Z_{r1} h_m^{(1)}(\tilde{k}_0 r_1) j_m'(\tilde{k}_1 r_1)] \\ & - [Z_{r1} j_m'(\tilde{k}_1 r_2) j_m(\tilde{k}_2 r_2) - Z_{r2} j_m(\tilde{k}_1 r_2) j_m'(\tilde{k}_2 r_2)] \\ & \times [h_m'^{(1)}(\tilde{k}_0 r_1) h_m^{(1)}(\tilde{k}_1 r_1) - Z_{r1} h_m^{(1)}(\tilde{k}_0 r_1) h_m'^{(1)}(\tilde{k}_1 r_1)]\}, \end{aligned} \tag{11}$$

where

$$Z_{r1} = \frac{\rho_0 c_0}{\rho_1 c_1}\left[\frac{1 + \frac{i\alpha_1 c_1}{\omega}}{1 + \frac{i\alpha_0 c_0}{\omega}}\right] \tag{12}$$

and

$$Z_{r2} = \frac{\rho_0 c_0}{\rho_2 c_2}\left[\frac{1 + \frac{i\alpha_2 c_2}{\omega}}{1 + \frac{i\alpha_0 c_0}{\omega}}\right]. \tag{13}$$

IV. VERIFICATION

One method of verifying the solution is to take the limit as the radius of the inner sphere, r_2, approaches the radius of the outer sphere, r_1. It can be shown that the coefficient for the scattered pressure in the infinite medium, A_m, takes on the following value as $r_1 \to r_2$:

$$A_m = \frac{(2m+1)\mathcal{P}_0\mathcal{L}_m[j_m'(\tilde{k}_2 r_1)j_m(\tilde{k}_0 r_1)Z_0 - j_m'(\tilde{k}_0 r_1)j_m(\tilde{k}_2 r_1)Z_2]}{-j_m'(\tilde{k}_2 r_1)h_m^{(1)}(\tilde{k}_0 r_1)Z_0 + h_m'^{(1)}(\tilde{k}_0 r_1)j_m(\tilde{k}_2 r_1)Z_2}. \tag{14}$$

If loss is no longer considered, $\tilde{k}_n$ becomes k, and Z_n becomes $\rho_n c_n$. Making these substitutions,

$$A_m = \frac{(2m+1)\mathcal{P}_0\mathcal{L}_m[j_m'(k_2 r_1)j_m(k_0 r_1)\rho_0 c_0 - j_m'(k_0 r_1)j_m(k_2 r_1)\rho_2 c_2]}{-j_m'(k_2 r_1)h_m^{(1)}(k_0 r_1)\rho_0 c_0 + h_m'^{(1)}(k_0 r_1)j_m(k_2 r_1)\rho_2 c_2}, \tag{15}$$

which is identical to Anderson's solution for the single fluid sphere[7] after some algebraic manipulation.

V. SIMULATION

Three simulations were performed to compare the solution found using the concentric fluid sphere model to the single fluid sphere model using a frequency of 12.5 kHz. The first simulation approximated the human head as an outer sphere of bone (r_1=75 mm, ρ_1=2000 kg/m^3, and c_1=2900 m/s) surrounding an inner sphere of water (r_2=65 mm, ρ_2=1000 kg/m^3, and c_2=1500 m/s) placed in an infinite medium of air (ρ_0=1.21 kg/m^3 and c_0=343 m/s). The second simulation used Anderson's single fluid sphere solution to simulate a fluid sphere of bone (a=75 mm, ρ'=2000 kg/m^3, and c'=2900 m/s) placed in an infinite medium of air (ρ=1.21 kg/m^3 and c=343 m/s). The third simulation used Anderson's single fluid sphere solution to simulate a fluid sphere of water (a=75 mm, ρ'=1000 kg/m^3, and c'=1500 m/s) placed in an infinite medium of air (ρ=1.21 kg/m^3 and c=343 m/s). The magnitude of the pressure along the z-axis is plotted in Fig. 2.

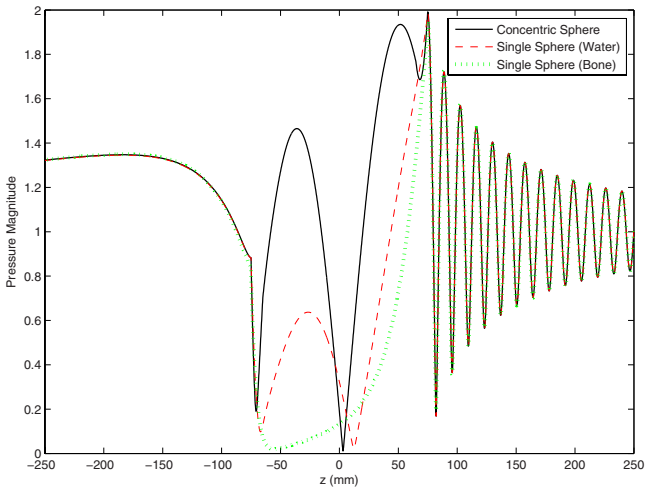


FIG. 2. (Color online) Pressure magnitude along the z-axis for concentric fluid sphere and single fluid sphere solutions.

As can be seen in Fig. 2, the pressure outside of the spheres is nearly identical for all three solutions. This is expected because for all cases the impedance mismatch between the scatterer and the background medium is very large, and therefore the scattered field approaches the limiting rigid sphere case. Inside of the spheres, however, the three solutions differ significantly.

VI. CONCLUSIONS

The solution to plane wave and point source scattering from two concentric fluid spheres was derived. The effects of differences in speed of sound, density, and attenuation coefficient were included. The coefficient required to solve for the scattered pressure was explicitly computed for cases only requiring the scattered pressure. The limit as the outer sphere becomes a thin shell is found and found to agree with Anderson's solution to a single fluid sphere. Finally, the solution is found to a concentric sphere approximation of the human head and compared to approximations of the human head as a single fluid sphere. It was found that outside of the scatterer, the solutions are similar but differ significantly inside the scatterer.

ACKNOWLEDGMENT

Work supported by AFOSR FA9550-06-0128.

[1]M. L. Oelze and W. D. O'Brien, Jr., "Application of three scattering models to the characterization of solid tumors in mice," Ultrason. Imaging **28**, 83–96 (2006).

[2]J. S. Allen, D. E. Kruse, P. A. Dayton, and K. W. Ferrara, "Effect of coupled oscillations on microbubble behavior," J. Acoust. Soc. Am. **114**, 1678–1690 (2003).

[3]W. D. O'Brien, Jr., and Y. Liu, "Evaluation of acoustic propagation paths into the human head," in *HFM Symposium on New Directions for Improving Audio Effectiveness* (RTO, Neuilly-sur-Seine, France, 2005), pp. 15-1–15-24; Available www.rto.nato.int/abstracts.asp (Last viewed August 2007).

[4]R. R. Goodman and R. Stern, "Reflection and transmission of sound by elastic spherical shells," J. Acoust. Soc. Am. **34**, 338–344 (1962).

[5]N. Kakogiannos and J. Roumeliotis, "Acoustic scattering from a sphere of small radius coated by a penetrable one," J. Acoust. Soc. Am. **98**, 3508–3515 (1995).

[6]J. Sinai and R. Waag, "Ultrasonic scattering by two concentric cylinders," J. Acoust. Soc. Am. **83**, 1728–1735 (1988).

[7]V. Anderson, "Sound scattering from a fluid sphere," J. Acoust. Soc. Am. **22**, 426–431 (1950).
[8]A. Gerard, "Scattering by spherical elastic layers: Exact solution and interpretation for a scalar field," J. Acoust. Soc. Am. **73**, 13–18 (1983).
[9]A. Gerard, H. Überall, and A. Guran, "Generalized series for acoustic scattering from objects of separable geometric shape," Acta Mech. **132**, 147–176 (1999).
[10]P. A. Martin, "Acoustic scattering by inhomogeneous spheres," J. Acoust. Soc. Am. **111**, 2013–2018 (2002).
[11]P. Morse, *Vibration and Sound*, 2nd ed. (McGraw-Hill, New York, 1948).
[12]R. F. Harrington, *Time-Harmonic Electromagnetic Fields* (McGraw-Hill, New York, 1961).

Toward the acoustical characterization of asphalt pavements: Analysis of the tire/road sound from a porous surface (L)

S. E. Paje,[a] M. Bueno, U. Viñuela, and F. Terán
Laboratory of Acoustics Applied to Civil Engineering (LA²IC), Department of Applied Physics, ETSI Caminos and E.S. Informática, Universidad de Castilla-La Mancha (UCLM), 13071 Ciudad Real, Spain

(Received 11 January 2008; revised 21 October 2008; accepted 21 October 2008)

Sound level in close proximity to the contact patch of the tire and road (L_{CPtr}) is analyzed as a function of the vehicle speed in the acoustic frequency range, showing different behavior depending on frequency. At least two regions are observed; one at low frequencies, where the variation in sound with speed (coefficient B) increases with frequency, and the other at higher frequencies, where such sound/speed variation does not increase significantly with frequency. The dependence of B at low frequencies seems to be correlated with the high sound absorption of this porous surface. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3025911]

PACS number(s): 43.50.Lj [KA] Pages: 5–7

I. INTRODUCTION

Close proximity measurements of the sound generated by the interaction between tire and road surface were carried out with an ordinary passenger car. The aim of this study was to investigate some characteristics of the rolling sound from a bituminous porous surface. The close proximity method provides a good measure of the tire/pavement sound levels (L_{CPtr}), in close proximity to the contact patch, and should provide insight into the acoustical characterization of asphalt pavements.[1]

A Pirelli P6000, 205/55 R16 V passenger car tire was selected for this study as reference tire, and a conventional bituminous porous surface PA-12 (Spanish denomination) on the road A-41 was selected as the asphalt pavement to be investigated. This type of porous surface could be considered representative of bituminous porous layers found in road surfaces throughout Spain and also in the rest of Europe.

The test section for the acoustic measurements was chosen so that the influence of other vehicles on the acoustical measurements was eliminated; the measurements were developed on a test section where the road was closed to traffic, and before its coming into service. All the close proximity tire/pavement sound tests were carried out in a georeferenced way by means of a global positioning system (GPS) card to improve subsequent analysis. The test section for this study was approximately 1 km long, from A-41 kilometric post 175 to A-41 kilometric post 174, in Ciudad Real, Spain. The road surface was a layer with an average thickness of 5 cm of bituminous materials of a coarse texture (see inset of Fig. 1) and nominal aggregate sizes of 6–12 (59%) mm and 12–18 (21%) mm. It had 4.5 wt % of a modified binder type Bm 3a and an air void content of around 21%. The mix was laid on the road one month before the measurements for this study. Details of the acoustical setup and measurement technique have been given elsewhere.[1–3]

The sound levels correspond to the average value between two microphones situated near the contact patch. During the acoustical tests, the one-third octave band levels [L_{CPtr}dB(A)] were continuously measured every 0.2 s. The levels throughout the test track were correlated with their global position coordinates for better post data processing.

II. CLOSE PROXIMITY SOUND MEASUREMENTS AND DISCUSSION

The tire/pavement sound level from the interaction of the reference tire and the porous surface PA-12 was analyzed in each third-octave band between 200 Hz and 16 kHz. Different tests of the evolution of L_{CPtr} with speed between 130 and 30 km/h were developed. From these measurements in which the speed decreased continuously, values of the coefficients B and A were obtained after the fitting to the relationship $L_{CPtr}=A+B\log_{10}(V)$, where B is the so-called speed constant. These parameters are used to acoustically characterize asphalt pavements and/or road surfaces as well as their respective conservation states.[1–3] The values of these coefficients are about 30 and 40 dB(A), respectively. Different values of coefficients B and A were obtained for an old PA-12 bituminous porous road surface.[2] Figure 1 shows that global levels generated by the interaction of the tire and the bituminous porous surface increase linearly proportional to approximately $30\log_{10}(V)$.

To provide further insight into the characterization of the sound from the bituminous porous surface, additional analysis of the vehicle speed effect is necessary. An additional study shows that the slope values of the logarithmic regression of the sound level and speed (coefficient B) for the 1250 and 2500 Hz bands (Fig. 2), which are attributed to a mechanism associated with air pumping and other mechanisms related to the flow of air in and around the tread grooves of the tire, are 35 and 39, respectively. This result is similar to the findings of the study carried out by Wik and Millar,[4] which showed that sound from air pumping increases proportional to $40\log_{10}(V)$. The sound levels of the 630 Hz band measured at the two microphone positions exhibit an almost con-

[a]Author to whom correspondence should be addressed. Electronic mail: santiago.exposito@uclm.es

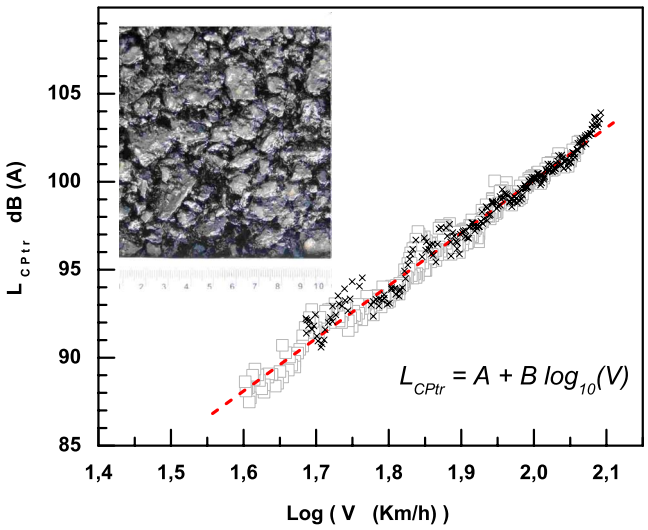


FIG. 1. (Color online) Linear regression of the global sound level of the tire/pavement interaction (□) vs speed. (×) Result of another measurement (run). The inset shows a photograph of the porous surface studied, and a rule is in centimeters.

stant relationship with the logarithm of vehicle speed (see Fig. 2), compared with the above mentioned bands, where the level increases relatively rapidly with speed. The relative importance of certain frequencies on the close proximity sound as a function of speed can be seen in this figure.

The spectrum of the close proximity sound-speed slopes (coefficient B) for the bituminous porous surface PA-12 is presented in Fig. 3. This figure indicates that the variation in the coefficient B increases nonlinearly with frequency. This is one of the main results of the present study. The coefficient B is greater at frequencies above around 1 kHz, implying that at high frequencies, the sound from a bituminous porous surface increases with speed at a greater rate than sound at lower frequencies. However, in the high frequency range, above 1250 Hz, where sound generation is considered to be dominated by air pumping mechanisms, the coefficient B has

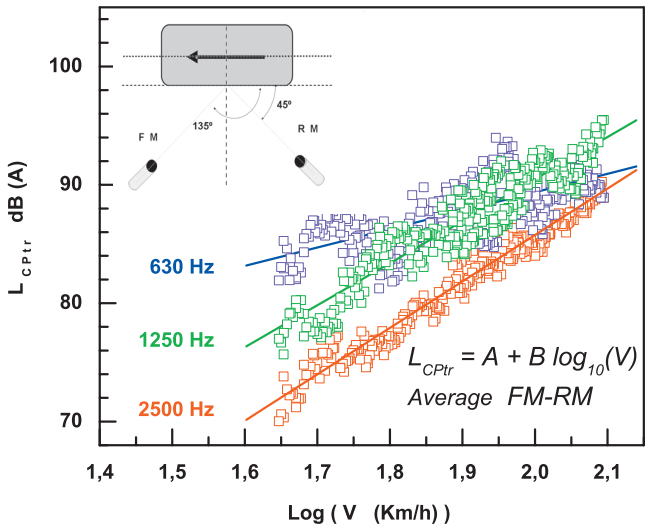


FIG. 2. Relationship between L_{CPtr} level and speed for different frequency bands, showing the slopes for the regressions (coefficients B). The levels are averaged values from two microphones, see inset.

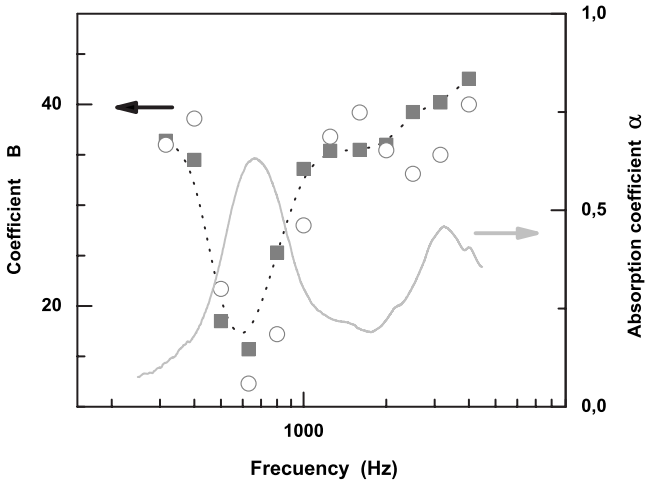


FIG. 3. Plot of coefficient B (sound-speed slope) vs frequency, B-coefficient spectrum. Gray line shows the averaged values of sound absorption from a number of core samples extracted from the road surface. (○) Result of another measurement (run).

an almost constant value. Between 500 and 1250 Hz, approximately, the coefficient B increases almost linearly with frequency.

It is important to note that this behavior characterizes a new porous asphalt. However, in order to establish a more general rule on the acoustical characterization of asphalt pavements, additional measurement on different asphalt mixes should be carried out.

Sound absorption spectra measured with an impedance tube for a number of core samples extracted from the test road showed an intense absorption band peaked at around 700 Hz, precisely where the spectrum of coefficient B presents a minimum (see Fig. 3). For this reason, the abrupt variation observed in the relationship between the coefficient B and speed at frequencies below 1 kHz could be influenced by the sound absorption mechanisms, particularly in our study for a nonclogged PA-12 surface with a high content of air voids (21%). To confirm that the values of coefficient B at low frequencies are correlated with sound absorption mechanisms, additional analysis of measurements on nonporous and/or dense pavement should be developed.

Figure 4 shows an x-y diagram of coefficient A versus coefficient B for each of the one third-octave bands studied (300 Hz–4 kHz). As is shown, the relationship between these coefficients is linear for a porous asphalt before its coming into service. This relationship could be used for asphalt pavement acoustical classifying purposes. This behavior of the coefficients A and B, for the frequencies studied, is considered the second major result of the present study.

In conclusion, it has been shown that a porous PA-12 pavement with a high air void content and before its coming into service could be acoustically characterized by different properties, first, via the coefficients A and B of the logarithmic regression $L_{CPtr}=A+B\ \log_{10}(V)$. In our case, for frequencies where the tire/pavement interaction dominates, the global sound increases linearly with the logarithm of speed, with $B \cong 30$ and $A \cong 40$. A detailed analysis of the sound evolution with speed (B) at the diverse frequency bands, represented graphically in the PA-12 sound/speed surface chart

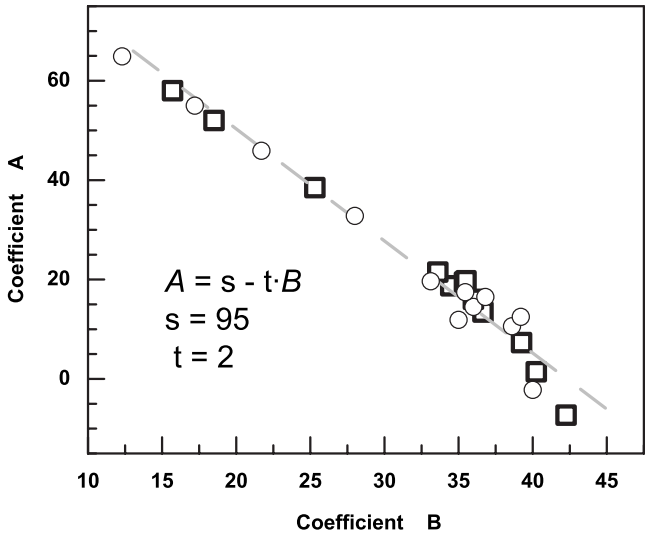


FIG. 4. Plot of the A and B coefficients (□) of each one-third octave band frequency (300 Hz–4 kHz), and the linear regression through all points (dashed lines). (○) Result of another measurement (run).

(see Fig. 5), shows the following: (1) at low frequencies (500–630 Hz), changes in L_{CPtr} with speed are not observed (the coefficient B is low), and it may be concluded that these frequencies are almost independent of speed; (2) at high frequencies (900 Hz–4 kHz), the variation in L_{CPtr} with speed is larger (the coefficient B is high); and (3) between these frequency regions, there is a linear increase in coefficient B. Figure 3 indicates that this behavior is correlated with the high sound absorption of the porous pavement studied at these frequencies. However, additional measurements with the same reference tire in different asphalt pavements are necessary to confirm this. On the other hand, a diagram of coefficients A and B in the range of frequencies studied shows a linear relationship between both. Further research is encouraged to confirm, for example, whether the linear relationship found between the coefficients A and B, $A=s-tB$ is an indication of how this surface would perform acoustically. For a new porous surface (PA-12), parameters $s=92$ and $t=2$ were found. Additional measurements on diverse pavements could serve to determine whether a surface is relatively quiet at low speeds for a low parameter t, and relatively noisy at high speeds, or if the best surface would be one with the lowest values of parameter s and t.

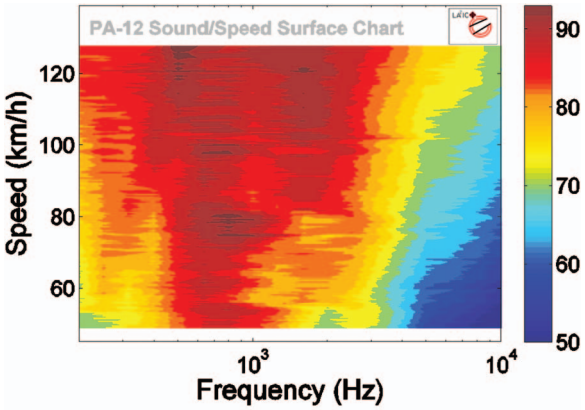


FIG. 5. The PA-12 sound/speed surface chart that represents graphically the evolution with speed of the L_{CPtr} as a function of frequency.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science, MEC, through Grant Nos. BIA2006-04783 and PAC07-0098-7433 of Junta de Comunidades de Castilla-La Mancha. We thank Juan Antonio Mesones, head of Unidad de Carreteras en Ciudad Real, General Directorate of Roads of the Spanish Ministry of Infrastructure and Transport for his valuable assistance.

[1]S. E. Paje, M. Bueno, F. Terán, U. Viñuela, and J. Luong, "Assessment of asphalt concrete acoustic performance in urban streets," J. Acoust. Soc. Am. **123**, 1439–1445 (2008).
[2]S. E. Paje, M. Bueno, F. Terán, and U. Viñuela, "Monitoring road surfaces by close proximity noise of the tire/road interaction," J. Acoust. Soc. Am. **122**, 2636–2641 (2007).
[3]U. Sandberg and J. A. Ejsmont, *Tyre/Road Noise Reference Book* (Informex, Harg, Kisa, Sweden, 2002).
[4]T. R. Wik and R. F. Miller, "Mechanisms of tire sound generation," Research Center, B. F. Goodrich Tire Co., SAE No. 720924, 2633–2643, 1972.

Feasibility of global-scale synthetic aperture communications (L)

H. C. Song[a] and M. Dzieciuch
Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0238

(Received 13 June 2008; revised 29 October 2008; accepted 30 October 2008)

A recent paper [Higley *et al.*, J. Acoust. Soc. Am. **118**, 2365–2372 (2005)] investigated synthetic aperture communications in shallow water exploiting the relative motion between a source and a receiver. This letter presents a feasibility study of synthetic aperture communications at global distances using the 57 Hz acoustic data from the Heard Island Feasibility Test conducted in January 1991 [Baggeroer and Munk, Phys. Today **45**, 22–30 (1992)]. Specifically, a reception at Ascension Island, about 9200 km from the source ship moving at about 3 kt near Heard Island, is analyzed by treating the 255-digit *m*-sequence tomography signal as a binary-phase shift-keying communication signal with an information rate of 11.4 bits/s. The performance using a single receiver combining three consecutive receptions spaced 5.2 min (~470 m) apart indicates that synthetic aperture acoustic communications is feasible at global distances.
[DOI: 10.1121/1.3035830]

PACS number(s): 43.60.Dh, 43.60.Gk, 43.60.Fg [DD] Pages: 8–10

I. INTRODUCTION

Beginning in the late 1960s, the U.S. Navy had an electromagnetic program (Project Sanguine/ELF) that involved using a huge Wisconsin-Michigan antenna system to communicate a minimal amount of information (e.g., three coded bits over 30 min) to submarines at speed and depth around the world oceans.[1] The project was based on the belief that such communications could not be accomplished acoustically, and eventually it was abandoned because of concerns over potential environmental and health problems. This letter demonstrates using archive data that indeed coherent acoustic communications at global scales is possible.

Synthetic aperture communications exploiting the relative motion between a source and a receiver has been investigated recently in shallow water.[2] Relative motion between the two generates a virtual horizontal array, which provides the spatial diversity required for multichannel time reversal combining. Two or three consecutive transmissions from a source moving at 2 kt over a 5 km range were combined after Doppler compensation, confirming the feasibility of synthetic aperture communications using time reversal.

This letter extends the concept of synthetic aperture communications to global scale by revisiting the low-frequency (57 Hz) acoustic data from the Heard Island Feasibility Test (HIFT) conducted in January 1991.[3] Specifically, a reception at Ascension Island is analyzed (Fig. 1). This high signal-to-noise ratio (SNR) reception (e.g., 19–30 dB after matched-filtering) traveled 9200 km from the source ship near Heard Island to bottom-mounted hydrophones located at the sound-channel axis depth around Ascension.[4,5] Among the three types of HIFT signals (cw, pentaline, and *m*-sequences), the 57 Hz, 255-digit *m*-sequence tomography signal is processed, with each digit consisting of five carrier cycles; thus it can be treated as a binary-phase shift-keying communication signal with an information rate of $R=57/5=11.4$ bits/s.

II. HIFT RECEPTION AT ASCENSION ISLAND

A detailed description of the HIFT can be found in Ref. 3; the main parameters relevant for discussion are briefly reviewed. Ascension Island is located in the southern Atlantic Ocean on the western edge of the Mid-Atlantic Ridge while Heard Island is in the southern Indian Ocean. Figure 1 shows the geodesic from the nominal position of the source ship R/V *Cory Chouest* near Heard Island (53°22′S, 74°30′E) to Ascension (7°57′S, 14°24′W).[4] The angles of the geodesic at the source ship and at Ascension are approximately 267° and 142°, respectively, measured relative to true North. The distance along the geodesic is approximately 9200 km, about one-quarter of the distance around the Earth.

The acoustic propagation turns out to be quite challenging since the ray-path goes through the dynamic Antarctic Circumpolar Current (ACC) front.[6] It is shown by simulations[7] that mode coupling at the ACC front has significant impact on modal dispersion and population, complicating the pulse-arrival sequence. Consequently there were no identifiable arrivals for tomography inversion. On the other hand, the subsequent tomography experiments (e.g., AET and ATOC) were carried out in the eastern North Pacific Ocean at the basin scale (e.g., 3 Mm) where the acoustic paths are found more stable and identifiable. Interestingly, basin-scale phase-coherent acoustic communications has been demonstrated, exploiting either spatial diversity (vertical array) or temporal diversity (ocean fluctuation) when both the transmitter and receivers were fixed.[8]

Acoustic transmissions (1 h long) were made with an acoustic-source array suspended from the research ship while underway into the wind and swell at about 3 kt (~1.5 m/s).

[a]Electronic mail: hcsong@ucsd.edu

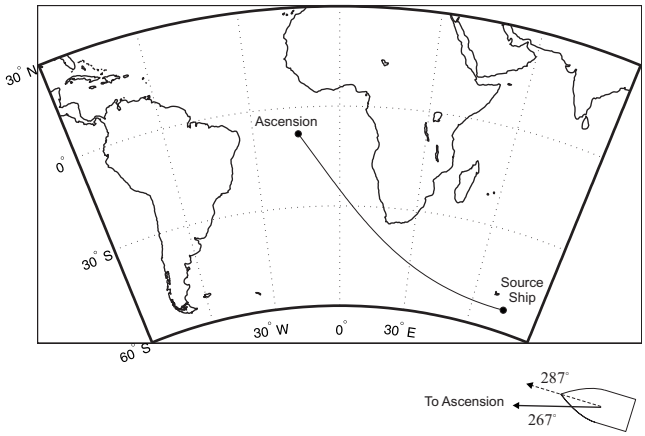


FIG. 1. Geodesic from the nominal position of the source ship near Heard Island to Ascension Island at approximately 9200 km range. The geodesic launch angles toward Ascension Island and the ship's course are 267° and 287°, respectively.

For a 57 Hz carrier signal, this yields a Doppler shift of $f_d = +0.057 \cos A$ Hz, where A is the angle between the azimuthal launch angle (267° toward Ascension Island) and the ship's course (287°), as shown in Fig. 1.[9] In practice, it is difficult to maintain a ship's speed and course during the entire 1 h transmission, resulting in a time-varying Doppler shift. Indeed the ship's GPS (global positioning system) positions measured at about 10 s intervals indicated time-varying offsets from a steady course on the order of 100 m in the direction of 267°, which was in good agreement with offsets inferred from the integrated Doppler measured at Ascension Island.[9] In this letter, a Doppler shift of $f_d = +0.057$ Hz (e.g., $\cos 20° \approx 1$) is chosen for Doppler compensation (resampling) in preprocessing, while the residual Doppler shift is compensated by a phase tracking algorithm after time reversal diversity combining.[10] Note that synthetic aperture communications normally requires estimation of a Doppler shift using either a separate signaling scheme (e.g., pilot tone) or initial phase tracking whose slope corresponds to a Doppler shift.

All transmissions from the HIFT were digitally recorded on eight bottom-mounted hydrophones at Ascension, most of them located near the depth of the sound-channel axis, about 800 m. We analyze a portion of the data (15 min long) received at one of the primary hydrophones (denoted hydrophone 23 in Refs. 4 and 5) from an acoustic run that started at 0600 UTC 26 January 1991. The travel time of the earliest arrival is about 1 h, 44 min, 17 s, and the delay spread at this range is about 12 s long (Fig. 2), resulting in the intersymbol interference spanning about 140 symbols at a symbol rate of 11.4 symbols/s. Coarse synchronization is accomplished by correlating with one period of the 255-digit m-sequence (i.e., 22.4 s). For communications processing, the Doppler-compensated data are complex demodulated to baseband and then sampled at twice the symbol rate (i.e., $2 \times 11.4 = 22.8$ Hz), facilitating use of a fractionally sampled decision-feedback equalizer (DFE) after time reversal combining.[10]

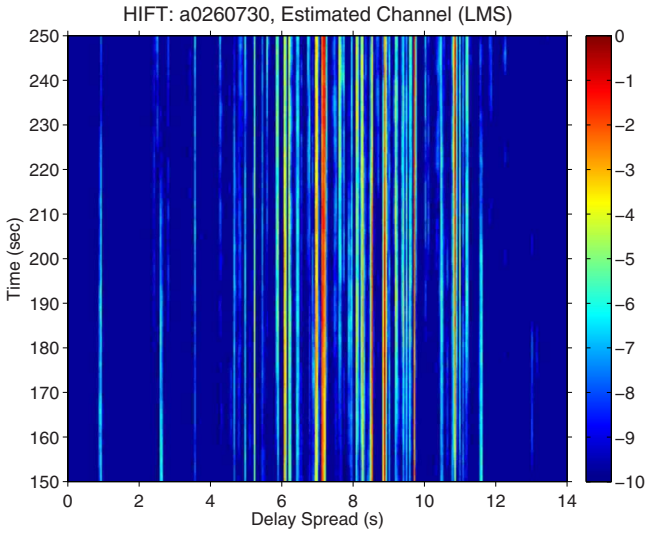


FIG. 2. (Color online) Temporal evolution of the channel response. The delay spread is about 12 s, resulting in the ISI spanning about 140 symbols at a symbol rate of 11.4 symbols/s.

III. SYNTHETIC APERTURE COMMUNICATIONS

Consider three blocks of signals separated by 5.2 min corresponding to 14 periods of the 255-digit m-sequence, as illustrated in Fig. 3. The corresponding travel distance of a source moving radially at 3 kt is about 470 m (~18 wavelengths). Here the same communication sequence is presumed to be transmitted periodically (5.2 min) at the cost of a reduction in the overall data rate by the number of transmissions (i.e., $M=3$) accumulated in generation of the synthetic aperture.[2] The three blocks of communication sequences are coherently combined by the time reversal approach followed by a single channel adaptive DFE.[10] A decision-directed carrier phase estimate based on the maximum likelihood method[11] is employed prior to the DFE.

Time reversal requires knowledge of the channel between a source and a receiver, usually provided by a channel probe at the beginning of the data packet. Since there is no explicit probe signal in the HIFT tomography signal, each block consists of five periods of training symbols ($N_T=5 \times 255$) for channel estimation and two periods of information symbols ($N=2 \times 255$) to be demodulated. The least mean square (LMS) algorithm is employed for channel estimation (see Fig. 2), whereas the recursive least squares (RLS) algorithm is employed for the single channel DFE after time reversal combining. The corresponding LMS step size and the RLS forgetting factor are $\Delta=0.001$ and λ

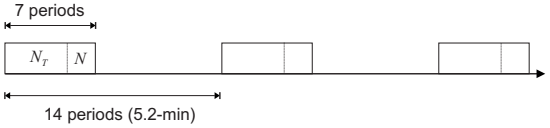


FIG. 3. Data structure emulating synthetic aperture communications. Each block consists of five periods of training symbols and two periods of information symbols. The blocks are separated by 14 periods of the 255-digit m-sequence (5.2 min).

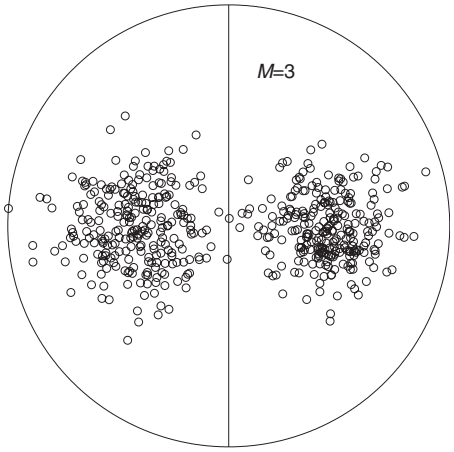


FIG. 4. Performance of synthetic aperture time reversal communications combining three consecutive receptions spaced 5.2 min (~470 m) apart after Doppler compensation (see Fig. 3). The output SNR is 6.6 dB with a bit error rate of 1/510.

=0.999, respectively. The number of fractionally spaced feedforward equalizer taps is 120 (=2×60), and the number of feedback equalizer taps is set to 80.

The resulting performance is shown in Fig. 4 as a scatter plot. The output SNR is 6.6 dB, and the bit error rate is 1/510. The performance using a single receiver combining three consecutive receptions spaced 5.2 min (~470 m) indicates that synthetic aperture communications is feasible at global distance using time reversal.

[1]L. Klessig and V. Strite, *The ELF Odyssey: National Security Versus Environmental Protection* (Westview, Boulder, CO, 1980).
[2]W. J. Higley, P. Roux, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, and T. Akal, "Synthetic aperture time-reversal communications in shallow water: Experimental demonstration at sea," J. Acoust. Soc. Am. **118**, 2365–2372 (2005).
[3]A. Baggeroer and W. Munk, "The Heard Island Feasibility Test," Phys. Today **45**, 22–30 (1992).
[4]D. Palmer, T. Georges, J. Wilson, L. Weiner, J. Paisley, R. Mathiesen, R. Pleshek, and R. Mabe, "Reception at Ascension of the Heard Island Feasibility Test transmissions," J. Acoust. Soc. Am. **96**, 2432–2440 (1994).
[5]T. Georges, L. R. Boden, and D. R. Palmer, "Features of the Heard Island signals received at Ascension," J. Acoust. Soc. Am. **96**, 2441–2447 (1994).
[6]G. L. Pickard and W. J. Emery, *Descriptive Physical Oceanography: An Introduction* (Butterworth-Heinemann, Oxford, 1990).
[7]E. C. Shang and Y. Y. Wang, "Dispersion and repopulation of Heard-Ascension modes," J. Acoust. Soc. Am. **96**, 2371–2379 (1994).
[8]H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "Basin-scale time reversal communications," J. Acoust. Soc. Am. **125**, 212–217 (2009).
[9]M. Dzieciuch, W. Munk, and A. Forbes, "Interpretation of GPS offsets from a steady course," J. Atmos. Ocean. Technol. **9**, 862–866 (1992).
[10]H. C. Song, W. S. Hodgkiss, W. A. Kuperman, W. Higley, K. Raghukumar, T. Akal, and M. Stevenson, "Spatial diversity in passive time reversal communications," J. Acoust. Soc. Am. **120**, 2067–2076 (2006).
[11]J. Proakis, *Digital Communications* (McGraw-Hill, New York, 2001).

Comment on "Measuring power production in the mammalian cochlea" [Curr. Biol. 17, 1340 (2007)] (L)

Fangyi Chen
Oregon Hearing Research Center, NRCO4, Department of Otolaryngology/Head & Neck Surgery, Oregon Health Sciences University, 3181 S.W. Sam Jackson Park Road, Portland, Oregon 97239

Alfred L. Nuttall[a)]
Oregon Hearing Research Center, NRCO4, Department of Otolaryngology/Head & Neck Surgery, Oregon Health Sciences University, 3181 S.W. Sam Jackson Park Road, Portland, Oregon 97239; Kresge Hearing Research Institute, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109-0506; Department of Otolaryngology, Renji Hospital, Shanghai Jiao Tong University, Shanghai, China 200001

(Received 13 March 2008; revised 30 May 2008; accepted 30 May 2008)

Recently, a paper by Lakashkin *et al.* (2007) ("Power amplification in the mammalian cochlea," Curr. Biol. **17**, 1340–1344) was published on how power can be measured in the mammalian cochlea. The general subject is of current widespread interest, so the question of whether the method used by Lakashkin *et al.* is valid may be of interest to the readers of this journal. Power generation in the cochlea can account for the extraordinary sensitivity of hearing. Lukashkin *et al.* claimed to provide a direct proof of cochlear power generation. A first-order spring-dashpot system was used to model the organ of Corti. The power flux direction can be derived from the sign of the phase difference between the force and displacement, which can be presented as a "hysteresis plot." Basilar membrane (BM) vibration near the characteristic frequency (CF) was measured while applying a low-frequency modulation tone together with the CF tone. A force was derived from the modulation profile of the BM CF vibration and when plotted versus the displacement at the modulation frequency, the function had a counterclockwise direction of hysteresis, suggesting power generation. In this letter, we present comments on the analysis in the report: (1) that it is not appropriate to analyze at the modulation frequency to derive the power generation at CF; (2) that the derivation of a force from just the displacement profile is not justified, followed by an alternative interpretation of the experimental data.
 [DOI: 10.1121/1.2950090]

PACS number(s): 43.64.Bt, 43.64.Kc, 43.64.Ld, 43.64.Yp [BLM] Pages: 11–14

I. INTRODUCTION

The cochlea possesses a cellular mechanism to enhance sound induced motion of the organ of Corti that is termed the "cochlear amplifier" (CA). The effect of cochlear amplification on auditory performance as well as its cellular basis has been documented by many studies (LeMasurier and Gillespie, 2005; Dallos *et al.*, 2006; de Boer, 2006). For example, it is known that the vibration amplitude of basilar membrance (BM) oscillation is enhanced by a factor of 100–1000 for frequencies matched to the spatial location in the cochlea where the frequencies have their best responses (Robles *et al.*, 1986; Ruggero *et al.*, 1997).

The cellular mechanism of amplification is based on the electromechanical activity of the outer hair cell which is capable of producing force by one or more molecular mechanisms involving specialized proteins (Martin *et al.*, 2000; Zheng *et al.*, 2000; Ricci *et al.*, 2002). It has been hypothesized that the force serves the purpose of countering energy loss from the viscous component of what is primarily a viscoelastic system. Thus the outer hair cells are hypothesized to produce power and it has been argued that power is the more appropriate variable to measure in order to characterize the CA (de Boer and Nuttall, 2001).

Testing the hypothesis of power generation by the organ of Corti is critically important but represents a vexing experimental problem. The direct experimental approach requires measurement of both the local acoustic force or pressure and the local velocity. To measure acoustical pressure inside the cochlea is exceedingly difficult and thus the data are confounded by large margins of error. Olson (2001) drew the tentative conclusion from such measurements that power is being produced by the organ of Corti. In an indirect approach, the power going into and outward from the cochlea as otoacoustic emissions has been estimated (Kemp, 1979). Another alternative is to infer the pressure from comparing mechanical data to models of cochlear mechanics. With the inverse solution method, it is possible to compute the BM's mechanical impedance from the measured velocity of the BM (de Boer and Nuttall, 1999). The method generates an impedance function of which the real part has negative values in the region basalward from the response peak. Since

[a)]Author to whom correspondence should be addressed. Electronic mail: nutall@ohsu.edu

the real part of an impedance is related to power (growth or decay), the negative values indicate generation of acoustical power by the BM and its associated structures.

II. INTRODUCTION TO THE REPORT PAPER

Lukashkin *et al.* (2007) (termed the "report" below) take another indirect approach using the modulation profile of the BM vibration to derive the active force on the BM. This force was then plotted versus the displacement to infer the power generation/consumption. In the report, a simple spring-dashpot model was first used to introduce the hysteresis idea. The phase difference between the driving force and displacement (or the force on the spring) results in a graphical pattern (a Lissajous figure) that is open (has area) when the displacement is plotted against the force over one vibration cycle. This open pattern is termed a hysteresis and the direction around the hysteresis loop can infer the direction of the power flux. In the experimental design, a low-frequency (LF) high-intensity tone was applied together with a high-frequency (HF) low-intensity tone. The LF tone causes change of the operating point, and thus the gain, of the CA. This gain change results in a modulation on the HF vibration. After demonstrating the sensitivity of the preparation in their Fig. 2, Lukashkin *et al.* (2007) presented a method of deriving the driving force from the modulation profile. Their Fig. 3A showed the measured HF and LF displacements. The HF displacement was modulated by the LF one, showing two peaks in its amplitude profile within one LF cycle. These two peaks correspond to the condition when optimal (maximum gain) operating position was approached twice: one in the rising edge of the LF vibration, and one in the falling edge. This is also plotted in their Fig. 3B. (Note that in Fig. 3B, the amplitude is plotted versus the modulation so that the *X*-axis is labeled as "Low-frequency bias.") In their Fig. 3B, the peak at the rising portion of the LF displacement (solid line) is at a positive displacement; the peak at the falling portion of the LF displacement (dash line) is at a negative displacement. Correspondingly in the report's Fig. 3A, these two peaks are at time instances later than the two zero crossings of the LF displacement. The amplitude of the HF BM vibration versus LF BM vibration was then fitted with the derivative of a second-order Boltzmann function. The dependent variable of the fitted Boltzmann function was presented in the report as the nonlinear force exerted on the organ of Corti. The graph obtained when this nonlinear force versus the LF BM displacement is plotted has the open form of a hysteresis. The counterclockwise rotation of the graphical path around this hysteresis was promulgated as a direct proof that power is generated by the organ of Corti.

III. THE HYSTERESIS PLOT

Since the hysteresis plot was used in the report to evaluate the direction of the power flux, here we present explanations in order to clarify the condition in which the hysteresis plot can be linked to the power generation. While plotting the two sinusoidal signals of the same frequency versus each other, phase difference between two signals will result in a Lissajous figure (a hysteresis plot). The hysteresis has clockwise direction if the signal at the *Y*-axis lags the one at the *X*-axis and vice versa. This was shown in Figs. 1C and 1D of the report. In general, a hysteresis plot between two signals in a system reflects only the time/phase difference between them. Furthermore, because the signals are sinusoidal, a phase lag more than 180° but less than 360° will be treated as a phase lead due to the periodic nature of the signal.

The direction of the hysteresis plot can be related to the direction of the power flux only when the two signals are the true driving force and the complete motion (or voltage and current in electrical system). This is essentially the same concept as measuring the driving point impedance of a system. The difference is that in the impedance measurement, velocity is used but displacement is used in the power (hysteresis) method. For a system that can be simplified into a driving point impedance, such as the spring-dashpot system in Fig. 1 of the report, the phase difference between the force applied to and the total velocity/displacement of the system is determined by the mechanical impedance of the system. There is a strict definition of "active" for a one port system in network theory (Guillemin, 1951; Raisbeck, 1954), but in effect a negative real value of the impedance indicates that the power is produced.

Two points are worth mentioning while applying the hysteresis method to study the organ of Corti. *First*, the power analyzed is that of the cycle frequency of analysis and thus when using the hysteresis method for the cochlea the frequency should be around the CF of the measured location. Whether this frequency should be at CF or at a frequency higher than CF (since it is commonly believed that power is generated at places basal to the CF location) is arguable, the frequency should not be at a much lower frequency—the modulation frequency. For a linear time-invariant system, where the impedance concept was developed, the real part of the impedance, and thus direction of the power flux, is usually frequency independent. Therefore, the hysteresis plotted at any frequency can be used to identify the power flux. However, the organ of Corti is a highly nonlinear system. It has been shown that the BM vibration is not amplified (Cooper and Rhode, 1997; Ruggero *et al.*, 1997) at frequencies outside the peaking region. The power studies (Allen and Fahey, 1992; de Boer *et al.*, 2005) suggested that the possible power generation, if there is such, should happen within the frequency range of half an octave from the CF. Therefore, the hysteresis plot should be generated using the driving force and the displacement at or above the CF. *Second*, there must be a correct force and displacement estimate. As mentioned above, in the impedance (and also the power hysteresis) method, the organ of Corti is simplified into a single port system. The motion of the organ is taken as BM displacement/velocity. Ideally one can measure the pressure difference between scala vestibuli and scala tympani as the driving force. Other methods of estimating the force need serious justification.

IV. THE PROBLEMS IN THE REPORT

(1) *The frequency used to plot the hysteresis curve is at the modulation frequency*. In the report, the hysteresis was

not generated from a vibration cycle at the CF but at the modulation frequency of 28 Hz. As mentioned in Sec. III, to study power generation, one would plot the curve over a cycle of the CF vibration and study the amount of area in the hysteresis loop and the direction of the path. However, in the report, the conclusion about power is from the hysteresis of the HF modulation profile versus LF displacement. This hysteresis is essentially obtained at the modulation frequency rather than at the CF. Although the *X*-axis of their Fig. 3C is labeled as "High-frequency displacement" and the report states "The axis in Figure 3C represents BM displacement at the CF," the derivation of Fig. 3C does suggest that it should be LF. In the report, Fig. 3C was derived from the integration of a fitted result of Fig. 3B. The *X*-axis in Fig. 3B is "Low-frequency displacement" and the integration operation cannot change the *X*-axis label.

(2) *The derivation of the force is not valid*. In the report, the mechanical driving force was derived from the modulation profile of the displacement. In the experiment, only the BM displacement was measured. It is intuitively difficult to imagine that the force can be derived from the displacement only. Actually, from the derivation shown in the *experimental procedures* section of the report, Lukashkin *et al.* (2007) seemed to believe that the modulation of the HF vibration can only be influenced by the mechanical force on the BM, so that they derived the driving force of the BM from it. However, other cellular biochemical processes can also influence the modulation of the BM vibration, especially the timing, which determines the direction of the hysteresis. Since the so-called force is not the actual mechanical force, the hysteresis plot of this so-called force versus the BM displacement cannot be related to the power generation.

V. WHAT COCHLEAR PROCESSES CAN PRODUCE THE HYSTERESIS?

Since hysteresis is caused by a phase/time difference, one possible source is the delay of the operating point changes, in response to the LF BM displacement. As indicated in Sec. II, the modulation peaks delayed the correspondent LF displacement zero crossings. Assuming that the operating point of the cochlea is optimized at rest, which corresponds to the zero crossing of the LF displacement, the fact that the peaks of modulation profile delayed from the zero crossing indicates that there is a delay from the stimulus (LF displacement) to the system response (amplitude modulation). It is this delay that produces the hysteresis loop, which was used by Lukashkin *et al.* (2007) to derive the power generation conclusion. However, this kind of delay is common for any realizable system, active or passive. This delay is also not unexpected inside the organ of Corti, considering that the translation of the LF BM displacement into change of the operating point of the system could involve complex mechanical, electrical, and chemical processes. All these processes will introduce delay between the LF BM vibration and the modulation profile, and this delay can produce the hysteresis, as shown in the report. For example, the relative motion (Mammano and Ashmore, 1993; Fridberger and de Monvel, 2003; Scherer and Gummer, 2004; Nowotny and Gummer, 2006; Tomo *et al.*, 2007) between different parts of the organ of Corti can introduce delays; the adjustment of the mechanoelectric transduction channel operating point can have an adaptation time constant from a few milliseconds to tens of milliseconds (LeMasurier and Gillespie, 2005; Fettiplace, 2006; Tinevez *et al.*, 2007). All those delays could add up to a significant amount of time. In the literature, a similar hysteresis analysis was used by Bian *et al.* (2004) where a LF tone modulated distortion product otoacoustic emissions. Although Bian *et al.* (2004) presented a model relating the power generation to the hysteresis [Sec. IV A in Bian *et al.* (2004)], they attributed the counterclockwise hysteresis to the delay mostly due to the complex bidirectional processes inside the organ of Corti [Sec. V A in Bian *et al.* (2004)]. Such a delay can also account for hysteresis in the Lukashkin *et al.* (2007) analysis.

VI. SUMMARY

In this letter, we discuss a recent paper [Lukashkin *et al.* (2007)] about the power generation in the organ of Corti. We pointed out that their conclusion about cochlear power generation is not valid because the hysteresis plot of the force versus BM displacement is at the modulation frequency and cannot be used to indicate the power generation at CF. We also questioned the force derivative. An alternative interpretation was provided to account for the experimental observation in Lukashkin *et al.* (2007).

ACKNOWLEDGMENTS

The authors wish to acknowledge Professors Egbert de Boer (University of Amsterdam) and Karl Grosh (University of Michigan) for extensive discussions on power analysis. The authors would also like to thank Dr. David Mountain and Dr. Allyn Hubbard (Boston University) for their contributions.

Allen, J. B., and Fahey, P. F. (**1992**). "Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane," J. Acoust. Soc. Am. **92**, 178–188.

Bian, L., Linhardt, E. E., and Chertoff, M. E. (**2004**). "Cochlear hysteresis: Observation with low-frequency modulated distortion product otoacoustic emissions," J. Acoust. Soc. Am. **115**, 2159–2172.

Cooper, N. P., and Rhode, W. S. (**1997**). "Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea," J. Neurophysiol. **78**, 261–270.

Dallos, P., Zheng, J., and Cheatham, M. A. (**2006**). "Prestin and the cochlear amplifier," J. Physiol. (London) **576**, 37–42.

de Boer, E. (**2006**). "Cochlear activity in perspective.," in *Auditory Mechanisms: Process and Models*, edited by A. L. Nuttall, T. Y. Ren, P. G. Gillespie, K. Grosh, and E. de Boer (World Scientific, Singapore), pp. 393–409.

de Boer, E., and Nuttall, A. L. (**1999**). "The "inverse problem" solved for a three-dimensional model of the cochlea. III. Brushing up the solution method," J. Acoust. Soc. Am. **105**, 3410–3420.

de Boer, E., and Nuttall, A. L. (**2001**). "Power gain of the "Cochlear Amplifier"," in *Physiological and Psychological Bases of Auditory Function: Proceedings of the International Symposium on Hearing*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht, NL), pp. 1–7.

de Boer, E., Nuttall, A. L., Hu, N., Zou, Y., and Zheng, J. F. (**2005**). "The Allen–Fahey experiment extended," J. Acoust. Soc. Am. **117**, 1260–1266.

Fettiplace, R. (**2006**). "Active hair bundle movements in auditory hair cells," J. Physiol. (London) **576**, 29–36.

Fridberger, A., and de Monvel, J. B. (**2003**). "Sound-induced differential motion within the hearing organ," Nat. Neurosci. **6**, 446–448.

Guillemin, E. A. (**1951**). "A summary of modern methods of network analysis," in *Advances in Electronics* (Academic, New York), pp. 261–303.
Kemp, D. T. (**1979**). "The evoked cochlear mechanical response and the auditory microstructure-evidence for a new element in cochlear mechanics," Scand. Audiol. Suppl. **9**, 35–47.
LeMasurier, M., and Gillespie, P. G. (**2005**). "Hair-cell mechanotransduction and cochlear amplification," Neuron **48**, 403–415.
Lukashkin, A. N., Walling, M. N., and Russell, I. J. (**2007**). "Power amplification in the mammalian cochlea," Curr. Biol. **17**, 1340–1344.
Mammano, F., and Ashmore, J. F. (**1993**). "Reverse transduction measured in the isolated cochlea by laser Michelson interferometry," Nature (London) **365**, 838–841.
Martin, P., Mehta, A. D., and Hudspeth, A. J. (**2000**). "Negative hair-bundle stiffness betrays a mechanism for mechanical amplification by the hair cell," Proc. Natl. Acad. Sci. U.S.A. **97**, 12026–12031.
Nowotny, M., and Gummer, A. W. (**2006**). "Nanomechanics of the subtectorial space caused by electromechanics of cochlear outer hair cells," Proc. Natl. Acad. Sci. U.S.A. **103**, 2120–2125.
Olson, E. S. (**2001**). "Intracochlear pressure measurements related to cochlear tuning," J. Acoust. Soc. Am. **110**, 349–367.
Raisbeck, G. (**1954**). "A definition of passive linear networks in terms of time and energy," J. Appl. Phys. **25**, 1510–1514.
Ricci, A. J., Crawford, A. C., and Fettiplace, R. (**2002**). "Mechanisms of active hair bundle motion in auditory hair cells," J. Neurosci. **22**, 44–52.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). "Basilar membrane mechanics at the base of the chinchilla cochlea. 1. Input-output functions, tuning curves, and response phases," J. Acoust. Soc. Am. **80**, 1364–1374.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.
Scherer, M. P., and Gummer, A. W. (**2004**). "Vibration pattern of the organ of Corti up to 50 kHz: evidence for resonant electromechanical force," Proc. Natl. Acad. Sci. U.S.A. **101**, 17652–17657.
Tinevez, J.-Y., Julicher, F., and Martin, P. (**2007**). "Unifying the various incarnations of active hair-bundle motility by the vertebrate hair cell," Biophys. J. **93**, 4053–4067.
Tomo, I., Boutet de Monvel, J., and Fridberger, A. (**2007**). "Sound-evoked radial strain in the hearing organ," Biophys. J. **93**, 3279–3284.
Zheng, J., Shen, W., He, D. Z., Long, K. B., Madison, L. D., and Dallos, P. (**2000**). "Prestin is the motor protein of cochlear outer hair cells," Nature (London) **405**, 149–155.

Reduced contribution of a nonsimultaneous mistuned harmonic to residue pitch: The role of harmonic number (L)

Hedwig E. Gockel[a] and Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Christopher J. Plack
School of Psychological Sciences, University of Manchester, Manchester M13 9PL, United Kingdom

(Received 15 May 2008; revised 22 August 2008; accepted 23 October 2008)

Listeners adjusted a matching sound so that its pitch equaled that of a subsequent 90-ms complex tone (12 harmonics of a 155-Hz *F*0), whose mistuned harmonic (±3%) was presented either simultaneously with or after the other harmonics. The third or the fourth harmonic was mistuned. For both harmonics, pitch shifts were significantly smaller when the mistuned harmonic was presented after rather than simultaneously with the remainder of the target complex and when the component corresponding to the mistuned harmonic was absent (as opposed to present) from the matching sound. Presenting the mistuned component after the remainder of the complex generally reduced the pitch shifts more for the third than for the fourth harmonic; the nonsimultaneous presentation reduced shifts by a factor of 2.8 (third) and 1.7 (fourth) in the presence, and 6.4 (third) and 1.9 (fourth) in the absence of the nominally mistuned harmonic from the matching sound. The results are discussed in terms of an analytical listening and matching strategy that results in larger pitch shifts for a nonsimultaneous mistuned harmonic that has an octave or double-octave relationship to the *F*0 than for one that does not, without it necessarily being temporally integrated into the residue pitch of the complex.
 [DOI: 10.1121/1.3026327]

PACS number(s): 43.66.Hg, 43.66.Mk, 43.66.Ba [RYL] Pages: 15–18

I. INTRODUCTION

Several researchers have explored the duration over which information is integrated in the determination of the residue pitch of a complex tone (Houtsma and Goldstein, 1972; Hall and Peters, 1981; Houtsma, 1984; Ciocca and Darwin, 1999; Grose *et al.*, 2002; Gockel *et al.*, 2005b). This question has implications for theories of pitch perception which specify the time constant over which pitch is calculated (Meddis and Hewitt, 1991a, 1991b; Wiegrebe, 2001).

Some of the studies addressing the question of the integration time for residue pitch exploited the fact that shifting the frequency of a single harmonic in an otherwise harmonic complex tone can lead to a shift in the residue pitch of the complex even when the amount of the mistuning is so large that the mistuned harmonic is heard out as a separate tone (Moore *et al.*, 1985, 1986). The largest shifts in residue pitch have been found when the frequency of the mistuned harmonic is shifted by about 3% from its harmonic value (Moore *et al.*, 1985; Darwin *et al.*, 1992). Ciocca and Darwin (1999) compared the size of the shifts in residue pitch obtained for a mistuned component presented simultaneously with the remainder of the complex with that obtained when the mistuned component was presented immediately after the remainder of the complex. The 90-ms harmonic complex had a fundamental frequency (*F*0) of 155 Hz and its fourth harmonic was mistuned by ±3%. When the mistuned component was presented after the remainder of the complex, it could clearly be heard as a separate tone "popping out," in addition to the richer percept evoked by the complex. Subjects were instructed to ignore this separate higher tone. In experiment 1, Ciocca and Darwin (1999) found no significant difference between the size of the pitch shifts observed with simultaneous presentation of the mistuned harmonic (mean shift of 0.7 Hz) and with post-remainder presentation (mean shift of 0.5 Hz). In experiment 2, they investigated the effect of silent gaps of various durations between the end of the remainder of the complex and the onset of the mistuned harmonic. The mean pitch shifts for simultaneous and post-target presentation with a 0-ms gap were virtually identical (0.52 and 0.49 Hz). The shifts for the 40- and 80-ms gap conditions were not significantly different from that for the 0-ms gap, but for the 160-ms gap the pitch shift was reduced to zero. Based on their results, Ciocca and Darwin (1999) estimated the integration time for pitch to be around 170–250 ms.

In contrast, Gockel *et al.* (2005b) recently observed that pitch shifts *were* significantly reduced when the mistuned harmonic was presented after the remainder of the complex. In addition, pitch shifts were reduced when the matching complex (the complex adjusted in *F*0 by the subjects) did not contain a component corresponding to the mistuned harmonic. However, Gockel *et al.* (2005b) mistuned the third rather than the fourth harmonic. They discussed two, not mutually exclusive, explanations for the discrepancy between their results and those of Ciocca and Darwin (1999). First, the fourth harmonic might have a special status due to

[a]Author to whom correspondence should be addressed. Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk

its double-octave relationship to the *F*0. For example, Houtsma (1979) reported that an isolated partial with an octave relationship to the fundamental, i.e., with a frequency two, four, or eight times the *F*0, has the same pitch chroma (see Bachem, 1950) as the harmonic complex. This identity of pitch chromas could lead to the fourth harmonic being more integrated into the pitch of the target complex than the third harmonic, even when slightly mistuned and presented after the remainder of the target. Second, rather than basing their judgements on a residue pitch which incorporates the mistuned component, subjects might have used a more analytical listening and matching strategy. This analytical strategy would involve: (a) comparing and trying to minimize the difference between the pitch of the mistuned component that pops out and either the pitch of the corresponding harmonic in the matching sound or the residue pitch of the matching sound, and, also (b) comparing and trying to minimize the difference between the residue pitch of the remainder of the target complex (without the mistuned harmonic) and that of the matching sound. One would have to assume that subjects do both (a) and (b) in order to explain why the observed pitch shifts first increase with increasing mistuning and, with further mistuning beyond about 3%, decrease again (Moore *et al.*, 1985; Darwin, 1992). The finding that pitch shifts were smaller in the absence of the third harmonic from the matching sound than in its presence lends some support to the idea that subjects might have used such an analytical listening and matching strategy. Note that: (1) such a strategy is more likely to be used in nonsimultaneous than in simultaneous conditions because mistuned components are perceptually segregated to a much higher degree in the former than in the latter case, and (2) the use of such a strategy would lead to larger pitch shifts for the fourth than for the third harmonic because the pitch chroma difference between the mistuned harmonic and the remainder of the target complex is smaller for the fourth than for the third harmonic.

The objective of the present study was to investigate the role of the number of the mistuned harmonic in the contribution of a nonsimultaneous harmonic to residue pitch. Specifically, the question was whether the difference in the conclusions reached by Ciocca and Darwin (1999) and by Gockel *et al.* (2005b) is due to effects specific to the fourth harmonic. For that purpose, in the present study the same subjects were tested with either the third or the fourth harmonic mistuned, and this harmonic could be presented either simultaneously with or after the remainder of the target complex. Additionally, the importance of the presence of the corresponding harmonic in the adjustable matching complex was assessed.

II. METHOD

The stimuli were similar to those used by Ciocca and Darwin (1999) and identical to those used by Gockel *et al.* (2005b), except that either the third harmonic or the fourth harmonic was mistuned by ±3%. The target was a 90-ms complex tone with an *F*0 of 155 Hz, containing the first 12 harmonics added in sine phase, each with equal amplitude [58 dB sound pressure level (SPL)]. The mistuned harmonic was presented either simultaneously with, or immediately after, the remainder of the target complex ("post-target condition"). The matching sound was a 90-ms harmonic complex, nominally containing the first 12 harmonics added in sine phase. Depending on the condition, the third or the fourth harmonic was either present in or absent from the matching sound.

The method was the same as that described by Gockel *et al.* (2005b). A two-interval, two-alternative adaptive procedure was used to obtain pitch matches between a matching sound (whose *F*0 was adjusted, and which was always presented first) and the target complex. Subjects were explicitly instructed to ignore any pure-tone percept in the interval containing the target sound and to match the pitch of the matching complex to that of the target complex.

Eight musically trained subjects participated in all conditions. Their ages ranged from 18 to 34 years, and their quiet thresholds at octave frequencies between 250 and 4000 Hz were within 15 dB of the ANSI (2004) standard. Stimuli were presented monaurally to the left ear for seven subjects and to the right ear for one subject. After about 2–3 h of practice, five subjects were tested first with the third harmonic mistuned and then with the fourth; the order was reversed for the other subjects. The five subjects who were tested first with the third harmonic also participated in part of Gockel *et al.*'s (2005b) study.

III. RESULTS AND DISCUSSION

The number of pitch matches collected was varied across subjects in order to compensate to a certain degree (depending on the availability of the subject) for differences in the variability of the matches across subjects. At least ten (usually 15) pitch matches were obtained for each subject and condition (and for each direction of mistuning) resulting in a typical standard error of about 0.06–0.07 Hz. The results are presented as the mean *absolute* pitch shift, calculated as half the difference between pitch matches for corresponding positive and negative mistunings.

Figure 1 shows the geometric means and the standard errors of the pitch shifts across subjects. For the third harmonic mistuned [Fig. 1(a)], the results are very similar to those reported for experiment 1A of Gockel *et al.* (2005b). This is not unexpected, as data for five of the eight subjects here contributed to the data for eight subjects shown in Fig. 2 of Gockel *et al.* (2005b). In the absence of the third harmonic from the matching sound, the present results are also very similar to those reported for experiment 2 of Gockel *et al.* (2005b), where either the matching sound was a sine tone (experiment 2B) or the target complex and the matching sound contained only nonoverlapping harmonics (experiment 2A). As discussed by Gockel *et al.* (2005b), the similarity between results obtained in experiment 2 and those obtained with a matching sound containing harmonics 1–2 and 4–12 indicates that the results with the latter (and by extension the present results in the absence of the nominally mistuned harmonic from the matching sound) were not due to *perceptual segregation* of the mistuned component from the target complex caused by the absence of that single specific component

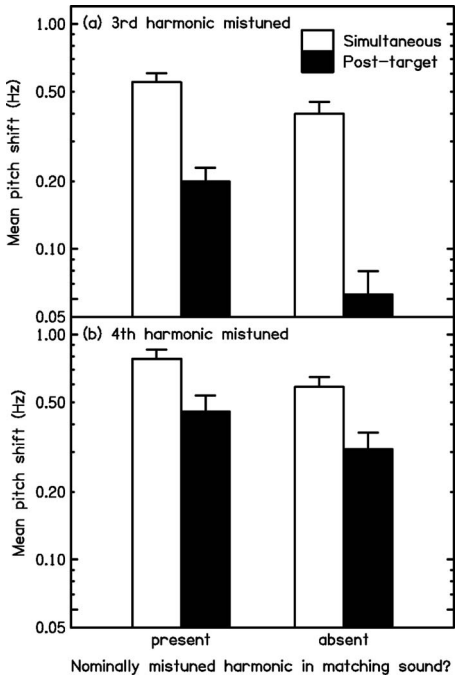


FIG. 1. Mean pitch shifts (log scale) and the corresponding standard errors across eight subjects. Panel (a) shows the data for the third harmonic mistuned. The left-hand group of two bars shows pitch shifts obtained for a matching complex that included the third harmonic. The right-hand group of two bars shows the pitch shifts obtained when the third harmonic was absent from the matching complex. The white and the black bars show results for simultaneous and post-target presentation of the mistuned component, respectively. Panel (b) as (a), but for the fourth harmonic mistuned.

in the matching sound. Rather the absence of the nominally mistuned harmonic from the matching sound *per se* seems crucial.

The mean pitch shifts were generally larger for the fourth [Fig. 1(b)] than for the third harmonic. When the fourth harmonic was present in the matching sound, the size of the pitch shift in the simultaneous condition (0.78 Hz) was bigger than but within the range of the standard error of the mean reported by Ciocca and Darwin (1999) for this condition in their experiment 1. The post-target condition (with the fourth harmonic present in the matching sound) led to a mean pitch shift of about 0.46 Hz, which corresponds well with Ciocca and Darwin's (1999) results.

A repeated-measures three-way analysis of variance, ANOVA (with factors harmonic number, timing of the mistuned component and presence of a component corresponding to the mistuned harmonic in the matching complex) was calculated, using the logarithm of the mean pitch shift for each subject and condition as input. Logarithms were used because we were interested in the relative changes across conditions. The ANOVA showed that the three main effects were highly significant [harmonic number: $F(1,7)=33.48$, $p=0.001$; timing: $F(1,7)=138.8$, $p<0.001$; presence: $F(1,7)=32.78$, $p=0.001$] and that all interactions were significant [timing × harmonic number: $F(1,7)=21.1$, $p<0.01$; timing × presence: $F(1,7)=13.05$, $p<0.01$; harmonic number × presence: $F(1,7)=9.36$, $p<0.05$; timing × harmonic number × presence: $F(1,7)=13.05$, $p<0.05$]. To assess whether the difference between the simultaneous and the post-target conditions were significant for each harmonic alone, repeated-measures two-way ANOVAs were calculated separately for the third and the fourth harmonics. For the fourth harmonic, the ANOVA showed significant main effects of timing [$F(1,7)=25.77$, $p=0.001$], and of presence [$F(1,7)=15.84$, $p<0.01$], but no significant interaction. For the third harmonic there were again significant main effects of timing [$F(1,7)=112.73$, $p<0.001$], and of presence [$F(1,7)=28.46$, $p=0.001$], and a significant interaction between timing and presence [$F(1,7)=9.31$, $p<0.05$].

In summary, pitch shifts were significantly larger for: (1) the fourth rather than the third harmonic being mistuned; (2) simultaneous rather than post-target presentation of the mistuned harmonic; and (3) presence rather than absence of the nominally mistuned harmonic from the matching sound. Additionally, all interactions were significant. The present results for the fourth harmonic replicate and extend the findings of Gockel *et al.* (2005b) and, in spite of the similarity of the means with those of experiment 1 (but not experiment 2) of Ciocca and Darwin (1999), do not replicate their finding of no significant difference between pitch shifts for simultaneous and post-target presentation of the mistuned component. For the fourth harmonic being mistuned and present in the matching sound, this discrepancy probably was caused by the standard errors within subjects and conditions being larger in Ciocca and Darwin's (1999) study (typically around 0.1 Hz, personal communication) than in the current study (about 0.06–0.07 Hz). For their experiment testing the influence of silent gaps of various durations between the mistuned harmonic and the remainder of the target complex, they reported that "The criterion for stopping after five matches was a standard error smaller than 0.4 Hz in all conditions."

Importantly, the reduction in the size of the pitch shifts in the post-target compared to the simultaneous condition was significantly smaller for the fourth (a factor of 1.8) than for the third harmonic (a factor of 4.2), as shown by the significant interaction between harmonic number and timing of the mistuned harmonic. This supports the hypothesis mentioned in the Introduction that, due to the double-octave relationship, the fourth harmonic has a special status with regard to pitch matches obtained when the mistuned harmonic is presented after the remainder of the complex.

The fact that pitch shifts were larger for the fourth than for the third harmonic mistuned theoretically could indicate that the fourth harmonic generally (whether presented after or simultaneously with the remainder of the complex) has a somewhat special status due to its (near) double-octave relationship to the residue pitch of the target. However, several earlier studies which investigated the relative dominance of individual partials in determining the pitch of complex tones using *simultaneous* presentation of all harmonics, did not report a higher dominance of the fourth harmonic (Moore *et al.*, 1985; Dai, 2000; Gockel *et al.*, 2005a). Thus, it seems most likely that, in the present study, for *simultaneous* presentation, the finding of larger pitch shifts for the fourth than for the third harmonic mistuned was due to a combination of factors such as the relatively low $F0$ of 155 Hz, the short stimulus duration of 90 ms, and the specific subjects used.

Pitch shifts were significantly smaller in the absence of the nominally mistuned harmonic from the matching sound than in its presence. This was true for both harmonic numbers tested and supports the idea that subjects may partly have used an analytical listening and matching strategy, as described in Sec. I. If the nominally mistuned harmonic is absent from the matching sound, the mistuned component cannot be compared with it, so the matches will be somewhat less influenced by the mistuned harmonic. When the nominally mistuned harmonic is present in the matching sound, the comparison between the mistuned harmonic and the corresponding harmonic in the matching sound may be partly affected by frequency-shift detectors (Demany and Ramos, 2005).

IV. SUMMARY AND CONCLUSIONS

(1) Pitch shifts were significantly smaller when the matching complex did not contain a component corresponding to the mistuned harmonic, suggesting that an analytical listening and matching strategy contributed to the pitch shifts measured.

(2) Shifts in matches to the residue pitch were significantly smaller when the mistuned component was presented after the remainder of the target complex than when it was presented simultaneously. This contrasts with Ciocca and Darwin's (1999) *statistical* result for the comparison of simultaneous and nonsimultaneous presentation of a mistuned fourth harmonic in the presence of the fourth harmonic in the matching sound; while the mean results of the present study for the fourth harmonic mistuned were similar to some of the results reported by Ciocca and Darwin (1999), the present standard errors within and across subjects were smaller. The significant reduction in pitch shifts for the post-target conditions replicates and extends the results of Gockel *et al.* (2005b), who mistuned the third harmonic.

(3) The factor by which the pitch shift decreased when changing from simultaneous to nonsimultaneous presentation of the mistuned component was greater for the third than for the fourth harmonic. This could indicate that nonsimultaneous components which have an octave or double-octave relationship to the *F*0 are integrated into the estimate of the residue pitch over a *longer* time window than other components. However, an alternative and perhaps more plausible explanation is that listeners use an analytical listening and matching strategy (especially) when the mistuned harmonic is presented nonsimultaneously and therefore pops out. The use of this strategy would result in larger pitch shifts when the mistuned harmonic has an octave or double-octave relationship to the *F*0 than when it does not (see Sec. I). In other words, nonsimultaneous harmonics that have an octave or double-octave relationship to the *F*0 affect pitch matches more than when they do not have this relationship, without actually being temporally integrated into the residue pitch. Thus, obligatory integration of post-target components for the purpose of determining the residue pitch might be less strong than originally estimated by Ciocca and Darwin (1999), even for the fourth harmonic.

ACKNOWLEDGMENTS

This work was supported by EPSRC Grant No. EP/D501571/1. We thank Brian Moore, Christian Lorenzi, Chris Darwin, Valter Ciocca, Ruth Litovsky, and an anonymous reviewer for comments on an earlier version of this paper.

ANSI (**2004**). *ANSI S3.6-2004, Specification of Audiometers* (American National Standards Institute, New York).

Bachem, A. (**1950**). "Tone height and tone chroma as two different pitch qualities," Acta Psychol. **7**, 80–88.

Ciocca, V., and Darwin, C. J. (**1999**). "The integration of nonsimultaneous frequency components into a single virtual pitch," J. Acoust. Soc. Am. **105**, 2421–2430.

Dai, H. (**2000**). "On the relative influence of individual harmonics on pitch judgment," J. Acoust. Soc. Am. **107**, 953–959.

Darwin, C. J. (**1992**). "Listening to two things at once," in *The Auditory Processing of Speech—From Sounds to Words*, edited by M. E. H. Schouten (Mouton de Gruyter, Berlin).

Darwin, C. J., Buffa, A., Williams, D., and Ciocca, V. (**1992**). "Pitch of dichotic complex tones with a mistuned frequency component," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany and K. Horner (Pergamon, Oxford).

Demany, L., and Ramos, C. (**2005**). "On the binding of successive sounds: Perceiving shifts in nonperceived pitches," J. Acoust. Soc. Am. **117**, 833–841.

Gockel, H., Carlyon, R. P., and Plack, C. J. (**2005a**). "Dominance region for pitch: Effects of duration and dichotic presentation," J. Acoust. Soc. Am. **117**, 1326–1336.

Gockel, H., Plack, C. J., and Carlyon, R. P. (**2005b**). "Reduced contribution of a nonsimultaneous mistuned harmonic to residue pitch," J. Acoust. Soc. Am. **118**, 3783–3793.

Grose, J. H., Hall, J. W., and Buss, E. (**2002**). "Virtual pitch integration for asynchronous harmonics," J. Acoust. Soc. Am. **112**, 2956–2961.

Hall, J. W., and Peters, R. W. (**1981**). "Pitch for nonsimultaneous successive harmonics in quiet and noise," J. Acoust. Soc. Am. **69**, 509–513.

Houtsma, A. J. M. (**1979**). "Musical pitch of two-tone complexes and predictions by modern pitch theories," J. Acoust. Soc. Am. **66**, 87–99.

Houtsma, A. J. M. (**1984**). "Pitch salience of various complex sounds," Music Percept. **1**, 296–307.

Houtsma, A. J. M., and Goldstein, J. L. (**1972**). "The central origin of the pitch of pure tones: Evidence from musical interval recognition," J. Acoust. Soc. Am. **51**, 520–529.

Meddis, R., and Hewitt, M. (**1991a**). "Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery. I: Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and Hewitt, M. (**1991b**). "Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery. II: Phase sensitivity," J. Acoust. Soc. Am. **89**, 2883–2894.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853–1860.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.

Wiegrebe, L. (**2001**). "Searching for the time constant of neural pitch extraction," J. Acoust. Soc. Am. **109**, 1082–1092.

Identification of synthetic vowels based on selected vocal tract area functions (L)

Kate Bunton[a] and Brad H. Story[b]
Speech Acoustics Laboratory, Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona 85721

(Received 3 June 2008; revised 14 October 2008; accepted 16 October 2008)

The purpose of this study was to determine the degree to which synthetic vowel samples based on previously reported vocal tract area functions of eight speakers could be accurately identified by listeners. Vowels were synthesized with a wave-reflection type of vocal tract model coupled to a voice source. A particular vowel was generated by specifying an area function that had been derived from previous magnetic resonance imaging based measurements. The vowel samples were presented to ten listeners in a forced choice paradigm in which they were asked to identify the vowel. Results indicated that the vowels [i], [æ], and [u] were identified most accurately for all of speakers. The identification errors of the other vowels were typically due to confusions with adjacent vowels. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3033740]

PACS number(s): 43.70.Bk, 43.70.Mn, 43.70.Aj [AL] Pages: 19–22

I. INTRODUCTION

Magnetic resonance imaging (MRI) has been widely used to acquire volumetric image sets of the head and neck from which vocal tract area functions can be directly measured. These collections of area functions, which are assumed representative of an individual speaker's production of a target vowel or consonant, have then been used in the development of speech production models and speech synthesizers (e.g., Ciocea, 1997; Story, 2005a, 2005b; Mullen *et al.*, 2007).

The similarity of speech sounds produced by area-function-based synthesis to natural speech has been typically assessed by comparing calculated formant frequencies to formant frequencies extracted from recorded speech (Story *et al.*, 1996, 1998; Story, 2005a). Reasonable similarity has been demonstrated; however, stimuli generated based on measured area functions have rarely been evaluated perceptually. This step is important before stimuli generated by simulation of the speech production process are used to answer questions about the perceptual relevance of various types of kinematic and structural variations of the vocal tract (Carré *et al.*, 2001).

Collections of volumetric image sets based on MRI and their analyses have been reported by Story (2005a) and Story *et al.* (1996, 1998) for eight speakers (four females and four males). A second set of data obtained from the speaker presented in Story *et al.*, 1996 has been published as well (Story, 2008). The inventories include area functions (area as a function of distance from the glottis) of a set of 10 or 11 American English vowels ([i,ɪ,e,ɛ,æ,ʌ,ɑ,ɔ,o,ʊ,u]), depending on the particular speaker. Across speakers, vocal tract area functions varied in vocal tract length and other idiosyncratic differences, but were similar with regard to gross shape for each of the target vowels and the location of major constrictions and expansions.

The measured area functions were subsequently used as input to a computer model of one-dimensional acoustic wave propagation in the vocal tract. The synthetic speech samples were then compared, in terms of location of the first three formant frequencies, to recorded natural speech from each speaker. The natural speech samples were recorded with the subject in a supine position with ear plugs in an attempt to simulate, as closely as possible, the conditions experienced in the MRI sessions. Subjects produced speech sounds that corresponded to the static shapes that were acquired with MRI. Percent error based on comparisons of measured and calculated formant frequencies from natural and simulated speech across speakers and formants (F1-F2-F3) ranged from 0.1% to 39%. Errors larger than 30% were calculated for only seven instances and were limited to two speakers. The majority (defined here to be 95%) of the calculated formants differed from those of natural speech by less than 10%. Overall, results indicated that formant locations of the synthesized samples were reasonably well represented compared to the natural productions that were recorded. These comparisons quantify the success of measurement of area functions from MRI images and speech modeling efforts. However, since one aim of developing a speech production model is to understand the relation between area functions and changes in the vocal tract shape result in acoustic characteristics indicative of a phonetic category, perceptual testing of simulated samples based on these area functions is needed. The purpose of the present study was to determine the vowel identification accuracy for simulated vowel samples of eight speakers based on previously reported vocal tract area functions derived from MRI image sets.

[a]Electronic mail: bunton@email.arizona.edu
[b]Electronic mail: bstory@u.arizona.edu

II. METHOD

A. Area function sets

Previously published area functions for eight speakers were used to synthesize vowel samples in the present study (Story *et al.*, 1996, 1998; Story, 2005a). This included four male (range 29–40 years) and four female (range 23–39 years) speakers. Speakers in Story's (2005a) article were identified as SF1, SF2, SF3, SM1, SM2, and SM3, where "F" denotes female and "M" denotes male. The two speakers presented in Story *et al.*, 1996, 1998 will be identified as SM0 and SF0, respectively. Finally, data for a second set of area functions obtained from speaker SM0 in 2002 will be identified as SM0-2.

B. Synthetic vowel samples

A synthetic vowel sample was generated for each area function of each speaker's inventory. Following Hillenbrand and Gayvert (1993), the duration of all samples was set at 0.3 s, and the fundamental frequency (F0) contour varied from 25% above an F0 target to 25% below that same target. The F0 targets for males and females were set at 110 and 220 Hz, respectively. The sample duration was chosen so that it would not be a primary cue in vowel identification (Hillenbrand *et al.*, 2000); that is, 0.3 s is on average shorter than long vowels and longer than short vowels. The samples were generated with a wave-reflection model of the trachea and vocal tract (Liljencrants, 1985; Story, 1995) that included energy losses due to yielding walls, viscosity, heat conduction, and radiation at the lips. The tracheal portion extended from the glottis to the bronchial termination. Its shape was idealized as a tube that is tapered from 0.3 cm^2 to just below the glottis to a constant area of 1.5 cm^2. All synthesized vowels were based on coupling this tracheal configuration to the respective measured area functions, which included their measured vocal tract lengths. The synthesis was driven by the respiratory pressure (P_R) assumed to exist at the bronchial termination of the trachea. In generating each sample for this study, P_R was ramped from 0 to 6000 dyn/cm^2 in 20 ms with a cosine function, similar to Hillenbrand and Gayvert's (1993) ramping of peak amplitude. The voice source was generated by a model of the time-varying glottal area for which wave shape parameters such as F0, amplitude, pulse skewing (skewing quotient), and duty cycle (open quotient) can be varied over the duration of the synthesized speech sound or held constant. The glottal area model was based on the glottal flow pulse model of Rosenberg (1971) but scaled in amplitude for glottal area. For each sample, the F0 followed either the male or female contour detailed above, the maximum glottal opening was set at 0.08 cm^2, the skewing quotient was held at a value of 2.4, and the open quotient was set to 0.6. The appropriateness of these values for both male and female speech might be questioned; however, they were chosen so that the energy in the harmonic components of the glottal flow wave would be similar for all samples. Although these parameters may reduce the femalelike quality of the samples produced with the SF-area functions, this was not considered to be problematic since the listening task was concerned only with phonetic identification.

In addition to the synthetic samples based on the original measured vowel area functions, a sample was also generated from each speaker's mean area function. That is, the mean of all 10 or 11 vowels measured for each speaker. These samples are effectively neutral vowels and were used as precursors to the other samples in the listening tests to provide a context for extrinsic normalization of each speaker (e.g., Ladefoged and Broadbent, 1957).

C. Listening Task

Ten listeners (mean age 26 years) participated in the present study. Listeners were native English speakers and native to Arizona and passed a hearing screening. All procedures were approved by the Institutional Review Board at the University of Arizona.

An ALVIN interface (Hillenbrand and Gayvert, 2005) was used to present samples via loudspeakers to listeners seated in a sound treated room. Samples were presented in pairs with the first sample being the mean vowel of a particular speaker followed by a target vowel from the same speaker. The computer screen displayed buttons for 11 English vowels that were labeled with both the phonetic symbol and an example "hVd" word. Listeners were asked to identify the second vowel in the pair. Vowel samples were blocked by speaker and each listener heard five repetitions of each vowel. The order of presentation for speaker and vowel samples was randomized. Each listening session lasted no longer than 30 min. A confusion matrix based on listener identification of the vowel samples was calculated separately for each speaker. Listeners also completed a training task with vowel samples recorded by a male speaker (second author) to assure they could identify all 11 English vowels. Accuracy was greater than 98% across vowels and listeners. Errors were limited to a confusion of [ɔ] and [ɑ].

III. RESULTS

Identification errors made for the vowels based on each speaker are indicated in the confusion matrices displayed in Tables I–III. In each matrix, the target vowel is listed in the leftmost column and the vowel identified is listed across the top of the columns. Accurate identification of target tokens can be seen along the diagonal in the boldface cells.

Accuracy across vowels varied from a low of 21% for female [ɛ] to a high of 98% for male [i]. Vowels with the highest accuracy rates across speakers (>89%) included three English corner vowels [i, æ, u]. Accuracies for the three vowels [ɛ, ɔ, ɑ] were greater than 50% for the male speakers and less than 50% for the female speakers. For the vowels [ɪ, e, o] and [ʌ] identification accuracy was less than 50% for both male and female speakers.

Although there was considerable variability in the identification accuracy, vowel confusions were typically between adjacent vowel categories in the vowel space. For example, the target vowel [ɪ] was identified as either [e] or [ɛ] for all of the speakers except SF2 whose [ɪ] targets were identified

TABLE I. Confusion matrices for synthesized vowels of speakers SM0 and SM0–2.

		Listener's identification											
		i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u	Total
Vowel intended by speaker SM0	i	**50**	0	0	0	0	0	0	0	0	0	0	50
	ɪ	0	**1**	32	17	0	0	0	0	0	0	0	50
	e	0	0	**0**	0	0	0	0	0	0	0	0	0
	ɛ	0	0	1	**25**	23	0	1	0	0	0	0	50
	æ	0	0	0	1	**48**	0	1	0	0	0	0	50
	ʌ	0	0	0	0	0	**22**	22	6	0	0	0	50
	ɑ	0	0	0	0	1	0	**29**	20	0	0	0	50
	ɔ	0	0	0	0	0	0	8	**28**	14	0	0	50
	o	0	0	0	0	0	0	1	1	**9**	6	33	50
	ʊ	0	0	0	0	0	0	0	0	0	**9**	41	50
	u	0	0	0	0	0	0	0	0	1	6	**43**	50
Vowel intended by speaker SM0-2	i	**50**	0	0	0	0	0	0	0	0	0	0	50
	ɪ	6	**23**	16	5	0	0	0	0	0	0	0	50
	e	0	8	**24**	18	0	0	0	0	0	0	0	50
	ɛ	0	0	2	**35**	13	0	0	0	0	0	0	50
	æ	0	0	0	0	**49**	0	1	0	0	0	0	50
	ʌ	0	0	0	0	0	**0**	10	39	0	1	0	50
	ɑ	0	0	0	0	0	0	**27**	23	0	0	0	50
	ɔ	0	0	0	0	0	0	5	**27**	16	2	0	50
	o	0	0	0	0	0	0	1	6	**43**	0	0	50
	ʊ	0	0	0	0	0	0	2	5	43	**0**	0	50
	u	0	0	0	0	0	0	0	0	0	3	**47**	50

TABLE II. Confusion matrices for synthesized vowels of speakers SM1, SM2, and SM3.

		Listener's identification											
		i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u	Total
Vowel intended by speaker SM1	i	**46**	3	0	0	0	0	0	0	0	0	1	50
	ɪ	0	**37**	1	2	0	0	0	0	0	5	5	50
	e	0	14	**4**	13	2	5	0	0	0	9	3	50
	ɛ	0	2	2	**21**	3	11	0	0	0	10	1	50
	æ	0	0	0	1	**48**	0	1	0	0	0	0	50
	ʌ	0	0	0	0	0	**41**	2	0	0	7	0	50
	ɑ	0	0	0	0	1	0	**34**	15	0	0	0	50
	ɑ	0	0	0	0	0	0	19	**31**	0	0	0	50
	o	0	0	0	0	0	0	0	0	**7**	25	18	50
	ʊ	0	0	0	0	0	0	0	0	0	**17**	33	50
	u	0	0	0	0	0	0	0	0	0	5	**45**	50
Vowel intended by speaker SM2	i	**50**	0	0	0	0	0	0	0	0	0	0	50
	ɪ	0	**33**	5	12	0	0	0	0	0	0	0	50
	e	4	22	**14**	10	0	0	0	0	0	0	0	50
	ɛ	0	0	0	**0**	0	0	0	0	0	0	0	0
	æ	0	0	0	0	**50**	0	0	0	0	0	0	50
	ʌ	0	0	0	0	0	**42**	0	0	1	6	1	50
	ɑ	0	0	0	0	0	0	**43**	7	0	0	0	50
	ɔ	0	0	0	0	0	0	4	**46**	0	0	0	50
	o	0	0	0	0	0	2	2	3	**14**	23	6	50
	ʊ	0	0	0	0	0	7	0	0	2	**38**	3	50
	u	0	0	0	0	0	0	0	0	0	8	**42**	50
Vowel intended by speaker SM3	i	**49**	1	0	0	0	0	0	0	0	0	0	50
	ɪ	0	**1**	37	12	0	0	0	0	0	0	0	50
	e	0	0	**26**	23	1	0	0	0	0	0	0	50
	ɛ	0	1	21	**26**	2	0	0	0	0	0	0	50
	æ	0	0	1	0	**49**	0	0	0	0	0	0	50
	ʌ	0	0	0	0	0	**0**	0	0	10	36	4	50
	ɑ	0	0	0	0	0	0	**35**	15	0	0	0	50
	ɔ	0	0	0	0	0	0	8	**42**	0	0	0	50
	o	0	0	0	0	0	13	8	9	**19**	0	1	50
	ʊ	0	0	0	0	0	2	0	0	1	**31**	16	50
	u	0	0	0	0	0	0	0	0	1	7	**42**	50

TABLE III. Confusion matrix for synthesized vowels of speaker SF0.

		Listener's identification											
		i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u	Total
Vowel intended by speaker SF0	i	**50**	0	0	0	0	0	0	0	0	0	0	50
	ɪ	0	**10**	9	26	5	0	0	0	0	0	0	50
	e	0	0	**0**	0	0	0	0	0	0	0	0	0
	ɛ	0	0	1	**8**	41	0	0	0	0	0	0	50
	æ	0	0	0	6	**44**	0	0	0	0	0	0	50
	ʌ	0	0	0	0	0	**0**	39	11	0	0	0	50
	ɑ	0	1	0	0	0	22	**19**	5	2	1	0	50
	ɔ	0	0	0	0	0	2	6	**23**	19	0	0	50
	o	0	0	0	0	0	12	5	3	**13**	17	0	50
	ʊ	0	0	0	0	0	0	25	25	0	**0**	0	50
	u	0	0	0	0	0	1	0	0	0	5	**44**	50
Vowel intended by speaker SF1	i	**49**	0	1	0	0	0	0	0	0	0	0	50
	ɪ	0	**23**	20	7	0	0	0	0	0	0	0	50
	e	0	1	**10**	36	2	1	0	0	0	0	0	50
	ɛ	0	3	0	**1**	0	5	0	0	0	33	8	50
	æ	0	0	0	0	**50**	0	0	0	0	0	0	50
	ʌ	0	0	0	0	0	**30**	3	0	5	11	1	50
	ɑ	0	0	0	0	46	0	**3**	1	0	0	0	50
	ɔ	0	0	0	0	0	0	16	**33**	1	0	0	50
	o	0	0	0	0	0	6	6	0	**24**	14	0	50
	ʊ	0	0	0	0	0	0	0	0	0	**29**	21	50
	u	0	0	0	0	0	0	0	0	0	6	**44**	50
Vowel intended by speaker SF2	i	**27**	22	0	1	0	0	0	0	0	0	0	50
	ɪ	0	**0**	5	0	0	35	0	0	0	10	0	50
	e	2	29	**7**	11	1	0	0	0	0	0	0	50
	ɛ	0	0	5	**1**	44	0	0	0	0	0	0	50
	æ	0	0	0	0	**49**	0	0	1	0	0	0	50
	ʌ	0	0	0	0	0	**39**	4	1	2	3	1	50
	ɑ	0	0	0	0	0	1	**12**	22	15	0	0	50
	ɔ	0	0	0	0	0	0	32	**18**	0	0	0	50
	o	0	0	0	0	0	2	0	0	**11**	34	3	50
	ʊ	0	0	0	0	0	0	0	1	0	**42**	7	50
	u	0	0	0	0	0	0	0	0	0	4	**46**	50
Vowel intended by speaker SF3	i	**48**	2	0	0	0	0	0	0	0	0	0	50
	ɪ	2	**9**	30	9	0	0	0	0	0	0	0	50
	e	1	21	**21**	7	0	0	0	0	0	0	0	50
	ɛ	0	4	11	**31**	0	3	0	0	0	1	0	50
	æ	0	0	0	0	**49**	0	0	1	0	0	0	50
	ʌ	0	0	0	0	0	**1**	0	0	3	38	8	50
	ɑ	0	0	0	0	0	0	**30**	20	0	0	0	50
	ɔ	0	0	0	0	0	0	22	**28**	0	0	0	50
	o	0	0	0	0	0	0	4	6	**22**	17	1	50
	ʊ	0	0	0	0	0	0	0	0	0	**10**	40	50
	u	0	0	0	0	0	0	0	0	0	6	**44**	50

as [ʌ]. A similar confusion was found for the target [e], which was identified as either [ɪ] or [ɛ] for all speakers. Identification of the target vowel [ɑ] included both [æ] and [ɔ] responses, and is the only vowel where listener identification was not an adjacent vowel category. The target [ɔ] was most commonly confused with [ɑ]. This confusion is not unexpected and the vowel [o] was confused with [ʊ] in a majority of cases. Confusion between [ɔ] and [ɑ] is not unexpected given that speakers in the southwest part of the United States tend to collapse these two categories (Labov, 1996).

Comparisons of vowel identification accuracy for the samples based on the data sets from the same speaker (SM0 and SM0-2) were similar, with overall accuracy slightly higher for the second data set (53% and 59%, respectively). The largest difference between samples was seen for the vowels [ɪ] and [o]. In both cases, confusions were between adjacent vowels.

IV. DISCUSSION

The confusion matrices suggest that most of the area functions from each speaker's inventory produce sound samples that can be expected to be identified as either the target vowel or as an adjacent vowel in the vowel space. Therefore, with a few exceptions, each area function is representative of the "neighborhood" of the target vowel. The modest accuracy rates for several vowels, however, also beg the question of why the identification accuracy is not better.

An obvious possibility is that some of the area functions are simply not good representations of the target vowels. In some cases, this is likely true. For example, the poor identification of SF0's [ʊ] vowel could have likely been predicted based on the fairly large errors found between the formant frequencies calculated from the [ʊ] area function and those measured from natural speech (Story *et al.*, 1998). In other cases, a presumably good area function representation of a particular vowel would not have predicted poor identification accuracy. SM2's [ɪ] area function produced formant frequencies with small error relative to natural speech and yet the identification responses indicated that listeners were correct only 66% of the time. Although area function quality is undoubtedly part of the problem, it would seem that other factors must also contribute.

The constant 0.3 s duration that was used to generate every sample may have affected some identification responses, especially for the "short" vowels. This duration was chosen as a compromise between short and long vowels (Hillenbrand and Gayvert, 1993), but may have been too long such that it inadvertently created a cue that conflicted with the typical duration of some of the shorter vowels.

Another possible reason for reduced identification accuracy is that each vowel sample was generated from a "static" area function. That is, each vowel was effectively produced without any change in vocal tract shape and, hence, no change in formant frequencies. In connected speech, vowels are typically embedded between consonants so that the formant frequencies are almost continuously in transition. Even productions of isolated vowels tend to have formant transitions over the course of the utterance (e.g., Story, 2007). There is much evidence that listeners use this dynamic spectral change for identification of vowels (Jenkins *et al.*, 1983; Strange *et al.*, 1983; Nearey, 1989; Hillenbrand and Gayvert, 1993; Nittrouer, 2007).

Finally, the listening paradigm, which consisted of presentations blocked by speaker and included a precursor mean vowel followed by the target, may have influenced the identification accuracy. This paradigm was implemented so that the precursor might allow for extrinsic normalization by the listener. Similar methods have been used with some success for vowel recognition algorithms (Pols and Weenink, 2005; Nearey and Assman, 2007).

The next steps in this research are to explore some of these possible influences on vowel identification; specifically, use of area functions for each speaker that have been "tuned" to produce formant frequencies directly aligned with those of recorded speech (Story, 2006), use of an area function model that allows for time variation of the vocal tract shape (e.g., Story, 2005b), build in natural vowel durations, and use of a listening paradigm that does not include a precursor vowel for normalization.

ACKNOWLEDGMENTS

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

Carré, R., Ainsworth, W. A., Jospa, P., Maeda, S., and Pasdeloup, V. (**2001**). "Perception of vowel-to-vowel transitions with different formant trajectories," Phonetica **58**, 163–178.

Ciocea, S. (**1997**). "Semi-analytic formant-to-area mapping," Ph.D. thesis, Université Libre de Bruxelles, Brussels, Belgium.

Hillenbrand, J., and Gayvert, R. T. (**1993**). "Identification of steady-state vowels synthesized from the Peterson and Barney measurements," J. Acoust. Soc. Am. **94**, 668–674.

Hillenbrand, J., Clark, M., and Houde, R. (**2000**). "Some effects of duration on vowel recognition," J. Acoust. Soc. Am. **108**, 3013–3022.

Hillenbrand, J., and Gayvert, R. T. (**2005**). "Open source software for experiment design and control," J. Speech Lang. Hear. Res. **48**, 45–60.

Jenkins, J. J., Strange, W., and Edman, T. R. (**1983**). "Identification of vowels in 'vowelless' syllables," Percept. Psychophys. **34**, 441–450.

Labov, W. (**1996**). "The organization of dialectic diversity in North America," presented at the Fourth International Conference on Spoken Language Proceeding, Philadelphia, 6 October; Available online at www.ling.upenn.edu/phono_atlas/ICSLP4.html (Last viewed 10/9/2008).

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Liljencrants, J. (**1985**). "Speech synthesis with a reflection-type line analog," DS thesis, Department of Speech Communication and Music Acoustic, Royal Institute of Technology, Stockholm, Sweden.

Mullen, J., Howard, D. M., and Murphy, D. T. (**2007**). "Real-time dynamic articulations in the 2-D waveguide mesh vocal tract model," IEEE Trans. Audio, Speech, Lang. Process. **15**, 577–585.

Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.

Nearey, T. M., and Assmann, P. (**2007**). "Probabilistic 'sliding-template' models for indirect vowel normalization," in *Experimental Approches to Phonology*, edited by M. Sole, P. Speeter Beddor, and M. Ohala (Oxford University Press, Oxford).

Nittrouer, S. (**2007**). "Dynamic spectral structure specifies vowels for children and adults," J. Acoust. Soc. Am. **122**, 2328–2339.

Pols, L., and Weenink, D. (**2005**). "Vowel recognition and (adaptive) speaker normalization," *Proceedings of the Tenth International Conference on Speech and Computer*, edited by G. Kokkinakis (University of Patras Press, Patras, Greece), Vol. **1**, 17–24.

Rosenberg, A. (**1971**). "Effect of glottal pulse shape on the quality of natural vowels," J. Acoust. Soc. Am. **49**, 583–590.

Story, B. H. (**1995**). "Speech simulation with an enhanced wave-reflection model of the vocal tract," .Ph.D. thesis, University of Iowa, Iowa City, IA.

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

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1998**). "Vocal tract area functions for an adult female speaker based on volumetric imaging," J. Acoust. Soc. Am. **104**, 471–487.

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

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

Story, B. H. (**2006**). "A technique for 'tuning' vocal tract area functions based on acoustic sensitivity functions," J. Acoust. Soc. Am. **119**, 715–718.

Story, B. H. (**2007**). "Time-dependence of vocal tract modes during production of vowels and vowel sequences," J. Acoust. Soc. Am. **121**, 3770–3789.

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

Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **74**, 695–705.

An analysis of the masking of speech by competing speech using self-report data (L)

Trevor R. Agus[a] and Michael A. Akeroyd
MRC Institute of Hearing Research (Scottish Section), Glasgow Royal Infirmary, Alexandra Parade, Glasgow G31 2ER, United Kingdom

William Noble and Navjot Bhullar
School of Behavioural, Cognitive and Social Sciences, University of New England, Armidale NSW 2351, Australia

(Received 19 March 2008; revised 20 October 2008; accepted 21 October 2008)

Many of the items in the "Speech, Spatial, and Qualities of Hearing" scale questionnaire [S. Gatehouse and W. Noble, Int. J. Audiol. **43**, 85–99 (2004)] are concerned with speech understanding in a variety of backgrounds, both speech and nonspeech. To study if this self-report data reflected informational masking, previously collected data on 414 people were analyzed. The lowest scores (greatest difficulties) were found for the two items in which there were two speech targets, with successively higher scores for competing speech (six items), energetic masking (one item), and no masking (three items). The results suggest significant masking by competing speech in everyday listening situations. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3025915]

PACS number(s): 43.71.Ky, 43.71.Lz, 43.66.Sr [RLF] Pages: 23–26

I. INTRODUCTION

In recent years there has been growing interest in the "informational masking" of speech (e.g., Freyman *et al.*, 1999; Brungart, 2001; Li *et al.*, 2004). This phenomenon is the extra masking observed when the identifiability of speech is measured in a background of competing speech versus that predicted from control conditions with comparable acoustics, such as a background of noise (cf. Carhart *et al.*, 1969). Informational masking is associated with uncertainty, target-masker similarity, and other "higher-level" aspects of the target and masker stimuli (cf. Durlach *et al.*, 2003). However, most of these measurements reflect situations where there are few cues available with which to distinguish the target and competing talkers. Laboratory measurements of the amount of informational masking of target speech by competing speech can be as much as 22 dB (Arbogast *et al.*, 2002), but is reduced considerably when supplementary cues are provided, including differences in talker gender (Brungart, 2001), pitch (Drullman and Bronkhorst, 2004), azimuth (Freyman *et al.*, 1999), and speech reading (Helfer and Freyman, 2005). Since many of these cues are typically available to most listeners in everyday listening situations, it seems unlikely that large informational-masking effects would be observed outside of the laboratory.

One approach to investigating the difficulties in real life due to competing speech is to actually ask participants about them. To this end, the items included in the "Speech, Spatial and Qualities of Hearing" scale (SSQ) (Gatehouse and Noble, 2004) are particularly useful because they ask about listening situations involving competing speech. The SSQ was designed to measure "auditory disability," the restriction of auditory ability to perform an activity considered to be normal for a human being, as distinct from "auditory impairment," the underlying loss of auditory function (cf. World Health Organisation, 1980). The items cover a wide range of auditory abilities, including speech perception (e.g., "Can you easily have a conversation on the telephone?"), spatial hearing (e.g., "Can you tell how far a bus or truck is, by the sound?"), and "qualities" of hearing (e.g., "Do other peoples' voices sound clear and natural?"). The questionnaire is designed to be completed by interview, with participants being read each vignette and then asked to respond on a visual scale from 0 ("not at all") to 10 ("perfectly"). Thus, lower scores correspond to greater auditory disability.

Here, we reanalyze some previously collected SSQ data and focus on the 14 items that relate to speech perception: many of them refer to situations involving the comprehension of speech in competing speech, and so responses might be expected to indicate the total masking of competing speech and to help distinguish informational masking from energetic masking.

II. METHOD

The data were collected as part of an earlier survey.[1] A short questionnaire was mailed to a random sample of 9000 individuals in the Glasgow area. The questionnaire used a set of screening questions including: (1) whether the individuals self-reported any hearing loss, the difficulties any hearing loss caused them, and whether they wore hearing aids, (2) a set of six of the SSQ items, with two from each of the speech, spatial, and qualities of hearing sections (Gatehouse and Noble, 2004), (3) what their opinions were about the provision of hearing aids by the UK National Health Service, (4) their age and their sex, and (5) their overall quality of life. The study was designed to stratify 50 people in each of the eight groups of a $2\times2\times2$ factorial design: 50–64 years versus 65–80 years; no hearing difficulty versus hearing dif-

[a]Electronic mail: trevor@ihr.gla.ac.uk

TABLE I. Mean and standard deviations of scored responses to speech items from the SSQ for those with and without self-reported hearing difficulties. Items are grouped by categorization then ordered by mean responses.

SSQ item	Vignette	Category	Mean (SD) hearing difficulties	
			Yes	No
10.	You are listening to someone talking to you, while at the same time trying to follow the news on TV. Can you follow what both people are saying?	Two targets	4.4 (*2.4*)	6.1 (*2.5*)
14.	You are listening to someone on the telephone and someone next to you starts talking. Can you follow what's being said by both speakers?	Two targets	4.8 (*2.6*)	6.2 (*2.4*)
6.	You are in a group of about five people in a busy restaurant. You cannot see everyone else in the group. Can you follow the conversation?	Multitalker masker	4.9 (*2.4*)	6.7 (*2.3*)
4.	You are in a group of about five people in a busy restaurant. You can see everyone else in the group. Can you follow the conversation?	Multitalker masker	5.4 (*2.5*)	7.3 (*2.2*)
11.	You are in conversation with one person in a room where there are many other people talking. Can you follow what the person you are talking to is saying?	Multitalker masker	5.7 (*2.4*)	7.3 (*2.3*)
8.	Can you have a conversation with someone when another person is speaking whose voice is the same pitch as the person you're talking to?	Single-talker masker	5.9 (*2.4*)	7.3 (*2.2*)
9.	Can you have a conversation with someone when another person is speaking whose voice is different in pitch from the person you're talking to?	Single-talker masker	6.0 (*2.2*)	7.4 (*2.1*)
1.	You are talking with one other person and there is a TV on in the same room. Without turning the TV down, can you follow what the person you're talking to says?	Single-talker masker	6.0 (*2.4*)	7.4 (*2.2*)
5.	You are talking with one other person. There is continuous background noise, such as a fan or running water. Can you follow what the person says?	Noise masker	6.4 (*2.3*)	7.8 (*2.1*)
3.	You are in a group of about five people, sitting round a table. It is an otherwise quiet place. You can see everyone else in the group. Can you follow the conversation?	No masker	6.8 (*2.5*)	8.2 (*2.0*)
13.	Can you easily have a conversation on the telephone?	No masker	7.9 (*2.2*)	8.8 (*1.9*)
2.	You are talking with one other person in a quiet, carpeted lounge-room. Can you follow what the other person says?	No masker	8.1 (*1.9*)	8.9 (*1.8*)
12.	You are with a group and the conversation switches from one person to another. Can you easily follow the conversation without missing the start of what each new speaker is saying?		*5.4 (2.5)*	*6.9 (2.5)*
7.	You are talking to someone in a place where there are a lot of echoes, such as a church or railway terminus building. Can you follow what the other person says?		*5.6 (2.5)*	*7.5 (2.2)*

ficulty; low score on the six-question form of the SSQ versus high score. There were 3824 respondents, of whom 1979 were 50–80 years old. Of these, 576 were selected for a face-to-face interview, to try to get 50 completed interviews per group. There, an interviewer conducted a hearing handicap questionnaire and the SSQ questionnaire (both are fully listed in Gatehouse and Noble, 2004), together with a small number of questions about the individual. The hearing difficulty was categorized by their yes/no response to the question "Do you have difficulty with your hearing?" In practice, 414 successful interviews were obtained, giving group sizes from 49 to 54. We used part of the stratification here, comparing the "younger" versus "older" groups (N=208, 206, respectively) and "no hearing difficulty" versus "hearing difficulty" groups (N=209, 205, respectively).[2]

For the present analysis, the speech items of the SSQ were split into six categories (see Table I) according to the configuration of target speech and competing speech described by the vignette. In four categories, the items asked about following one person in a variety of types of masker: no maskers (item Nos. 2, 3, and 13), a noise masker, interpreted as purely energetic (No. 5), single-talker maskers (Nos. 1, 8, and 9), and multitalker maskers (Nos. 4, 6, and 11). In a fifth category, the items asked about following two targets at the same time (Nos. 10 and 14). The sixth category contained two items that were omitted from the analysis because they were felt to be ambiguous about whether or not the masker should be considered to be competing speech (Nos. 7 and 12).

III. RESULTS

Table I reports the scores for each of the 14 speech items in the questionnaire separately for the normal-hearing and hearing-disabled participants. The items are ordered by mean score in the normal group (note that the ordering is the *same* for the hearing-disabled group). Figure 1 shows the data averaged across both groups, with items ordered in the same way: the highest mean responses (i.e., the least disability) were for the no-masker items and the lowest for the two-

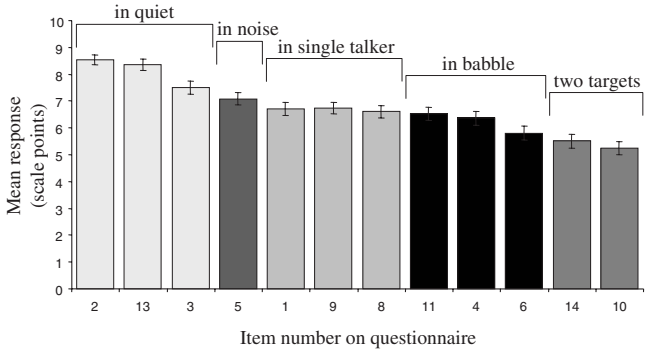


FIG. 1. Mean scores (out of 10) for each of the SSQ items included in the analysis, reverse-ordered by mean score, and grouped by categorization. Error bars are 95% confidence intervals.

target items. The mean responses for the other items fell between those two in an ordered manner, from noise-masker items, single-talker items, to multitalker items. Note that there was no overlap between the five categories of items.

The score for each item was lower for the hearing-disabled listeners than for the normal-hearing listeners by, on average, 1.4 scale points.[3] Accordingly, to compare the effects for these groups we calculated a within-listener "single-talker effect," "multitalker effect," and "double-target effect," as the difference between the score for the respective category and the baseline of the noise-masker category. Figure 2 shows the results for the normal-hearing listeners (open bars) and hearing-disabled listeners (filled bars). The ordering of the magnitudes of the three effects followed that expected from Fig. 1—the single-talker effect showed the least, the double-target effect showed the greatest—but only the multitalker effect showed a difference across normal-hearing versus hearing-disabled (a subsequent MANOVA demonstrated that this was a significant effect: $[F(1,406)=4.31,\ p=0.04]$). None of these three effects were correlated with the participants' ages, handicap, or overall quality of life.

The mean responses of the SSQ and the handicap questionnaire were negatively correlated $[r(410)=-0.60,\ p<0.001]$, showing that the participants who reported the greatest disability also reported the greatest handicap, replicating Gatehouse and Noble's (2004) result.

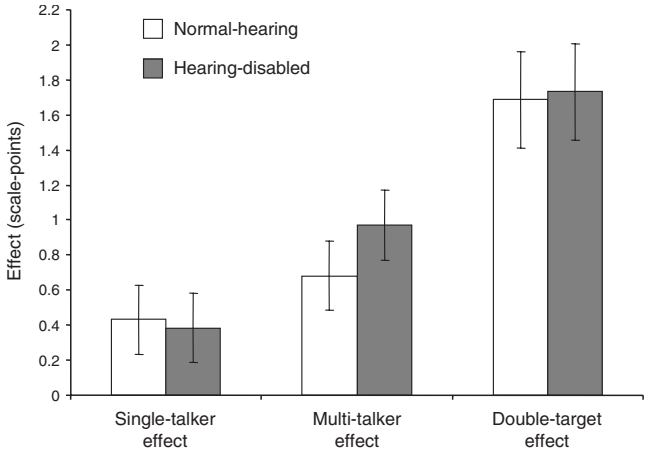


FIG. 2. Mean single-talker, multitalker, and double-target effects for normal-hearing and hearing-disabled listeners, relative to the score for the noise masker item (item 5). Error bars are 95% confidence intervals.

IV. DISCUSSION

This analysis has shown that the scores on those items of the SSQ questionnaire that involved competing-speech situations (i.e., listening in multitalker situations, or to two simultaneous targets) were lower than the scores of those items that involved other, arguably simpler situations. Given that the dataset came from face-to-face interviews with the general public, it clearly demonstrates that competing speech is a problem for many people and that it contributes to auditory difficulties in the real world. It also shows that the respondents are at least aware of the difficulties they have in the everyday listening situations that could involve informational masking.

Nevertheless, while it is possible that a source of these competing-speech effects is informational masking, we cannot necessarily be certain that it actually is. In none of the auditory circumstances described by these vignettes can we distinguish with confidence the informational-masking component of the total masking from the energetic-masking component; indeed, if a pair of vignettes adequately described two situations with acoustically equivalent maskers that differed only in their informational content, then it seems likely that the descriptions would have to be so carefully written and prescriptive that they would not be meaningful to most participants. Such is one of the costs of a questionnaire: it is not possible to describe a situation with the same control and precision as would be expected in a laboratory experiment. But such also is the power of the questionnaire: one can describe quickly and effectively a situation that many people would recognize and understand, and so obtain data on larger, more representative samples of people than would normally be possible experimentally. To distinguish informational masking from energetic masking, it is perhaps necessary to return to laboratory situations, but it is not clear which simple laboratory experiments would best reflect the extent of the role of informational masking in everyday listening situations, especially as different laboratory measures lead to vastly different estimates of informational masking (compare Brungart, 2001; Arbogast *et al.*, 2002; Wu *et al.*, 2004).

Some of the vignettes in the questionnaire parallel laboratory experiments on informational masking, such as those comparing single-talker maskers with multitalker maskers (Carhart *et al.*, 1975; Brungart *et al.*, 2001). In the current analysis, listeners reported more difficulty understanding one of many talkers than one of two talkers. This could be either a reflection of the expected extra difficulty when the number of maskers are increased or a reflection of the extra total power of the maskers when there are more talkers. It is not even clear whether the multitalker maskers would cause more informational masking than single-talker maskers: if the vignettes were interpreted as many voices each at similar levels, this might result in a semicontinuous speech babble with properties similar to a noise (see Carhart *et al.*, 1975; Brungart *et al.*, 2001), but in many real-life situations, some of the nearest talkers can be more clearly heard, and could cause informational masking (note also that one's attention could be taken by competing talkers, so reducing, at least momentarily, what is heard from the target). The present results demonstrated that the multitalker effect was larger for the hearing-disabled people than for normal-hearing people; perhaps the hearing-disabled participants were most aware of their hearing difficulties in what were arguably the most acoustically challenging situations with the likely greatest amount of power in the maskers.

The tasks for which the greatest hearing difficulties were reported involved following two simultaneous talkers. It is likely that this task has greater cognitive demands than following just one of two simultaneous talkers. This cognitive effect is more apparent in a comparison of two of the items (1 and 10) whose vignettes describe a person talking at the same time as a television is on in the room. They are distinguished by whether a listener can follow just the talker (item

1) or both the talker and the TV (item 10). Thus, the two situations were acoustically similar yet with different tasks with seemingly different cognitive demands. The participants on average reported lower scores by 1.4 scale-points when following both talkers than when following just the talker [6.7 versus 5.2 scale-points; $t(403)=13.49$, $p<0.001$]. Since the two situations are acoustically similar, this effect is likely to be due to the extra cognitive challenges involved in trying to understand two simultaneous talkers. In the laboratory, it is possible to observe listeners' ability to correctly report one or both of two talkers. The results are varied: some show decreased performance with the increased cognitive load (e.g., Broadbent, 1954; Humes *et al.*, 2006), but others do not (Ihlefeld and Shinn-Cunningham, 2008a, 2008b). Since the self-report data reported here came from elderly participants, it may exaggerate any difficulty with divided attention in the general population (Humes *et al.*, 2006).

There are further informative contrasts available in the SSQ data. First, items 8 and 9 of the SSQ involved following speech in competing speech when the pitches of the two talkers were either the same or different. This is perhaps analogous to the psychophysical measurements of speech understanding that have shown that competing speech is a less effective masker when it is spoken by a talker of the opposite gender (e.g., Brungart, 2001) or a different pitch to the target speech (Darwin *et al.*, 2003). There was a small but significant difference of 0.2 scale-points between these two items [$t(396)=2.34$, $p=0.02$], such that the listeners' responses were lower for same-pitch competing speech than for different-pitch competing speech. Thus, this contrast in the SSQ data is in the same direction as would be predicted from the psychoacoustical data. Second, items 4 and 6 of the SSQ describe the same scene (of following a conversation in a busy restaurant) and only differ in that the listener can see everyone in the group (item 4) or cannot see everyone in the group (item 6). MacLeod and Summerfield (1990) showed that speech-reception thresholds were on average 11 dB lower when visual speech cues were provided. Thus, a higher response would be expected for item 4 than for item 6. The mean response to item 4 was indeed significantly higher than the mean response to item 6 [6.4 versus 5.8; $t(398)=8.45$, $p<0.001$]. Again, the direction of the effect supports the link between the SSQ and the experimental data.

In summary, the respondents generally reported the most hearing difficulties for speech items with more than one talker. The largest effects were for items that involved two targets, and the smallest effects were for items with a single competing talker. For multiple competing talkers, the effects were larger for hearing-disabled listeners than for normal-hearing listeners. The results show that masking by competing speech is tangible in everyday listening situations. The descriptions of these situations in the SSQ questionaire suggest that they could involve informational masking, but there may also be some contribution from energetic masking.

ACKNOWLEDGMENTS

The first author was funded by a Ph.D. studentship from the Medical Research Council Institute of Hearing Research.

[1]The postal survey and face-to-face interviews were commissioned by Stuart Gatehouse (died, February 2007) and conducted by the National Centre for Social Research, London, in 2004. The full results are currently unpublished.

[2]A further two participants were excluded from the analysis because they omitted to answer the majority of the SSQ items relating to speech. There were 50 participants included in the analysis who omitted to answer one or more questions (a total of 103 missing responses). Where possible, means and other statistics were calculated from the data provided, without any corrections. In the rare cases where a participant answered none the items crucial to calculating a certain statistic, such a participant was excluded for the calculation of that statistic.

[3]This difference in overall SSQ score is somewhat surprising, as the original selection for interviews was designed to give approximately equal distributions of SSQ scores across the normal hearing versus hearing-disability stratification. We believe that the effect was due to the initial selection of participants being based on a much-reduced, 6-item, version of the SSQ, rather than the full questionnaire. The difference partly motivated our analysis of the relative effects, rather than mean scores, in Fig. 2.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Broadbent, D. E. (**1954**). "The role of auditory localization in attention and memory span," J. Exp. Psychol. **47**, 191–196.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Carhart, R., Johnson, C., and Goodman, J. (**1975**). "Perceptual masking of spondees by combinations of talkers," J. Acoust. Soc. Am. **58**, 35.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple background sounds," J. Acoust. Soc. Am. **45**, 694–703.

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.

Drullman, R., and Bronkhorst, A. W. (**2004**). "Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers," J. Acoust. Soc. Am. **116**, 3090–3098.

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

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Gatehouse, S., and Noble, W. (**2004**). "The speech, spatial and qualities of hearing scale (SSQ)," Int. J. Audiol. **43**, 85–99.

Helfer, K. S., and Freyman, R. L. (**2005**). "The role of visual speech cues in reducing energetic and informational masking," J. Acoust. Soc. Am. **117**, 842–849.

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.

Ihlefeld, A., and Shinn-Cunningham, B. (**2008a**). "Spatial release from energetic and informational masking in a selective speech identification task," J. Acoust. Soc. Am. **123**, 4369–4379.

Ihlefeld, A., and Shinn-Cunningham, B. (**2008b**). "Spatial release from energetic and informational masking in a divided speech identification task," J. Acoust. Soc. Am. **123**, 4380–4392.

Li, L., Daneman, M., Qi, J. G., and Schneider, B. A. (**2004**). "Does the informational content of an irrelevant source differentially affect spoken word recognition in younger and older adults," J. Exp. Psychol. Hum. Percept. Perform. **30**, 1077–1099.

MacLeod, A., and Summerfield, Q. (**1990**). "A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," Br. J. Audiol. **24**, 29–43.

World Health Organisation (**1980**). *International Classification of Impairments, Disabilities and Handicaps* (World Health Organisation, Geneva).

Wu, X., Wang, C., Chen, J., Hongwei, Q., Li, W., Wu, Y., Schneider, B. A., and Li, L. (**2004**). "The effect of perceived spatial separation on informational masking of Chinese speech," Hear. Res. **199**, 1–10.

Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal (L)

Christian Lorenzi[a)]
Equipe Audition (LPP CNRS, Universite Paris Descartes, and DEC, Ecole Normale Superieure), 29 rue d'Ulm, 75005 Paris, France and GRAEC, CNRS (Groupement de Recherche en Audiologie Experimentale et Clinique, GDR #2967), 29 rue d'Ulm, 75005 Paris, France

Louis Debruille, Stéphane Garnier, and Pierre Fleuriot
GRAEC, CNRS (Groupement de Recherche en Audiologie Expérimentale et Clinique, GDR #2967), 29 rue d'Ulm, 75005 Paris, France and Groupement d'Audioprothésistes Entendre, 65 Rue des 3 Fontanot, 92000 Nanterre, France

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 2 November 2007; revised 27 March 2008; accepted 6 May 2008)

The identification of nonsense syllables that were lowpass filtered at 1.5 kHz was compared for subjects with normal hearing and subjects with mild-to-severe hearing loss at high frequencies but with normal or near-normal hearing at low frequencies. Absolute thresholds were mostly within the normal range (<20 dB hearing level) for both groups for frequencies below 1.5 kHz. Performance was assessed with intact speech, speech that had been processed to preserve only temporal envelope cues in a few frequency bands (E speech), and speech that had been processed to remove envelope cues as far as possible while preserving temporal fine structure cues, again in a few frequency bands (TFS speech). For the intact speech and E speech, the hearing-impaired subjects performed slightly more poorly than the normal-hearing subjects, but this effect was significant only for the intact speech. For the TFS speech, the hearing-impaired subjects performed significantly more poorly than the normal-hearing subjects, with 12 out of 16 of the former performing at chance. The results indicate that, for people with hearing loss at medium to high frequencies, the processing of the TFS of speech can be degraded for frequencies where absolute thresholds are within the normal range.
 [DOI: 10.1121/1.2939125]

PACS number(s): 43.71.Ky, 43.71.Rt, 43.66.Sr [MW] Pages: 27–30

I. INTRODUCTION

Within the cochlea, broadband sounds like speech are decomposed by the "auditory filters" (Moore, 2003) into a series of narrowband signals, each evoked at a different place on the basilar membrane. Each signal can be considered as a "carrier"—the temporal fine structure (TFS), which is determined by the dominant frequencies in the signal that fall close to the center frequency of the band—and a temporal envelope (E), which corresponds to the relatively slow fluctuations in amplitude superimposed on the carrier. Information about speech sounds may be carried both by TFS and by E cues (Shannon *et al.*, 1995; Gilbert and Lorenzi, 2006; Hopkins *et al.*, 2008; Gilbert *et al.*, 2007).

Previously, we explored the role of TFS and E cues in speech perception for normal-hearing (NH) subjects and for young and older subjects with "flat" moderate hearing loss (Lorenzi *et al.*, 2006). Speech was filtered into 16 adjacent 0.35-oct-wide frequency bands spanning the range 0.08–8.02 kHz. The signal in each band was processed to preserve only E cues (by replacing the TFS with a sinewave at the center frequency of the band) or only TFS cues (by dividing the bandpass filtered signal by the envelope). The processed signals from the bands were then recombined. All groups scored 90–95% correct with E cues alone. After training, the NH group achieved about 90% correct with TFS cues alone, but both hearing-impaired (HI) groups performed very poorly, scoring between chance and 35% correct. The scores with TFS speech were highly correlated with the limited ability to take advantage of temporal dips when trying to understand unprocessed speech presented in a fluctuating background noise. We concluded that, regardless of age, hearing impairment adversely affects the ability to use TFS cues, and that such cues are important for the ability to listen in the dips of a competing background.

It is widely assumed that hearing impairment acts independently in different frequency regions, and that a "normal" audiometric threshold at a specific frequency, which is defined clinically as a threshold better than 20 dB hearing level (HL), implies normal auditory processing at that frequency. However, there is some evidence to the contrary. For example the ability to detect interaural time differences in low-frequency tones can be adversely affected by high-frequency cochlear hearing loss (Smoski and Trahiotis, 1986); for a review, see Moore (2007). Also, the ability to identify low-pass filtered speech can be adversely affected by high-frequency hearing loss (Horwitz *et al.*, 2002).

[a)]Author to whom correspondence should be addressed. Electronic mail: lorenzi@psycho.univ-paris5.fr

TABLE I. Age (years) and audiometric thresholds dB HL for each HI subject.

		Frequency (kHz)								
Listener	Age	0.25	0.5	0.75	1	1.5	2	3	4	6
HI1	67	20	25	30	30	30	35	55	55	55
HI2	53	15	10	15	10	15	20	35	50	35
HI3	46	30	30	35	30	35	35	45	60	60
HI4	39	20	15	20	25	25	30	45	40	85
HI5	58	30	35	35	35	40	40	35	35	30
HI6	54	25	30	30	30	35	35	40	50	75
HI7	38	20	15	15	15	25	50	55	50	50
HI8	61	15	15	15	20	25	20	30	30	55
HI9	63	15	15	10	10	20	25	45	55	85
HI10	61	15	15	15	20	20	20	50	55	55
HI11	60	15	15	20	20	25	30	40	55	35
HI12	68	10	5	10	15	15	35	65	70	70
HI13	59	15	10	15	20	20	45	55	70	100
HI14	57	15	15	20	15	15	35	50	50	65
HI15	55	5	5	10	10	20	30	35	45	55
HI16	56	15	15	10	15	15	45	55	60	70

Here we present further evidence that the processing of speech by HI listeners can be degraded when the speech is filtered so as to restrict its spectrum to the range where absolute thresholds are within normal limits. We also show that the degradation is greater for the processing of TFS cues than for the processing of E cues in speech.

II. METHOD

A. Subjects

Twelve subjects with normal hearing (NH1–NH12, age: 21–46 years, mean: 29 years) and 16 subjects with high-frequency cochlear hearing loss HI1–HI16 were tested. All NH subjects had audiometric thresholds between 0 and 20 dB HL at audiometric frequencies between 0.25 and 4 kHz at the test (right) ear. The ages of the HI subjects and the audiograms for the test (right) ears are shown in Table I. The HI subjects were chosen to have normal, or near-normal low-frequency hearing. However, the HI subjects had slightly poorer low-frequency hearing than the NH subjects. An analysis of variance (ANOVA) of the audiometric thresholds for frequencies below 2 kHz (0.25–1.5 kHz) with factors listener group (NH and HI), and audiometric frequency (0.25–1.5 kHz) showed significant main effects of group, $F(1,25)=26.734$, $p<0.0001$, and audiometric frequency, $F(4,100)=8.92$, $p<0.0001$, and a significant interaction between group and audiometric frequency, $F(4,100)=3.477$, $p<0.05$. The average threshold across all frequencies from 0.25 to 1.5 kHz (the PTA) was 6.8 dB HL for the NH group and 19.4 dB HL for the HI group. Despite this, 11 of the HI subjects (HI2 and HI7–HI16) had PTAs better than 20 dB HL, which is usually taken as the boundary between normal and impaired hearing.

B. Stimuli

Subjects were required to identify VCV nonsense syllables that were either intact, or were processed to preserve mainly envelope cues (E speech) or temporal fine structure cues (TFS speech). One set of 48 unprocessed VCV stimuli was recorded. There were three utterances of each of the 16 /aCa/ syllables (C=/p, t, k, b, d, g, f, s, ʃ, v, z, ʒ, m, n, r, l/) read by a French female speaker and digitized at a 44.1 kHz sampling frequency. Signals were digitized via a 16 bit analog-to-digital converter at a 44.1 kHz sampling frequency.

The processing was the same as described by Lorenzi *et al.* (2006) and by Gilbert and Lorenzi (2006). For all conditions (intact, E and TFS), each VCV signal was initially bandpass filtered using zero-phase, 3rd-order Butterworth filters into 16 adjacent, 0.35-octave-wide frequency bands spanning the range 0.08–8.02 kHz. This bandwidth corresponds to approximately 2 ERB_N (Glasberg and Moore, 1990) for midrange frequencies, but to about 1 ERB_N for low frequencies (Gilbert and Lorenzi, 2006). To create intact speech, all the band signals were summed. For the E and TFS conditions, the Hilbert transform was applied to the signal in each band to decompose the signal into its E (modulus of the analytic signal) and TFS (cosine of the argument of the analytic signal).

For the E condition, the envelope in each band was downsampled to 0.441 kHz, and then passed through a zero-phase, 6th-order Butterworth lowpass filter (cutoff frequency=64 Hz). The resulting envelopes were upsampled back to 44.1 kHz and used to amplitude modulate sinusoidal carriers with frequencies at the arithmetic center frequencies of the bands, and with random starting phase. The modulated signals were summed over the 16 frequency bands to form the E signal. For the TFS condition, the TFS for each band was multiplied by the root mean square (rms) power of the VCV in that band. This was done to avoid amplifying frequency bands carrying little or no speech information (e.g., bands dominated by recording noise). The "power-weighted" TFS signals were summed over the 16 frequency bands to form the TFS signal.

Following processing, the stimuli were lowpass filtered at 1.5 kHz (−72 dB/oct slope, Butterworth filter), so as to

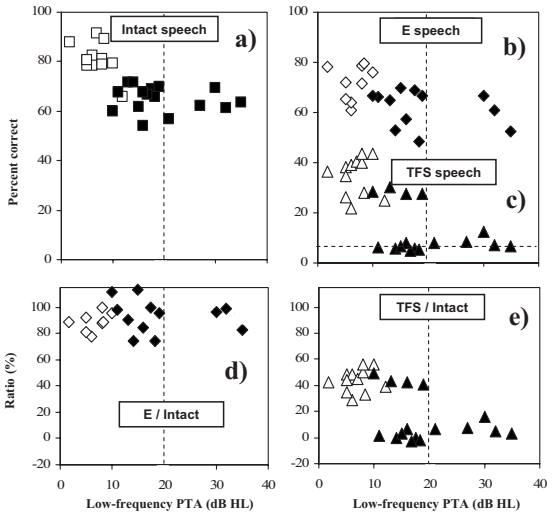


FIG. 1. Individual identification scores as a function of the PTA, for NH (open symbols), and HI subjects (filled symbols) for: (a) intact speech (squares); (b) E speech (diamonds); (c) TFS speech (triangles). For the ratio scores shown in panels (d) and (e), scores for intact, E and TFS speech were first transformed to RAU. The RAU-transformed scores for E and TFS speech were expressed relative to RAU-transformed scores for intact speech, and expressed as a percentage. Vertical dashed lines indicate the clinically accepted boundary between normal and impaired hearing.

restrict the spectrum to a frequency range where audiometric thresholds were <20 dB HL for 11 of the HI subjects and near normal for the rest. An advantage of this approach is that it eliminates possible confounding effects of differences in the widths of the auditory filters between groups; although auditory filters tend to broaden with increasing audiometric threshold, the broadening is typically small for hearing losses up to 30 dB (Moore, 2007). One problem with TFS speech is that the envelope in different frequency bands may be partly reconstructed in the auditory system (Gilbert and Lorenzi, 2006). The sharper the auditory filters, the greater is the degree of reconstruction. Given that the auditory filter widths were likely to be similar for the NH and HI subjects, it is very unlikely that any differences in sensitivity to TFS between the two groups could be attributed to differences in the degree of envelope reconstruction.

For the NH subjects, stimuli were presented at 75 dBA. For the HI subjects, a half-gain rule was applied, i.e., a gain equal to half of the average PTA was used. The resulting level was between 78 and 88 dBA, depending on the subject. The level was sufficient to ensure that all frequency components below 1.5 kHz were fully audible. Stimuli were presented using Sennheiser HD580 earphones. All subjects were trained for six, 5 min sessions, and performance was assessed over an additional four sessions. Subjects were tested first using intact and TFS speech, and were later recalled for the tests using E speech. Nine of the 12 NH subjects and 12 of the 16 HI subjects were able to return for the tests with E speech.

III. RESULTS

Figure 1 shows scores for the intact speech (a, squares) and the E and TFS speech (b, diamonds, and c, triangles, respectively), plotted as a function of the PTA. Scores for the NH and HI subjects are shown by open and filled symbols, respectively. A within-subjects ANOVA was conducted on the RAU-transformed scores with factors group (NH or HI) and processing (intact, E or TFS) (with missing values for the subjects who were not able to return for the tests with E speech). The main effect of group was significant: $F(1,19)=46.30$, $p<0.001$. The main effect of processing was also significant: $F(2,38)=272.5$, $p<0.001$. Finally, the interaction was significant: $F(2,38)=4.89$, $p=0.013$. *Post hoc* comparisons were based on the Tukey test.

For the intact speech [Fig. 1(a)], the NH subjects typically performed somewhat better than the HI subjects and the difference in mean scores was statistically significant ($p=0.02$). Similar moderate deficits in understanding lowpass filtered speech have been reported previously for HI subjects with good low-frequency hearing (Horwitz *et al.*, 2002). For the E speech [Fig. 1(b)], there was a slight tendency for the HI subjects to perform more poorly than the NH subjects, but the difference in mean scores was not statistically significant. This is consistent with previous work showing that HI subjects do not show a deficit in the processing of envelope cues in lowpass filtered speech (Turner *et al.*, 1995), although they do show a deficit for broadband speech processed using a large number of channels (Baskent, 2006), perhaps as a consequence of reduced frequency selectivity.

For the TFS speech [Fig. 1(c)], the NH subjects scored between 20 and 50%. These scores are much lower than those obtained when broadband TFS speech is used (Lorenzi *et al.*, 2006). This suggests that substantial TFS information can be used by the auditory system for frequencies above 1.5 kHz. Four of the HI subjects (HI2, HI11, HI13, and HI15) achieved scores of 25–30%, but most scored close to the chance level (6.25%), indicated by the horizontal dashed line. The difference in mean scores for the two groups was statistically significant: $p=0.0003$. These results indicate that many of the HI subjects had little or no ability to use TFS cues in lowpass filtered speech, despite their near-normal absolute thresholds.

Figure 1(d) shows, for each subject, the RAU-transformed scores for E speech expressed as a percentage of RAU-transformed scores for intact speech. These relative scores are somewhat more scattered for the HI than for the NH subjects, but most of the scores for both groups fall close to but a little below 100%. The means were 88% and 93% for the NH and HI groups, respectively. This indicates that the most salient cues for identification were the E cues, and removing TFS cues had only a small effect. Figure 1(e) shows, for each subject, the RAU-transformed scores for TFS speech expressed as a percentage of the RAU-

TABLE II. Mean identification scores across subjects (in percent correct, with standard deviation in parentheses) for each speech processing condition (Intact, E- and TFS-speech) and each filter slope/masker condition.

	Filter slope and masker	
Processing condition	−72 dB/oct (without masker)	−216 dB/oct (with masker)
Intact speech	85(4)	79(2)
E-speech	80(5)	68(5)
TFS-speech	46(9)	43(7)

transformed scores for intact speech. These relative scores for the NH subjects were much lower than the relative scores shown in panel d, confirming the greater salience of E cues. As expected, the relative scores for most of the HI subjects were very low, consistent with a greatly reduced ability to use TFS cues. The means were 44% and 14% for the NH and HI groups, respectively.

The scores of the HI subjects with TFS speech were not related to age ($r=0.016$, not significant: NS). They were also not significantly related to the PTA ($r=-0.31$, NS), although it is noteworthy that the four subjects who achieved scores for TFS speech within the normal range all had PTA values below 20 dB. There were seven HI subjects with PTA values below 20 dB who scored close to the chance level with TFS speech. The four subjects who scored within the normal range for TFS speech did not differ from the other HI subjects in terms of their scores for intact speech or E speech.

IV. SUPPLEMENTARY EXPERIMENT: THE ROLE OF THE TRANSITION BAND

Although our stimuli were lowpass filtered at 1.5 kHz, subjects may have been able to use information from the "transition band" extending upwards from 1.5 kHz, where speech information was attenuated, but not removed completely. If the NH subjects could use this information more effectively than the HI subjects, this could account for the difference between the groups. This possibility was tested by running a new group of eight NH subjects (absolute thresholds below 20 dB HL at audiometric frequencies between 0.250 and 6 kHz) using conditions similar to those of experiment 1, and an additional condition in which the transition slope was steeper and a highpass-filtered noise was added to mask the transition band. The subject's ages ranged from 20 to 48 years.

The intact, E- and TFS-coded VCV stimuli were either generated exactly as for experiment 1 (with a lowpass filter slope of −72 dB/oct), or were lowpass filtered using a zero-phase, Butterworth lowpass filter with a cutoff frequency of 1.5 kHz and slope of −216 dB/oct and presented together with a speech-shaped noise masker that was highpass filtered (zero-phase, Butterworth filter with cutoff frequency of 1.5 kHz and slope of 108 dB/oct) at a signal-to-noise ratio of +12 dB (in terms of overall level). The subjects were initially trained with unprocessed, E and TFS broadband VCV stimuli, generated using a 16 band vocoder, until they achieved better than 70% correct for both E and TFS speech.

The procedure was identical to that for experiment 1. Mean results across listeners (based on 96 items per condition) are presented in Table II. A *t*-test with Bonferroni correction was used to estimate the significance of the differences in RAU scores across conditions. Preventing use of information from the transition band impaired speech identification for E-speech by 12 percent points ($p<0.05$). In contrast, the effects were smaller for intact speech (6 percent points, NS) and TFS speech (3 percent points, NS).

Overall, these data suggest that NH subjects made little or no use of TFS information from the transition band. This indicates that the difference between the NH and HI subjects in their ability to use TFS information, as found in experiment 1, cannot be accounted for by the ability of the former to use information from the transition band.

V. CONCLUSIONS

Subjects with impaired hearing at high frequencies but normal or near-normal hearing at low frequencies often show a complete inability to use TFS information at low frequencies for speech reception. TFS information in the auditory nerve is probably not completely absent, since subjects with audiograms similar to those of the present subjects have some ability to discriminate changes in interaural phase of the TFS at the two ears for low-frequency amplitude-modulated tones (Lacher-Fougère and Demany, 2005). It may be that problems in using TFS occur when the TFS is complex and time varying, as in our TFS speech. In any case, the results demonstrate that normal audiometric thresholds at a specific frequency can be associated with strong abnormalities in the processing of TFS cues in speech at that frequency.

ACKNOWLEDGMENTS

The work of B.C.J.M. was supported by the MRC (UK). The authors thank Hedwig Gockel and two anonymous reviewers for helpful comments.

Baskent, D. (**2006**). "Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels," J. Acoust. Soc. Am. **120**, 2908–2925.

Gilbert, G., and Lorenzi, C. (**2006**). "The ability of listeners to use recovered envelope cues from speech fine structure," J. Acoust. Soc. Am. **119**, 2438–2444.

Gilbert, G., Bergeras, I., Voillery, D., and Lorenzi, C. (**2007**). "Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues," J. Acoust. Soc. Am. **122**, 1336–1339.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

Horwitz, A. R., Dubno, J. R., and Ahlstrom, J. B. (**2002**). "Recognition of low-pass-filtered consonants in noise with normal and impaired high-frequency hearing," J. Acoust. Soc. Am. **111**, 409–416.

Lacher-Fougère, S., and Demany, L. (**2005**). "Consequences of cochlear damage for the detection of interaural phase differences," J. Acoust. Soc. Am. **118**, 2519–2526.

Lorenzi, C., Gilbert, G., Carn, C., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).

Moore, B. C. J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*, 2nd ed. (Wiley, Chichester).

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Smoski, W. J., and Trahiotis, C. (**1986**). "Discrimination of interaural temporal disparities by normal-hearing listeners and listeners with high-frequency sensorineural hearing loss," J. Acoust. Soc. Am. **79**, 1541–1547.

Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.

Sound transmission loss of rectangular and slit-shaped apertures: Experimental results and correlation with a modal model

Nicolas Trompette and Jean-Louis Barbry
Laboratoire Réduction du Bruit au Travail, Institut National de Recherche et de Sécurité (INRS), Avenue de Bourgogne, 54500 Vandoeuvre, France

Franck Sgard and Hugues Nelisse
Institut de Recherche Robert-Sauvé en Santé et Sécurité au Travail (IRSST), 505 Boulevard de Maisonneuve Ouest, Montréal, Québec H3A 3C2, Canada

(Received 28 May 2008; revised 10 September 2008; accepted 29 September 2008)

Among noise control techniques, enclosures are widely used. It is known that enclosure acoustic efficiency is strongly influenced by the presence of openings or leaks. Modeling of diffuse field sound transmission loss (TL) of apertures and slits is therefore critical when the enclosure acoustic performance characteristics need to be predicted with confidence either for design or for modifying existing enclosures. Recently, a general model for diffuse field sound TL of rectangular and circular apertures has been developed and validated with respect to existing analytical or numerical models. This paper presents an experimental validation of this new model. The aim was to develop a simple, reliable tool for predicting enclosure insertion loss using statistical energy analysis. Twelve out of the 15 test configurations were found to be reliable and were compared with theoretical models, which in fact correlate closely (without adjustment) with the experimental work.
 [DOI: 10.1121/1.3003084]

PACS number(s): 43.20.El, 43.55.Ti, 43.55.Fw, 43.55.Rg [KA] Pages: 31–41

I. INTRODUCTION

Among noise control techniques, enclosures are widely used. An enclosure must be absolutely airtight to ensure maximum acoustic efficiency. Even a small aperture (opening or leak) will significantly affect the acoustic efficiency of an enclosure.[1] Miller and Montone[2] revealed the effect of apertures on enclosure acoustic efficiency and their results are illustrated in the well-known graph (Fig. 1). Enclosure openings are usually designed for material flow, while leaks (slits or holes) result from poor assembly or engineering choices, such as using sliding gates. When the opening size is smaller than the acoustic wavelength, it is typically referred to as a leak; the term "opening" is generally used for larger opening sizes. Apertures must be prevented, minimized, or lined with sound absorption systems (sound traps and silencers) to ensure a good enclosure acoustic performance. In practice, however, opening influence and leak prevention are often overlooked. Quantification of opening and leak impact would be helpful when existing enclosure efficiency is insufficient and requires appropriate modification. For this reason, modeling of diffuse field sound transmission loss (TL) of apertures (openings, slits, and holes) can be considered a key process for correctly predicting enclosure acoustic performance at design stage and for assisting noise prevention specialists in performing appropriate design modifications to existing enclosures. This work forms part of a research project aimed at developing simple, efficient design tools for machinery enclosures based on statistical energy analysis (SEA). System responses are often analyzed in third-octave-bands, when applying SEA, whose parameters include nonresonant coupling loss factors based on evaluating sound TL between two subsystems. Therefore, the focus here is made on measuring and calculating third-octave-band sound TLs. Sgard *et al.*[3] recently undertook a comprehensive review of existing models and proposed a general, efficient, and rigorous numerical method based on a modal approach for predicting diffuse field TL of apertures with rectangular and circular cross sections. These authors also provided numerical results for this indicator and its relationship with the normal incidence case for various geometrical configurations. They concluded that the conventional third-octave normal incidence TL could be substituted for the diffuse field TL if adjusted by a correction factor.

Only comparisons between existing models and numerical results were provided in the paper of Sgard *et al.*[3] The present paper is intended to complement the work of Sgard *et al.*[3] by performing new third-octave-band experimental validation tests for rectangular openings and slits and to compare experimental results with the model of Sgard *et al.*[3] and other analytical models. It should be noted that while circular apertures have been presented in Ref. 3, these are not considered here for conciseness, but their testing is indeed planned for the future. The model agrees closely with experimental data for rectangular openings; as the paper subsequently shows, it is expected to also compare well for circular openings. This paper firstly provides a review of the literature to reveal the lack of available experimental data concerning rectangular opening and slit models. Experimental setup and investigated configurations are described in a second part. Third-octave-band experimental results are then presented and compared with existing values. Theoretical results derived from the model of Sgard *et al.*[3] for openings

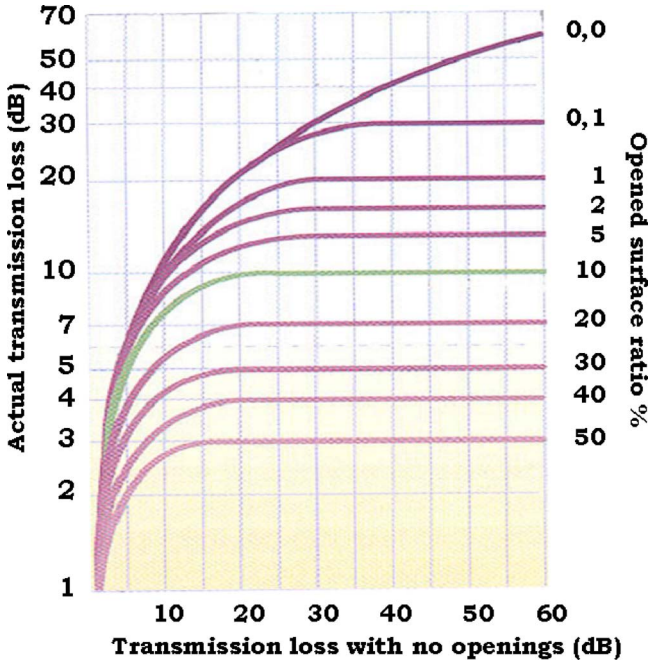


FIG. 1. (Color online) Maximum insertion loss of an enclosure as a function of the opened surface taken from Ref. 2.

and slits and from the model of Mechel[4] for slits, are subsequently compared with the experimental data. The main findings of this work are consolidated in our conclusion.

II. LITERATURE REVIEW

The relevant literature does not provide many diffuse field TL test results for apertures. Gomperts conducted a comprehensive study using two reverberant chambers interconnected by an 8 cm deep aperture.[5] Reproducibility, geometrical shape, and aperture location were all closely examined in this study. The aperture cross-sectional area varied fairly widely between 2 and 100 cm^2, and both rectangular and circular apertures were tested. Measurements were performed in third-octave bands from 630 Hz to 2 kHz. Measurement uncertainties were acceptable, except for the 1.6 kHz third-octave band, in which resonance of the first longitudinal mode occurs. These experiments were perfectly controlled and will be further investigated. Using the same test setup, Gomperts and Kihlman extended the previous survey to 2 m wide slits[6] with depths ranging from 1.5 to 100 mm and widths ranging from 0.5 to 8 mm. The results were less convincing than those of previous tests. Nevertheless, these 2 m wide slit results will be compared with measurements. Wilson and Soroka[7] also correlated their aperture model with experimental data. Several tests were conducted, but, unfortunately, only four are described in the authors' paper. The apertures studied were of circular cross section with 2 and 4 in. diameters and 3 and 12 in depths. The results of these tests are rather surprising: the TL frequency response appears correct, but the amplitudes are very high. This was noticed by the authors, who pointed out the small degree of damping in the recording system. These experiments must therefore be viewed with caution. Investigations were pursued by Sauter and Soroka,[8] applying exactly the same procedure to rectangular apertures. Twelve apertures, divided into four sets of three rectangular steel tubes, were tested. All tubes were 30 cm deep. The cross-sectional dimensions of each set were $\pi/4$, π, and 4π $in.^2$. The authors introduced an aspect ratio M (long side to short side ratio) and a geometrical parameter $\beta=h/(S^*\pi)^{1/2}$ (h=depth, S=cross-sectional area) for each set. Results were given for only five test configurations, corresponding to different combinations of M and β. Experimental data agreed closely with the models and can be used for further comparisons. More recently, Gibbs and Balilah[9] applied an impulse technique to measuring sound TL through circular holes. However, this technique involves meeting a number of strict conditions governing time signal generation and acquisition, and, more importantly, it is restricted to plane wave excitation. Furue[10] undertook an experimental study involving TL measurements of an aperture between a reverberant chamber and an anechoic room. However, this study focused on diffraction theory, so only high frequencies were considered. Oldham and Zhao[11] measured the circular aperture and slit TL of circular apertures and slits in a reverberant field source and compared their results with those of Wilson and Soroka's[7] and Gompert's[5] analytical models. These authors used the same experimental setup that used by Furue but took measurements in the receiving room with an intensity probe. Very small (4.5 and 10.2 mm diameters) circular apertures and 1 to 10 mm wide slits of different depths were tested. The narrow band results agree closely with the model predictions for the circular cross section case. Theoretical comparisons were less close for the slit case. The authors highlighted a discrepancy between theoretical and measured TLs at the fundamental resonance frequency for both circular cross section and slit cases. The deeper the aperture and the smaller the width or radius, the larger this discrepancy. This phenomenon was attributed to viscosity effects inside the aperture. It is interesting to note that the size of the tested circular apertures is unrealistic for enclosures. In addition, all results start at 1 kHz (small source room of 3.3 m^3) and are given in narrow bands, so they are not used in the present study. Finally, Chen[12] used a similar experimental setup for validating an infinitely long slit model similar to Mechel's model.[4] In this study, slit depth and length were 0.3 and 1.2 m, respectively (i.e., infinite width) for three widths: 7.8, 15, and 66 mm. Measurements were compared with both proposed normal incidence and diffuse field models for frequencies ranging from 100 to 1250 Hz. Differences between calculations and experiments were found to be smaller than 4 dB over the entire frequency range. According to Chen, the results seem reliable above 200 Hz. However, it is worth noting that the use of the normal incidence model leads to smaller discrepancies with respect to experimental data than the diffuse field theory. These results would therefore be questionable.

Table I consolidates papers available in the literature, which provide the reader with measurements that are usable for both openings and slits. Authors names, aperture shape, dimensions, and frequency range of interest are provided in different columns of this table. The table shows that there are a limited number of experimental data in the literature for

TABLE I. Summary of measurement data available in literature and usable.

Openings					
Author	Shape	Length or radius (cm)	Width (cm)	Depth (cm)	Frequency range (third-octave band)
Gomperts (Ref. 5)	Circular	0.2–100	⋯	11	630 Hz–2 kHz
	Rectangular	0.2–100	0.2–100	11	630 Hz–2 kHz
Wilson and Soroka (Ref. 7)	Circular	2.5/5	⋯	7.5/30	200 Hz–10 kHz
Sauter and Soroka (Ref. 8)	Rectangular β=6, M=2,4,8	3.2/4.5/6.4	25.6/18/12.8	30.48	200 Hz–6.3 kHz
	Rectangular β=6,12,24, M=8	3.2/1.6/0.8	25.6/12.8/6.4	30.48	200 Hz–6.3 kHz
Oldham and Zhao (Ref. 11)	Circular	0.45/1.02/2.04/3.98	⋯	3.6/7.26/14.4	~1 kHz–12 kHz

Slits					
Author	Shape	Length (m)	Width (mm)	Depth (cm)	Frequency range (third-octave band)
Gomperts and Kihlman (Ref. 6)	Rectangular	2	0.5/1/2/4	1.5/20/50/100	100 Hz–8 kHz
	Rectangular	0.2/0.5/1/2	4.5	50	100 Hz–8 kHz
Chen (Ref. 12)	Rectangular	1.2	0.78/1.5/6.6	30	100 Hz–5 kHz
Oldham and Zhao (Ref. 11)	Rectangular	Not specified	1/1.5/3/6	5.08/7.62/15.24	~1 kHz–12 kHz

validating aperture and slit models. New experiments therefore appeared necessary for validating the model proposed in Ref. 3. Some of these validation tests are similar to those contained in the literature. Apertures sizes were set according to dimensions tested by Sauter and Soroka.[8] Aperture depth was kept the same (30 cm). Two out of six configurations are effectively identical. Slits are identical to those used by Gomperts and Kihlman,[6] except for the 1 and 4 mm widths, which have not been tested. The dual purpose of this paper is, in fact, to reproduce data found in the literature and increase the number of available test results.

III. EXPERIMENTAL PROCEDURE

The purpose of the measurements was to obtain the diffuse field sound TL of openings and slits. Apertures were therefore located between an emission chamber and a receiving chamber.

The emission chamber was a 8 m^3 reverberant room, which was mounted on isolators to prevent flanking transmissions. The emission chamber was separated from the receiving chamber by a wall, part of which was concrete and part of which was steel plate (see the wall in Fig. 2). The two chambers were interconnected by a 0.41×0.55 m^2 aperture located in the center of the steel plate wall section. The diffuse sound field was produced by four noise sources, i.e. three Omni–Power sound sources and a compression chamber loudspeaker. Sound source generators were all uncorrelated, and the resulting noise spectrum was equalized, so that the acoustic energy was equally distributed in the third-octave bands. Sound pressure was measured at four different points in the chamber, which allowed its uniform spatial distribution to be verified (see Fig. 3). Incident acoustic energy was deduced from the sound pressure level averaged for the four microphone positions. The number of microphones was chosen based on Ref. 7.

The receiving room was an anechoic chamber. Transmitted acoustic energy was deduced from sound intensity point mapping measurements (see Fig. 2). Meshing was dependent on the size of the test aperture. Intensity measurement point spacing was coarser for large apertures and finer for slits. For large openings, 100 uniformly distributed measurement points were used with a 10 cm spacing in both aperture length and width directions. Measurement points were distributed on two different grids for the smallest slits. A fine grid for measurement points closest to the slit center was

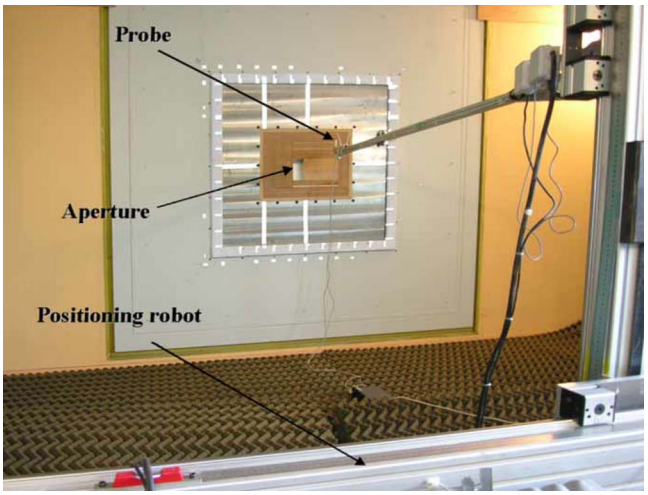


FIG. 2. (Color online) Receiving chamber with the intensity scanning device.

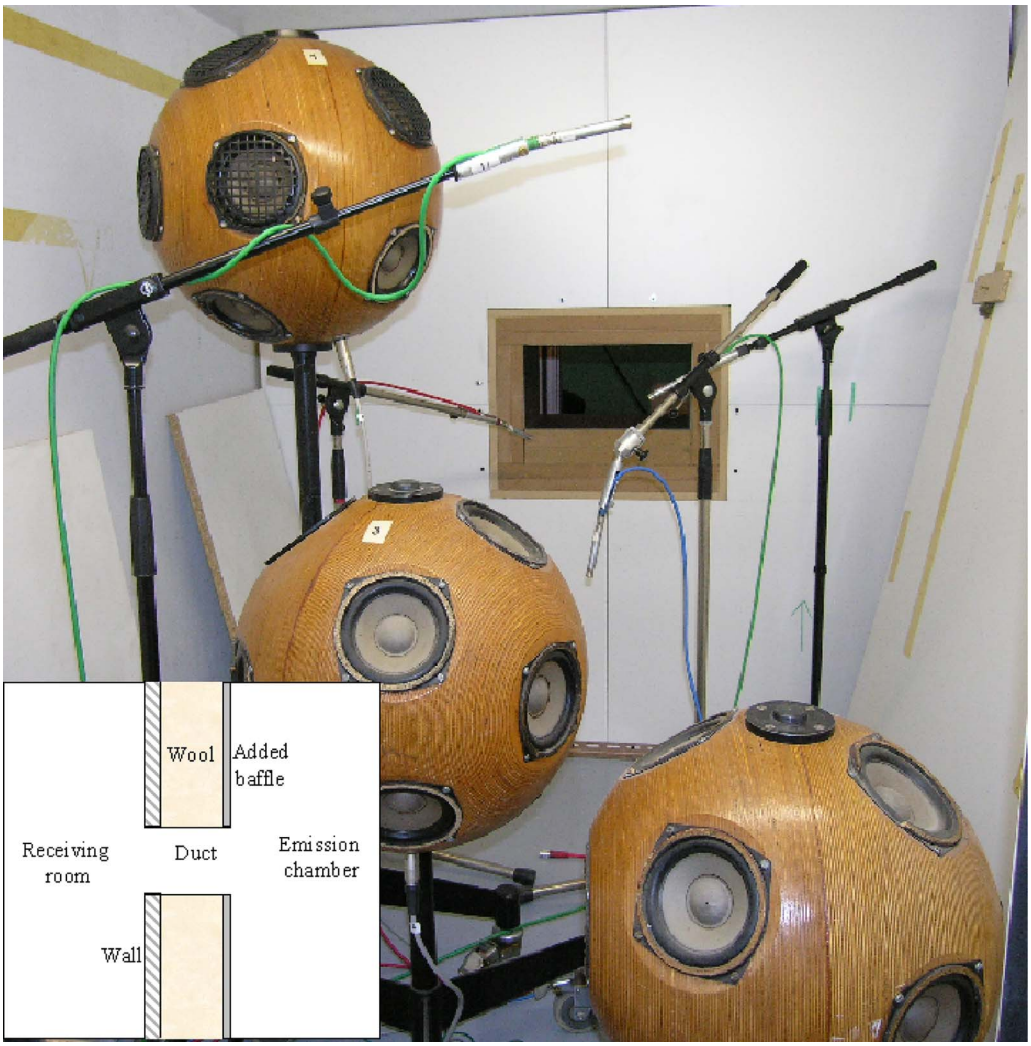


FIG. 3. (Color online) Emission chamber with three omni power sound sources and four microphone positions.

complemented by a coarser grid for more distant measurement points. 70 measurement points were used in the finer grids with a 4 cm spacing in the aperture length direction and a 2 cm spacing in the aperture width direction. 58 measurement points were used for the coarsest grid with a 8 cm spacing in the aperture length direction and a 6 cm spacing in the aperture width direction. At least 100 measurement points were used irrespective of the configuration. In the reverberant room, the sound power incident on the aperture (or the slit) W_i is deduced from the average pressure using the relation

$$W_i = \frac{\langle p^2 \rangle}{4\rho c} S, \tag{1}$$

where

- $\langle p^2 \rangle$ is the mean square pressure averaged for the four microphone positions
- ρ and c are the air density and sound speed, respectively
- S is the aperture cross-sectional area

In the anechoic chamber, the sound power W_t transmitted by the aperture was calculated from the sound intensity field using the relation

$$W_t = \sum_{j=1}^{N} I_j \Delta S_j, \tag{2}$$

where

- I_j is the sound intensity at point j
- ΔS_j is the surface area element around point j
- N is the number of the measurement grid points

The TL of the aperture is defined as

$$\mathrm{TL} = 10 \log \frac{W_i}{W_t}. \tag{3}$$

Measurements are valid as long as the acoustic field can be considered diffuse and uniform in the emission chamber. Several investigations have been performed to evaluate the frequency validity range. The Schröeder frequency[13] was calculated from the measured reverberation time and was found to be 705 Hz. Subsequently, an array of eight microphones spaced at 160 mm was used to check acoustic field uniformity (by comparing third-octave-band levels). The field was found to be uniform above 630 Hz. Differences in levels between microphones were smaller than ±2 dB in the

TABLE II. Geometrical characteristics of test apertures.

Aspect ratio M b/a	Geometrical parameter β $\beta=h/(S^*\pi)^{1/2}$	Width a (mm)	Height b (mm)	Depth h (cm)	Third octaves
2	2	190	380	30	500 Hz–6.3 kHz
1	2	270	270	30	500 Hz–6.3 kHz
2	6	60	120	30	500 Hz–6.3 kHz
1	6	90	90	30	500 Hz–6.3 kHz
8	6	30	240	30	500 Hz–6.3 kHz
1	12	45	45	30	500 Hz–6.3 kHz

250–630 Hz range. The coherence between a microphone positioned near the aperture and the microphone array was also examined. The field, in fact, became incoherent between 250 and 315 Hz, depending on the microphone considered. These results indicate that measurements cannot be representative of diffuse field conditions below 250 Hz. However, these conditions are found to be acceptable in the frequency range (250–630 Hz) and good above 630 Hz.

The receiving room volume was 54 m^3. It was lined with 20 cm thick mineral wool. Its absorption reaches 1 at 400 Hz, and it can therefore be considered anechoic above this frequency. In addition, measurements were taken using the intensity method. Standard ISO 9614-2 states that the intensity method can be extended down to 100 Hz with a 12 mm microphone spacing.

The six openings were formed by 30 cm deep rectangular steel ducts. Table II shows the different cross-sectional areas of these ducts. The apertures needed to be baffled to comply with the model hypothesis. The duct depth exceeded the thickness of the separating wall between the emission and the receiving rooms, so an additional baffle was added inside the reverberant chamber with the aperture located at its center (Fig. 3). The cavity between the baffle and the chamber separating wall was filled with mineral wool to eliminate the cavity mode effects. The steel duct was positioned inside a larger rectangular wooden frame (0.51 $\times$0.4 m^2) to facilitate aperture installation. The gap between the duct and the wooden frame was made up with additional pieces of wood, which could be easily adjusted to prevent leaks (see Fig. 4).

Slits were built using two steel plates (1.5 and 20 mm thick) or wooden (50 mm thick) panels inserted into the primary aperture and separated by spacers. This arrangement ensured a slit depth equal to the plate thickness and a slit width equal to the spacer width (Fig. 5). The plates were 50 cm long, and the slits could therefore be considered infinite in the studied frequency range. This assembly had the advantage of a simple design. However, unlike the mounted aperture, leaks could occur at fixing points, and plate or panel TL could influence results. Special care was therefore taken to prevent leaks around fixing points, and additional tests were conducted to evaluate the contribution of the plate or panel to the acoustic power generated by the system. Table III shows the nine configurations, which were tested.

In addition, three transmitted sound power measurements were taken with the aperture or slit closed. In these configurations, the transmitted sound power was effectively inherent to the separating wall and the flanking transmission, and it represented the part of the sound not transmitted by the opening or slit. Measurements were used only when this transmitted sound power measurement was lower (at least 6 dB) than that transmitted by the opening or slit.

IV. EXPERIMENTAL RESULTS

The presentation of results is divided into two sections: the first focusing on openings and the second on slits.

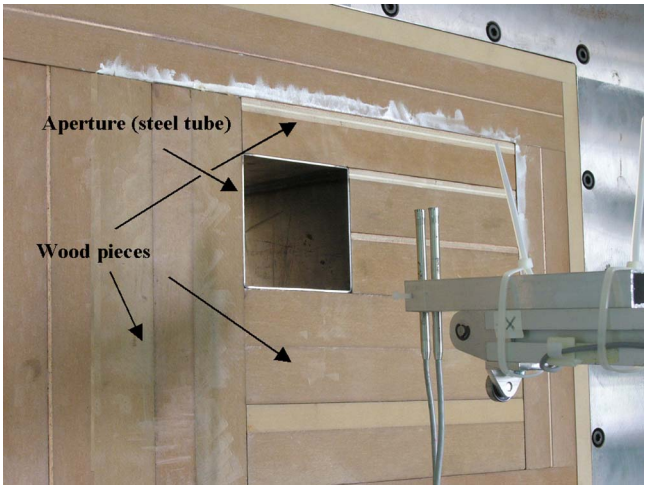


FIG. 4. (Color online) Aperture assembly, constituted of a steel tube and some pieces of wood filling a larger wood tunnel.

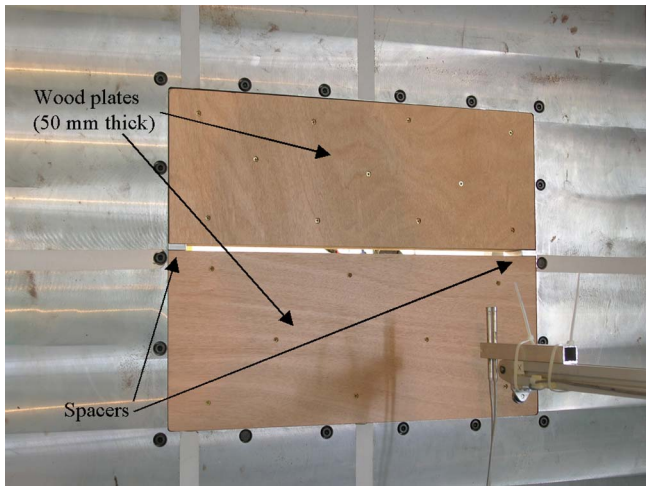


FIG. 5. (Color online) Leak made up by two plates separated by spacers.

TABLE III. Geometrical characteristics of test slits.

Length (mm)	Width (mm)	Depth (mm)	Third-octave central frequencies
500	0.5/2/8	1.5	500 Hz–6.3 kHz
500	0.5/2/8	20	500 Hz–6.3 kHz
500	0.5/2/8	50	500 Hz–6.3 kHz

A. Opening case

Literature-based data refer to rectangular apertures 18 and 30.48 cm deep.[5,8] One of the measurement aims was to compare experimental results with the literature. 30 cm deep openings (similar to experiments described in Ref. 8) only were studied because each opening depth change requires fabrication and assembly of a baffled transmission tunnel.

Reference 5 states that the cross-sectional shape has no practical consequences on sound transmission. Theoretically, this assumption is valid up to the aperture cutoff frequency. It was therefore interesting to test large apertures and to compare those with the same cross section but different shapes. The tested apertures thus have three different cross-sectional areas to (i) cover a wide range of aperture aspect ratios and dimensions, (ii) compare results for apertures of the same cross-sectional area, and (iii) compare results with those of Ref. 8.

Experimental results for the first two configurations in Table II ($M=2$, $\beta=2$) and ($M=1$, $\beta=2$), corresponding to the largest apertures, are shown in Fig. 6, and the corresponding TL values are listed in Table IV. Despite the small size of the emission chamber, the sound TL for these first two apertures exhibit a typical behavior above 250 Hz. Below the first longitudinal mode resonance, the TL is indeed positive and fairly large. It decreases rapidly at the mode resonance and becomes negative before finally tending to zero at high frequencies. The TL is zero above 800 Hz, as can be expected for the large aperture size. The differences between the two configurations do not exceed 2 dB, and the mean difference between the two configurations is 0.6 dB, which is insignificant.

Figure 6 also illustrates the experimental TL for the next three apertures in Table II, ($M=2$, $\beta=6$), ($M=1$, $\beta=6$), and ($M=8$, $\beta=6$). Experimental values are tabulated in Table IV. These apertures have the same cross-sectional area but different aspect ratios. The same conclusions apply. The TL maximum difference between the three configurations is 4 dB and is lower than 2 dB above 500 Hz. These aperture sizes are typical of those found in the industry. TL amplitude is high and can be negative, which underlines the importance of good aperture design.

The last aperture tested is the smallest. For conciseness, the results are also plotted in the top graph in Fig. 6 and tabulated in Table IV. This aperture is too small to be realistic, but this configuration will be the most relevant for comparing with the numerical models because the TL plot appears to have many dips and troughs. The reason for this is the aperture duct modal density, which is lower than for larger apertures.

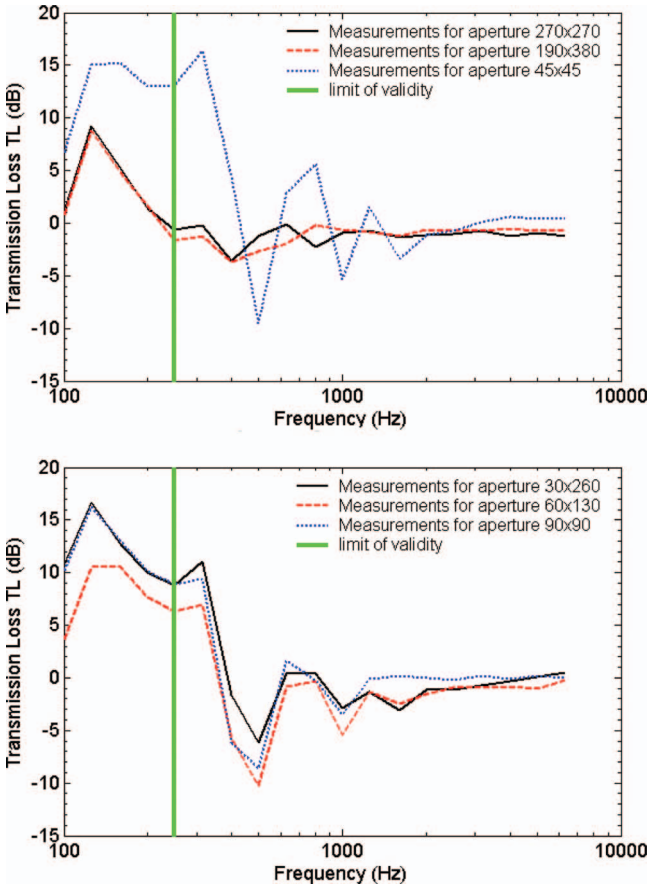


FIG. 6. TL of apertures—experimental results.

Intermediate size apertures (i.e., cross-sectional areas 30×260 and 60×130 mm^2) were also tested by Sauter and Soroka.[8] Figure 7 provides a comparison. It should be noted that there is an error in Ref. 8: the axis is, in fact, reversed in Fig. 5. Measurement correlation is very good, particularly if one considers the fact that the dimensions were not exactly the same and that the test setup was different.

B. Slit case

References 6, 11, and 12 contain some experimental data on slits. In this study, tests were only conducted on some of the configurations considered in Ref. 6. However, it is explained later that the data provided in Ref. 6 does not appear to be reliable.

Approximately half of the configurations tested by Gomperts and Kihlman[6] were retested (Table III). Figure 8 displays the results, which are also tabulated in Table V. It is important to emphasize that the transmitted acoustic power is very low, especially for the 50 mm deep slits. Flanking transmission and the transmission through the wall were both measured for these configurations. When the difference with respect to transmission through the slit was lower than 6 dB, an adjustment was made by subtracting the flanking transmission contribution to the radiated acoustic power from the overall radiated acoustic power. This is the case for the two 50 mm deep, 2 and 0.5 mm wide slits. Differences varied from 5 dB at 250 Hz to 15 dB at 6.3 kHz for the 2 mm wide slit and from 0.5 dB at 315 Hz to 6 dB at 6.3 kHz for the

TABLE IV. Experimental results for openings.

Aperture size	90×90	130×60	270×270	390×190	260×30	45×45
Frequency (Hz)	Transmission loss (dB)					
250	8.8	6.4	−0.6	−1.7	8.7	13.1
315	9.4	7.0	−0.2	−1.2	11.0	16.4
400	−6.1	−5.7	−3.6	−3.7	−1.6	4.4
500	−8.6	−10.2	−1.2	−2.6	−6.1	−9.6
630	1.6	−0.8	−0.1	−2.0	0.4	2.9
800	−0.2	−0.3	−2.2	−0.2	0.4	5.6
1000	−3.5	−5.4	−0.9	−0.6	−2.9	−5.3
1250	−0.1	−1.4	−0.8	−0.8	−1.3	1.5
1600	0.2	−2.5	−1.4	−1.2	−3.1	−3.4
2000	0.0	−1.5	−1.1	−0.6	−1.1	−1.1
2500	−0.2	−0.9	−1.0	−0.7	−1.2	−0.7
3150	0.2	−0.9	−0.7	−0.8	−0.7	0.1
4000	−0.1	−0.8	−1.2	−0.5	−0.3	0.6
5000	0.2	−1.1	−1.0	−0.8	0.1	0.4
6300	0.0	−0.2	−1.2	−0.7	0.5	0.4

0.5 mm wide slit. For these two slits, flanking transmissions prevailed below 250 and 315 Hz, respectively. The results for these two test configurations must therefore be accepted with caution. Comparisons between slits of the same depth show that the configuration differences agree with the width difference and the theoretical models (in particular, see Ref. 12).

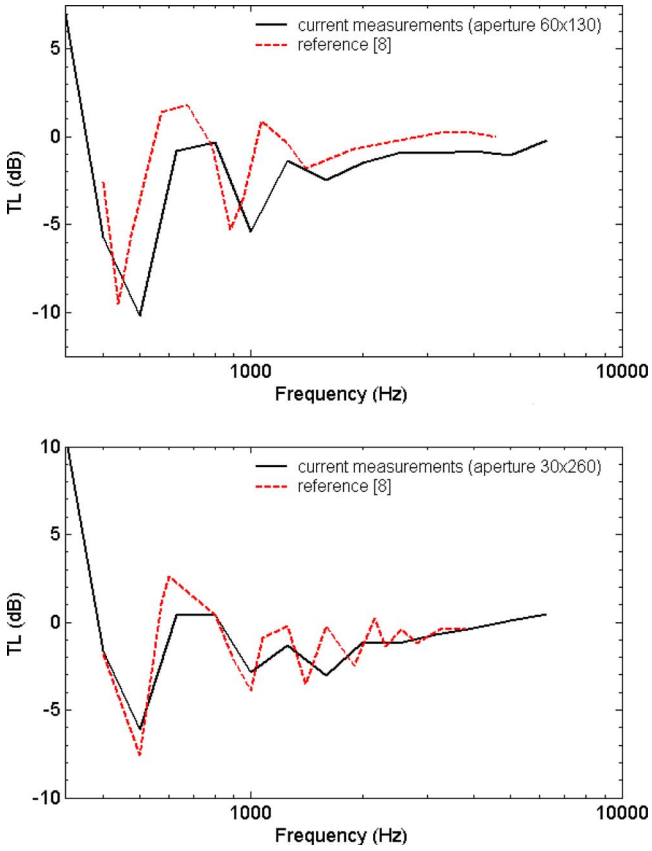


FIG. 7. (Color online) Comparison between the current measurements and data from Ref. 8 for rectangular apertures of dimensions $60\times130\times300$ mm^3 ($M=2$, $\beta=2$) (top) and $30\times260\times300$ mm^3 ($M=8$, $\beta=2$) (bottom).

This is particularly clear in the top graph in Fig. 8, which corresponds to the only configuration for which flanking transmission can be neglected. Based on Refs. 6 and 12,

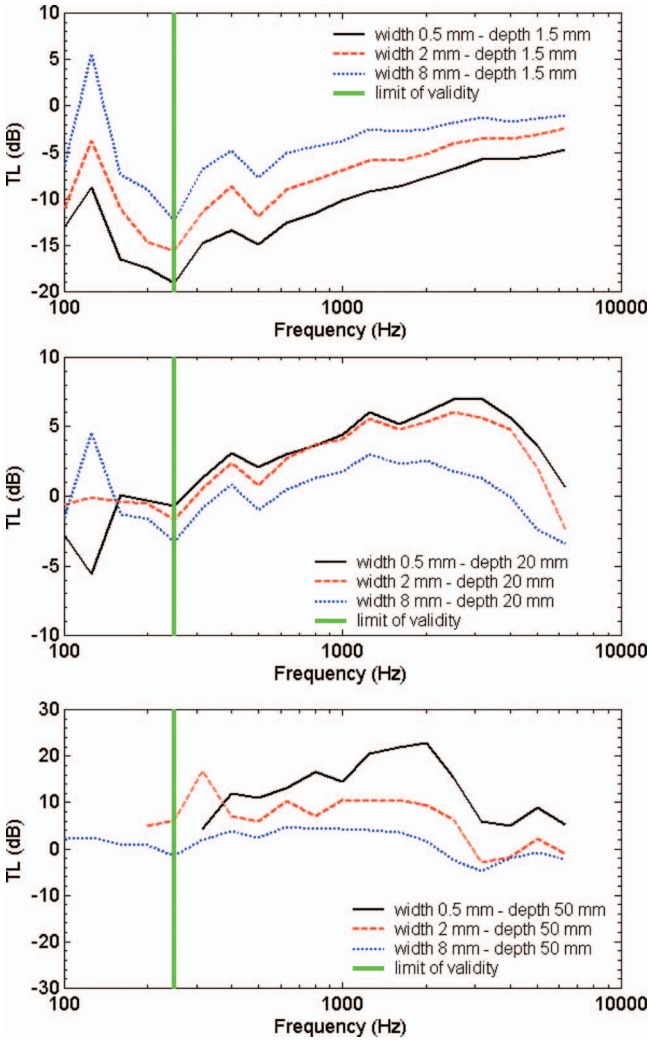


FIG. 8. TL of slits—experimental results.

TABLE V. Experimental results for slits.

Slit size	50×8 ×500 mm^3	50×2 ×500 mm^3	50×0.5 ×500 mm^3	20×8 ×500 mm^3	20×2 ×500 mm^3	20×0.5 ×500 mm^3	1.5×8 ×500 mm^3	1.5×2 ×500 mm^3	1.5×0.5 ×500 mm^3
Frequency (Hz)	Transmission loss (dB)								
250	−1.4	6.0	—	−3.3	−1.7	−0.7	−12.3	−15.6	−19.1
315	2.0	16.8	4.2	−0.8	0.5	1.4	−6.8	−11.5	−14.9
400	3.8	7.0	12.0	0.9	2.4	3.1	−4.9	−8.6	−13.4
500	2.4	5.9	10.9	−1.0	0.8	2.1	−7.7	−11.9	−15.0
630	4.7	10.4	13.8	0.4	2.7	3.0	−5.1	−9.0	−12.6
800	4.2	7.2	16.6	1.3	3.7	3.6	−4.4	−8.0	−11.5
1000	4.2	10.5	14.5	1.8	4.1	4.4	−3.8	−6.9	−10.2
1250	4.1	10.2	20.5	3.0	5.6	6.0	−2.5	−5.9	−9.3
1600	3.5	10.5	22.0	2.4	4.8	5.2	−2.8	−5.8	−8.6
2000	1.8	9.3	22.8	2.5	5.4	6.1	−2.5	−5.1	−7.7
2500	−2.5	6.4	15.5	1.8	6.1	7.0	−1.8	−4.1	−6.8
3150	−4.8	−2.8	5.9	1.3	5.7	7.0	−1.2	−3.6	−5.7
4000	−1.9	−1.6	4.9	−0.1	4.8	5.7	−1.6	−3.5	−5.8
5000	−0.9	2.2	8.9	−2.4	2.0	3.6	−1.3	−3.1	−5.4
6300	−2.3	−1.1	5.1	−3.4	−2.4	0.6	−1.0	−2.4	−4.7

the TL for this small depth (1.5 mm) should also show negative values and should behave smoothly as a function of frequency. This has been confirmed above 500 Hz, in accordance with the fact that the incident sound field can only be considered perfectly diffuse at 630 Hz. In the bottom graph in Fig. 8, the differences between 2 and 0.5 mm widths for a 500 mm deep slit are not as large as it would have expected, in view of the trend observed in the previous configuration (middle graph in Fig. 8) and in Ref. 12. This is probably due to the fact that the flanking transmission and the transmission through the separating wall were large for the 0.5 mm slit width configuration, particularly at third-octave center frequencies between 800 Hz and 1.25 kHz. For this depth, the TLs exhibit an erratic behavior due to both the adjustment referred to in the previous paragraph and the low transmitted power. It is important to highlight that this adjustment was not applied for the 8 mm slit width.

As mentioned above, slit dimensions were chosen to allow a comparison with the results published in Ref. 6. No such comparison is shown here because there were discrepancies up to 10 dB. It should be noted that this was already the case when Gomperts and Kihlman[6] failed to obtain satisfactory agreement between their experiments and the numerical models. The authors believe that there must be a problem in the experimental data presented in Ref. 6; this will be confirmed in the following section. While it is unfortunate that the results do not correlate, this does demonstrate how relevant it was to perform this new series of experiments

V. COMPARISONS WITH NUMERICAL MODELS

The model used for the comparison is the one proposed by Sgard *et al.* in Ref. 3, in which the authors reviewed all the main existing models for rectangular and circular apertures and developed a new model based on the modal approach. This model is briefly recalled in the following section. It is valid for both rectangular and circular openings and for leaks. Sgard *et al.*[3] showed that Mechel's analytical model for infinitely wide slits[14] compares perfectly well with the modal approach when the width of the slit is taken to be sufficiently large. The modal approach deals with finite width slits and becomes time consuming for slits that are very wide compared to the acoustic wavelength. Mechel's model can therefore be used as an alternative, thereby reducing computation time.

The model developed by Sgard *et al.*[3] is based on modeling the sound field within the aperture in terms of propagating and evanescent acoustic modes. Aperture radiation is considered by means of a modal radiation impedance matrix. The coupled problem is then solved in terms of modal contribution factors to obtain the transmission coefficient for any plane wave angle of incidence. The diffuse field (or field-incidence) transmission coefficient is finally obtained by integrating over all angles of incidence. Model details can be found in Ref. 3. To summarize, the sound transmission coefficient for a plane wave with an angle of incidence (θ, ϕ) is given by

$$\tau(\theta_i, \varphi_i) = -\frac{\rho_0}{k_0 \cos\theta_i \rho_f^* |\hat{A}_i|^2 S} \Re\left(\sum_M N_M \hat{k}_M^* \hat{C}_M \hat{D}_M^*\right), \quad (4)$$

where the summation is performed over the aperture lateral modes. Modal coefficients $\hat{C}_M$ and $\hat{D}_M$ are related to modal radiation impedances, the area of the aperture, and the impedance of the fluid within the aperture. Coefficients $\hat{k}_M$ and N_M are the modal wave number and the modal norm, respectively. Amongst the remaining factors, it is worth mentioning that $\hat{A}_i$ represents the amplitude of the incident plane wave.

The diffuse field (or field-incidence) TL of the aperture is therefore given by

$$\mathrm{TL} = -10 \log_{10}(\tau_d), \quad (5)$$

where

$$\tau_d = \frac{\int_0^{2\pi}\int_0^{\theta_{\lim}} \tau(\theta_i,\varphi_i)\sin\theta_i \cos\theta_i d\theta_i d\varphi_i}{\pi \sin^2\theta_{\lim}}. \quad (6)$$

The diffuse field TL is obtained by setting the angle $\theta_{\lim}$ to 90°, while the field-incidence TL is defined by setting $\theta_{\lim}$ to 78°. The integrals in Eq. (6) are performed numerically. The model results presented in the following sections use the field-incidence TL.

In the case of Mechel's model[4] for slits of width b and depth d, the transmission coefficient is relatively easy to obtain and is given by

$$\tau(\theta_i,\varphi_i) = \frac{Z_0}{\cos\theta_i}\Re[\hat{Z}_{R2}]p_g^2 \left| \frac{\hat{Z}_a}{\hat{Z}_a(\hat{Z}_{R1}+\hat{Z}_{R2})\cos(\hat{k}_a d) + j(\hat{Z}_a^2+\hat{Z}_{R1}\hat{Z}_{R2})\sin(\hat{k}_a d)} \right|^2, \quad (7)$$

where $Z_0=\rho_0 c_0$ is the characteristic impedance of the external fluid, $\hat{Z}_{Ri}=Z_0\hat{Z}_{\text{slit},m}$ is the radiation impedance at the front ($i=1$) and rear faces ($i=2$) of the aperture, $\hat{Z}_f$ is the characteristic impedance of the fluid within the aperture, $\hat{k}_f$ is the wave number within the aperture, $p_g=2\sin_c(k_y b)$, $\hat{Z}_a=\hat{Z}_f/\cos\theta_f$, $\hat{k}_a=\hat{k}_f\cos\theta_f$, $k_y=k_0\sin\theta_i\sin\varphi_i$, k_0 is the wave number in the fluid outside the aperture, $\cos\theta_f=\sqrt{1-(k_0\sin\theta_i\cos\phi_i/\hat{k}_f)^2}$, and

$$\hat{Z}_{\text{slit},m} = 2k_0 b\left[H_0^{(2)}(u) + \frac{\pi}{2}[H_1^{(2)}(u)S_0(u) - H_0^{(2)}(u)S_1(u)] - \frac{1}{u}H_1^{(2)}(u) + \frac{2j}{\pi u^2}\right], \quad (8)$$

with $u=2b\sqrt{k_0^2-k_x^2}$, $k_x=k_0\sin\theta_i\cos\varphi_i$, and $H_0^{(2)}(u)$ and $H_1^{(2)}\times(u)$ as the zeroth and first order Hankel functions of the second kind, respectively.

Note that for very small cross-sectional areas and large depths, viscosity effects in the aperture may become significant and should be included in the calculation, as pointed out in Ref. 11. These effects can be included for both rectangular and circular cross sections using a simple model such as the one developed by Pierce.[14] This is based on expressing $\hat{k}_f$ as

$$\hat{k}_f = \frac{\omega}{c_f} + (1-i)\alpha_{\text{walls}}, \quad (9)$$

where

$$\alpha_{\text{walls}} = 2^{-3/2}\sqrt{\frac{\omega\mu}{\rho_f c_f^2}}\left[1+\frac{\gamma-1}{\sqrt{\text{Pr}}}\right]\frac{L}{A}, \quad (10)$$

in which μ is the dynamic viscosity, Pr is Prandlt number, ρ_f and c_f are the density and the sound velocity in the fluid within the aperture, respectively, and L and A are, respectively, the perimeter and the aperture cross-sectional area.

Equation (9) is valid for frequencies satisfying the condition

$$\left(\frac{L}{A}\right)^2\frac{\mu}{8\rho_f} < \omega < \left[\frac{9}{32}\left(\frac{L}{A}\right)^2\frac{\rho_f c_f^4}{\mu}\right]^{1/3}.$$

A. Opening case

Figure 9 illustrates the experimental data and the modal-based results for cases featuring two rectangular apertures with dimensions of $45\times45\times300$ and $60\times130\times300$ mm^3. It is clear that the experimental data match closely the calculation-based results, even in the low frequency range in which the incident field cannot be considered diffuse. No adjustments were made, and dimensions are those given in the tables. Modes have been kept up to a 10 kHz truncation frequency. An excellent correlation between the experimental data and the calculations confirms that these results can con-

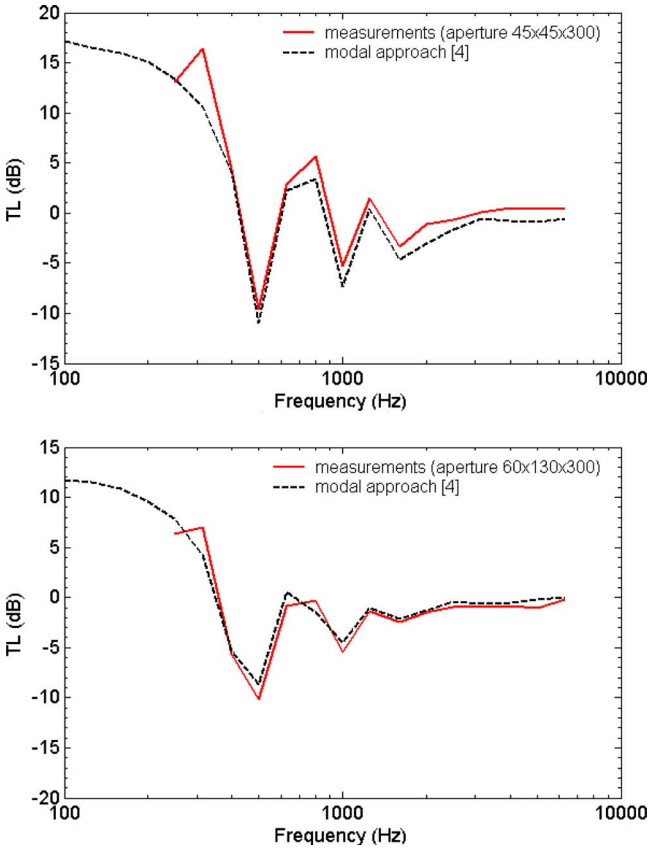


FIG. 9. (Color online) Comparison between experimental results and calculations for apertures of dimensions $45\times45\times300$ mm^3 (top figure) and $60\times130\times300$ mm^3 (bottom figure).

stitute a reliable set of experiments for the TL of baffled apertures exposed to a diffuse field. Correlation also underlines the accuracy of the model proposed in Ref. 3.

B. Slit case

The experimental results were compared with the two models (Mechel's[4] and modal approach[3]). It should be noted that viscous effects are taken into account in the calculations using Eq. (9). These effects are mainly important for slits with the smallest widths (0.5 and 2 mm) and the largest depth (50 mm) but are negligible otherwise.

For the smallest depth (1.5 mm), the experiments correlate fairly closely with the modal-based calculations above 500 Hz. Discrepancies can be attributed to nonuniformity of the incident field (see Sec. IV) below 630 Hz, although such discrepancies were not observed in larger openings (Fig. 9). In the slit case, the cutoff frequency is dictated by the slit length, which is very small. Oblique modes (which are numerous) can be excited in the frequency range of interest. The TL will therefore be more sensitive to nonuniformity of the incident acoustic field in this range. For openings, the cutoff frequency is thus higher at low frequencies, and it is primarily the plane wave mode that contributes to TL. However, the generalized force exerted on this mode is independent of the angle of incidence (see Ref. 3), and the TL is little influenced by nonuniformity of the incident acoustic field. Comparison leads to the same conclusions, except for the smallest width (not shown here for conciseness) for the medium slit depth (20 mm). As explained above, the difference between transmission through the slit and transmission through the wall in addition to flanking transmission is too low at this width, which is the reason for discrepancies. Measurements for this particular configuration cannot be used. The observation is the same for the largest slit depth (50 mm) and widths of 0.5 and 2 mm. The ratio between acoustic power transmitted through the slit and lateral transmissions is even worse. The discrepancies therefore accentuate and are unacceptable. Only the results for the 8 mm slit width are reliable for this 50 mm depth.

Correlation between experiments and modal-based calculations is acceptable above 500 Hz for the remaining six configurations. Figure 10 illustrates this for three of these configurations. This holds true irrespective of the considered model, i.e., Mechel's or modal approach. Correlation with Gomperts' model (not shown here for conciseness) is also acceptable. This tends to confirm that the series of tests described in Ref. 6 is unreliable.

VI. CONCLUSION

The purpose of this paper is to report new experimental diffuse field sound TL tests for rectangular openings and slits to validate the model developed in Ref. 3. This would provide additional reliable third-octave-band experimental data, which could be used in the framework of SEA.

This objective was fulfilled, although about half of the tests conducted on slits could not be used because of some uncertainties caused by flanking transmission paths. In the case of openings, experimental results were found to be reliable over almost all the frequency range of interest, namely, 250–5000 Hz. In the case of slits, only the results above the reverberant chamber cutoff transmission (630 Hz) are reliable.

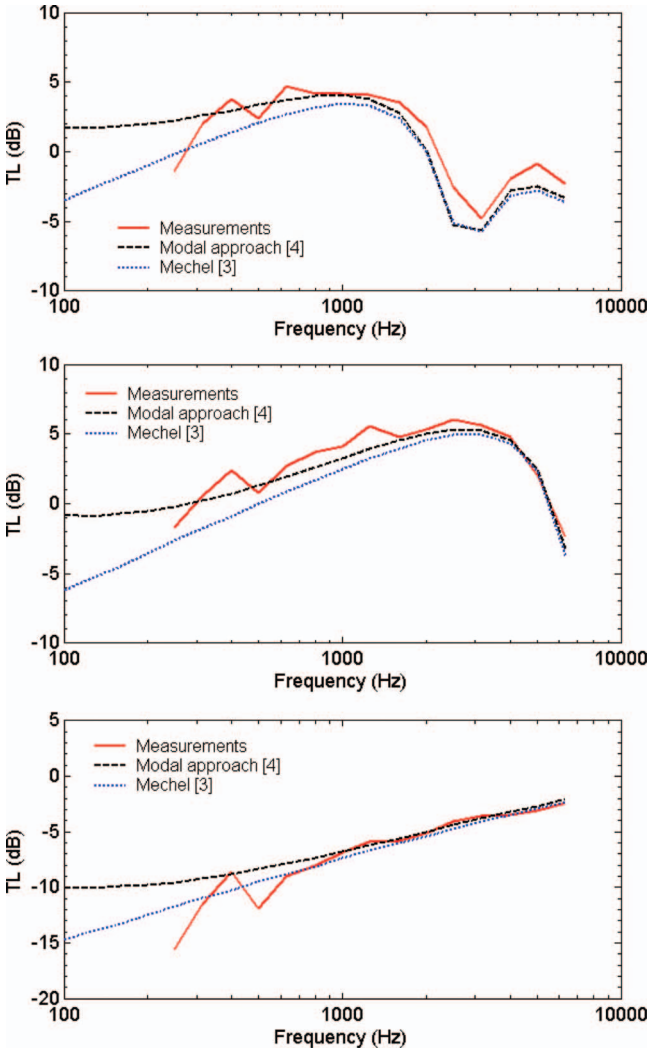


FIG. 10. Comparison between experimental results and calculations for slits of dimensions $500\times8\times5$ mm^3 (top figure) and $500\times2\times20$ mm^3 (middle figure), and $500\times2\times1.5$ mm^3 (bottom figure).

It was found that the modal approach correlates very closely with the experiments performed for all chosen configurations. This result, obtained with no adjustments, validates both the model and the tests. In addition, for sufficiently wide slits, Mechel's model[4] proves to be a faster alternative to the modal model, as observed in Ref. 3.

Future work will deal with modeling of apertures with acoustic linings, a configuration often encountered in industry, when the silencing equipment is fitted to apertures or when they comprise a simple hole in an acoustic panel.

ACKNOWLEDGMENTS

This research was undertaken within the framework of cooperation between IRSST and INRS. The research project scope involves developing accurate, easily applicable methods for modeling enclosures.

[1]H. Kuttruff, *Room Acoustics*, 2nd ed. (Applied Science, London, 1979).
[2]R. K. Miller and W. V. Montone, *Handbook of Acoustical Enclosures and Barriers* (Fairmont Press Inc., Atlanta, 1978).
[3]F. Sgard, H. Nelisse, and N. Atalla, "On the modelling of diffuse field sound transmission loss of finite thickness apertures," J. Acoust. Soc. Am. **122**, 302–313 (2007).
[4]F. P. Mechel, "The acoustic sealing of holes and slits in walls," J. Sound Vib. **111**, 297–336 (1986).
[5]M. C. Gomperts, "The sound insulation of circular and slit-shaped apertures," Acustica **14**, 10–16 (1964).
[6]M. C. Gomperts and T. Kihlman, "The sound transmission loss of circular and slit-shaped apertures in walls," Acustica **18**, 144–150 (1967).
[7]G. P. Wilson and W. W. Soroka, "Approximation to the diffraction of sound by a circular aperture in a rigid wall of finite thickness," J. Acoust. Soc. Am. **37**, 286–297 (1965).
[8]A. Sauter and W. W. Soroka, "Sound transmission through reverberant slots of finite depth between reverberant rooms," J. Acoust. Soc. Am. **47**, 5–11 (1970).
[9]B. M. Gibbs and Y. Balilah, "The measurement of sound transmission and directivity of holes by an impulse method," J. Sound Vib. **133**, 151–162 (1989).
[10]Y. Furue, "Sound propagation from the inside to the outside of a room through an aperture," Appl. Acoust. **31**, 133–146 (1990).
[11]D. J. Oldham and X. Zhao, "Measurement of the sound transmission loss of circular and slit-shaped apertures in rigid walls of infinite thickness by intensimetry," J. Sound Vib. **161**, 119–135 (1993).
[12]F.T. Chen, "Study of acoustic transmission through apertures in wall," Appl. Acoust. **46**, 131–151 (1995).
[13]J. M. Bruneau, "Acoustic of close space," *Fundamentals of Acoustic* (Hermes, Paris, 1998), Sec. 9.
[14]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, Woodbury, 1981), pp. 531–534.

T-matrix methods in acoustic scattering

P. C. Waterman[a)]
54 Chickadee Lane, West Yarmouth, Massachusetts 02673-2502

(Received 1 August 2008; accepted 3 November 2008)

Some further refinements are described for the *T*-matrix approach to acoustic scattering. From the structure of the matrices involved, one can infer the Rayleigh limit explicitly for objects having no density contrast. One finds $T^{\mathrm{Ray}}=iR-R^2$, where the *R*-matrix involves integrals of the regular spherical wave functions over the object's surface. The index of refraction and loss factor can be chosen as desired, and energy balance and reciprocity requirements are found to be met. The derivation can be extended to obtain the Rayleigh expansion, effectively describing *T* as a series in ascending powers of the ratio of object size to wavelength. In trial cases, the series converges throughout the Rayleigh region and somewhat beyond. Bodies of high aspect ratio are also considered, where difficulties arise due to precision loss during numerical integration. Loss ranges from 4 or 5 significant figures (2:1 spheroid) to 22 figures (40:1 spheroid) or more. A class of surfaces has been found for which this problem can be avoided, however, enabling one to treat a variety of body shapes up to aspect ratios of 100:1 with no difficulty. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3035839]

PACS number(s): 43.20.Fn, 43.58.Ls, 43.58.Ta [RLW] Pages: 42–51

I. INTRODUCTION

The *T*-matrix has been found useful in describing scattering of waves impinging on objects of nonspherical shape whose bulk properties differ from those of the surroundings. It was recognized quite early that the approach could equally well be applied in acoustics, electromagnetics, and linear elasticity, as described in earlier work.[1] Further applications followed—there are now two databases available, one general,[2] and a second,[3] with more recent updates,[4] dealing exclusively with electromagnetic problems. Excellent review articles have been given for the electromagnetic case by Mishchenko *et al.*[5]

Much of this work was concerned primarily with applications. Here we look at some further refinements of the theory. A recent paper examined these ideas for two dimensional problems in the electromagnetic case;[6] the present work considers the acoustic case in three dimensions. From the structure of the equation determining the *T*-matrix one can obtain an analytical formula for *T* in the Rayleigh limit, provided no density disparities are present. Alternately one can obtain the Rayleigh expansion, enabling computation of *T* at higher frequencies without requiring matrix inversion. The expansion generally converges throughout the Rayleigh region and somewhat beyond—just how far must be determined individually for each case.

The other problem of interest concerns objects of high aspect ratio, for which numerical difficulties are encountered—specifically, in the numerical integrations needed to obtain elements of the *Q*-matrix (which is then inverted to obtain the *T*-matrix). Some improvement is obtained by carrying out the computations at higher precision.[7,8] Another alternative was found by Sarkissian *et al.*[9] at the Naval Research Laboratory (NRL), who obtained results for thin rigid rods by first modifying the integrands of the *Q*-matrix. Here that idea is extended to a class of body shapes, for both penetrable objects and surface scatterers.

II. BASIC EQUATIONS

The basic equations for the null-field approach to the *T*-matrix have been given elsewhere.[6,10] They are summarized briefly here, including a few recent results, along with a change in normalization of the basis functions, the purpose of which will become clear.

Consider a homogeneous obstacle having density and stiffness (inverse compressibility), ρ' and μ', respectively, embedded in a uniform space ρ,μ. The boundary of the obstacle is specified in spherical coordinates by $r=r(\theta,\phi)$, with origin in the interior [see Fig. 1, where for simplicity a spheroid $r=r(\theta)$ is shown], and may have edges. With a time factor $\exp(-i\omega t)$ suppressed, the outgoing wave functions are given by

$$\psi_{\sigma mn}(k\mathbf{r})=(\varepsilon_m\gamma_{mn})^{1/2}h_n(kr)Y_{\sigma mn}(\theta,\phi),\quad m=0,1,\ldots,n,n=0,1,\ldots, \tag{1a}$$

with $\sigma=e/o$ (even/odd) in terms of the Hankel functions $h_n(kr)=j_n(kr)+in_n(kr)$ and the spherical harmonics

$$Y_{\sigma mn}(\theta,\phi)=(\cos/\sin)m\phi P_n^m(\cos\theta). \tag{1b}$$

Here the P_n^m are associated Legendre functions, $\varepsilon_0=1$, $\varepsilon_m=2$ otherwise, is the Neumann factor, and the normalizing constants are given by

$$\gamma_{mn}=(2n+1)(n-m)!/(n+m)!. \tag{1c}$$

The field in the interior of the obstacle will be represented by the regular functions

a)Electronic mail: pcwaterman@comcast.net

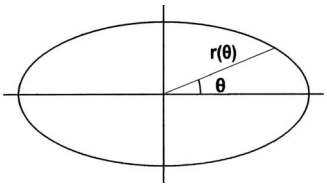


FIG. 1. A cross-sectional view of the object is shown, assuming rotational symmetry. In this and subsequent figures, the object is generated by rotation about the horizontal axis.

$$\chi_{\sigma mn}(k'\mathbf{r}) = 2(1+\rho'/\rho)^{-1} q^{-n} (\varepsilon_m \gamma_{mn})^{1/2} j_n(k'r) Y_{\sigma mn}(\theta,\phi). \tag{2}$$

Note that the relative density ρ'/ρ and the (possibly complex) index of refraction $q=k'/k$ have been included in the normalization here. Finally, the incident wave is represented by the functions $\chi_{\sigma mn}(k\mathbf{r})$ (for which q and ρ'/ρ become unity). It is also convenient to put the indices in one-to-one correspondence with a single index ν. The incident, scattered, and internal fields are now given explicitly by

$$\psi^{\text{inc}}(k\mathbf{r}) = \sum_{\nu=1} a_\nu \chi_\nu(k\mathbf{r}), \tag{3a}$$

$$\psi^{\text{sca}}(k\mathbf{r}) = \sum_{\nu=1} f_\nu \psi_\nu(k\mathbf{r}), \tag{3b}$$

$$\psi^{\text{int}}(k'\mathbf{r}) = \sum_{\nu=1} \alpha_\nu \chi_\nu(k'\mathbf{r}). \tag{3c}$$

Given the incident wave (the a_ν), we want to compute the scattered wave (f_ν) and possibly the interior field (α_ν). To this end, define the Q-matrix having elements[6]

$$Q_{u\nu} = (k/4\pi) \int d\boldsymbol{\sigma} \cdot [(\rho'/\rho)\chi_u(k'\mathbf{r}) \nabla \psi_v(k\mathbf{r}) - \nabla\chi_u(k'\mathbf{r})\psi_\nu(k\mathbf{r})], \quad u,\nu = 1,2,\ldots. \tag{4a}$$

Q is resolved into regular and singular parts R and S, with

$$Q = R + iS, \tag{4b}$$

by replacing the Hankel functions with Bessel or Neumann functions, respectively. The scattered wave is now computed for a given incident wave by the T-matrix, i.e.,

$$f = Ta, \tag{5a}$$

where (assuming T is symmetric, see below)

$$T = -Q^{-1}R. \tag{5b}$$

Three consistency checks are available here. First, the equations must be truncated at finite size in order to compute T, and one can then check for constancy of the elements versus truncation size. Second, reciprocity requires that T be symmetric, which also can be checked. Energy balance provides the third constraint: The absorption matrix A is defined in terms of the integral of the inward normal component of energy flux over the object surface, i.e. (dagger denotes conjugate transpose),

$$A = \tfrac{1}{4}(Q^{-1})^\dagger (W + W^\dagger) Q^{-1}, \tag{6a}$$

where elements of W are given by

$$W_{\mu\nu} = (k/4\pi) \int d\boldsymbol{\sigma} \cdot (\rho/\rho_0) \chi_\mu(k'^*\mathbf{r}) \nabla \chi_\nu(k'r), \tag{6b}$$

and now T must obey the detailed energy balance equations

$$-\operatorname{Re} T = T^\dagger T + \tfrac{1}{4}A. \tag{6c}$$

As noted elsewhere,[6] Eq. (6c) is a generalized optical theorem; for plane-wave incidence, it requires that (minus) the real part of the forward amplitude is equal to the sum of the scattering and absorption cross sections. On the other hand, for the scattering matrix $\mathcal{S}=I+2T$ connecting incoming and outgoing waves, where I is the identity matrix with elements $\delta_{vv'}$ (the Kronecker delta), one writes

$$\mathcal{S}^\dagger \mathcal{S} = I - A. \tag{6d}$$

The association of A with absorption is now clear—in order for $\mathcal{S}$ to be unitary, A must vanish.

One other constraint is implied by the above equations. Because T is symmetric, it follows formally from Eq. (5b) that the product (prime denotes transpose)

$$Z = QR' \tag{6e}$$

is also symmetric. This may be useful in (i) indicating that the computation has been programed correctly, and (ii) giving a rough estimate of the truncation size (of the infinite matrices) needed to obtain convergent results in any specific case, all *before* any matrix inversion is done. Unpublished numerical results confirm this.

For truncation size N, the scattered wave itself is given by

$$\psi^{\text{sca}}(k\mathbf{r}) = \sum_{\nu,\nu'=1}^{N} a_\nu T_{\nu\nu'} \psi_{\nu'}(k\mathbf{r}) \sim f(\theta,\phi)(1/ikr)\exp(ikr), \quad kr \gg 1, \tag{7a}$$

where the far-field amplitude is

$$f(\theta,\phi) = -\sum_{\nu,\nu'=1}^{N} a_\nu T_{\nu\nu'} (-i)^{\nu'} Y_{\nu'}(\theta,\phi). \tag{7b}$$

Finally, the scattering cross section is given by

$$\sigma^{\text{sca}} = (1/4\pi) \int d\Omega |f(\theta,\phi)|^2. \tag{7c}$$

For simplicity, the present work will consider primarily objects with rotational symmetry, for which a great deal of decoupling occurs. For Q, for example, the general matrix element reduces to

$$Q_{\sigma mn\sigma' m' n'} \doteq \delta_{\sigma\sigma'} \delta_{mm'} Q^{\sigma m}_{nn'}. \tag{8}$$

Here each fixed pair of superscripts σm gives an independent matrix, which we write simply as $Q_{nn'}$ when ambiguity is unlikely. If the object has a plane of mirror symmetry normal to the rotational symmetry axis, then in addition $Q_{nn'}=0$ for

$n+n'$ odd, and there is no coupling between even-even and odd-odd elements.

In order to write out the matrix elements in detail, note that

$$kd\boldsymbol{\sigma}\cdot\nabla = d\phi d\theta \sin\theta[x^2\,\partial/\partial x - x'(\theta)\,\partial/\partial\theta], \tag{9}$$

where x is set equal to $x(\theta)\equiv kr(\theta)$ after the derivatives are taken. For derivatives of the wave functions take

$$\sin\theta dP_n^m(\cos\theta)/d\theta = (n-m+1)P_{n+1}^m(\cos\theta) - (n+1)\cos\theta P_n^m(\cos\theta),$$

$$dj_n(x)/dx = (n/x)j_n(x) - j_{n+1}(x),$$

$$dn_n(x)/dx = -(n+1)(1/x)n_n(x) + n_{n-1}(x). \tag{10}$$

Putting these results in Eq. (4) for the S-matrix and integrating over azimuth, one obtains

$$\begin{aligned} S_{nn'} = d_{nn'}\int_0^\pi d\theta\{&j_n(qx)n_{n'}(x)G_{nn'}(\theta) \\ &+[(\rho'/\rho)j_n(qx)n_{n'-1}(x) \\ &+qj_{n+1}(qx)n_{n'-1}(x)]\}H_{nn'}(\theta), \end{aligned} \tag{11a}$$

where we have defined angular functions

$$\begin{aligned} G_{nn'}(\theta) = \{&[(\rho'/\rho)(n'+1)-(n+1)]r'(\theta)\cos\theta \\ &-[(\rho'/\rho)(n'+1)+n]r(\theta)\sin\theta\} \\ &\times P_n^m(\cos\theta)P_{n'}^m(\cos\theta) \\ &+r'(\theta)[(n-m+1)P_{n+1}^m(\cos\theta)P_{n'}^m(\cos\theta) \\ &-(\rho'/\rho)(n'-m+1)P_n^m(\cos\theta)P_{n'+1}^m(\cos\theta)], \end{aligned}$$

$$H_{nn'}(\theta) = r^2(\theta)\sin\theta P_n^m(\cos\theta)P_{n'}^m(\cos\theta), \tag{11b}$$

and the normalization constants are given by

$$d_{nn'} = 2(1+\rho/\rho_0)^{-1}q^{-n}(\gamma_{mn}\gamma_{mn'})^{1/2} \tag{11c}$$

(note that the Neumann factor has dropped out due to the azimuthal integration). For the R-matrix, replace $n_s(x)$ with $j_s(x)$ in Eq. (11a) wherever it occurs.

III. RAYLEIGH REGION AND BEYOND

Some interesting simplifications occur at the lower frequencies, provided we restrict ourselves to objects with no density contrast relative to their surroundings. Thus assume *for this section only* that $\rho/\rho'=1$. Defining $\delta=|qkr_{\max}|$, with $r_{\max}$ the maximum value of $r(\theta)$, we begin with the limiting case $\delta\ll 1$. Keeping only the lowest order term in each element of R, one finds the form (see Appendix A)

$$R_{nn'} = (q^2-1)r_{nn'}\delta^{n+n'+3}, \tag{12a}$$

where r is a real symmetric matrix, given by comparison with Eq. (A3). A similar analysis for S gives

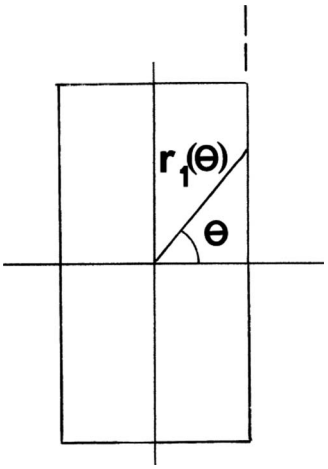


FIG. 2. The right circular cylinder is shown, with physical and nonphysical parts of the end-plane given by the solid and dashed lines, respectively.

$$S_{nn'} = I_{nn'} + (q^2-1)s_{nn'}\delta^{n-n'+2} \tag{12b}$$

with s real [but not in general symmetric, see Eq. (A5)]. As will be seen later, there is an exceptional case, the spheroid (or ellipsoid), for which S takes the form

$$S_{nn'} = I_{nn'} + (q^2-1)s_{nn'} \times \begin{cases} \delta^{n-n'+2} & \text{for } n' \le n \\ 1 \text{ (or } \delta) & \text{for } n+n' \text{ even (or odd), otherwise.} \end{cases} \tag{12c}$$

In these equations $n,n'=m,m+1,\ldots$.

Just as in the cylindrical case,[6] the identity matrix is separated out by writing

$$Q_0 = Q - iI. \tag{13}$$

For the transition matrix one now has $T=-(Q_0+iI)^{-1}R=(I-iQ_0)^{-1}iR\doteq(I+iQ_0)iR$, where in the last step we have used the binomial approximation, assuming Q_0 is small in an appropriate sense. The last equality can be written as

$$T = i[R-(S-I)R]-R^2.$$

Now, as one can verify from the preceding discussion, corresponding elements of the product $(S-I)R$ are smaller than those of R by a factor of order δ^2 and can be neglected (provided the product sr exists). Thus

$$T^{\mathrm{Ray}} = iR - R^2 \tag{14}$$

in the Rayleigh limit $\delta\ll 1$. Correct to leading order, the matrix product should also be replaced with scalar products, i.e., $(R^2)_{nn'}\to R_{n0}R_{0n'}$. Equation (14) clearly satisfies the reciprocity requirement. The second term is always higher order, but serves to adjust the phase in cases where no loss is present—note that to within leading order the energy balance equations, Eq. (6c), are satisfied by inspection in that event.

As an example, take the right circular cylinder

$$r(\theta) = \begin{cases} r_1(\theta), & 0\le\theta\le\theta_0 \\ r_2(\theta), & \theta_0\le\theta\le\pi/2, \end{cases} \tag{15}$$

with $r_1(\theta)=a/\cos\theta$ and $r_2(\theta)=b/\sin\theta$, as shown in Fig. 2,

TABLE I. Rapid falloff in magnitude of the matrix elements of T is seen for a "square" cylinder with size parameter $ka=0.1$, using six equations and unknowns (Rayleigh region, rounded logarithmic scale). The dashes indicate elements that vanish due to symmetry.

−3	—	−7	—	−10	—
—	−6	—	−10	—	−14
−7	—	−10	—	−14	—
—	−10	—	−13	—	−17
−10	—	−14	—	−17	—
—	−14	—	−17	—	−21

with cylinder half-length a, radius b, and $\theta_0=\tan^{-1} b/a$ Consider the "square" cylinder $b/a=1$, with $ka=0.1$ and $q=1+i$ (high loss). Computations have been done in double precision (15 decimal digits) on a PC using MATHCAD 11. The standard equations (4) and (5) were truncated at six equations in six unknowns, and magnitudes of the resulting T-matrix elements are shown in Table I. This along with other tables herein are laid out in the usual fashion with the (0, 0) element in the upper left corner, with entries rounded to integers, and for this case elements vanish whenever the sum of their indices is odd. Note the rapid falloff in magnitude, so that only the (0, 0) element is significant in the far field. In Table II, energy balance is checked by computing (minus) the log of the magnitude of the relative difference between elements on the two sides of Eq. (6c). At least four significant figure agreements are seen throughout.

The limiting form T^{Ray} was also computed from Eq. (14), and a similar comparison of T and T^{Ray} in Table III shows agreement typically within two significant figures. This is exactly what one would expect because T (but not T^{Ray}) includes, in addition, all higher order terms, beginning with those a factor $\delta^2\approx 10^{-2}$ smaller. Reducing the object size by an order of magnitude increases every entry of Table III by 2, again as would be expected.

The isotropic term is readily obtained analytically. From Eq. (14) $T_{00}^{\mathrm{Ray}}=iR_{00}-(R_{00})^2$, and the limiting value of R_{00} is evident from the volume integral of Eq. (A1). Letting V equal the object volume, one has in agreement with earlier work[10]

TABLE II. A comparison of the lhs and the rhs of the energy balance (6c) for the T-matrix of Table I, showing the number of significant figures of agreement for each entry.

9	—	4	—	4	—
—	9	—	4	—	4
4	—	7	—	5	—
—	4	—	9	—	9
4	—	5	—	7	—
—	4	—	5	—	7

TABLE III. Significant figures of agreement are shown for the matrices T and T^{Ray} given in Eqs. (5b) and (14), respectively, for the case of Table I. Agreement increases rapidly as the object size is further reduced (see text).

2	—	2	—	2	—
—	2	—	2	—	2
2	—	3	—	2	—
—	2	—	3	—	3
2	—	2	—	3	—
—	2	—	3	—	3

$$T_{00}^{\mathrm{Ray}}=i(q^2-1)(k^3V/4\pi)-(q^2-1)^2(k^3V/4\pi)^2$$
$$\equiv -(9.9925+4.99i)\times 10^{-4}. \tag{16}$$

These are terminating decimals—the numerical values shown are exact. The computer program can be checked by evaluating Eq. (A3) numerically and employing Eq. (14). Results of the computer run agree with Eq. (16) to 15 significant figures, a good consistency check.

Because Eq. (14) involves no inverse powers of r, it can be used even when the origin lies on the surface. An interesting case of this is the inverted spheroid given by

$$r(\theta)/a=(\cos^2\theta+\alpha^2\sin^2\theta)^{1/2},\quad 0\le\theta\le\pi. \tag{17a}$$

For aspect ratio $0\le\alpha\le 1$ this represents an inverted oblate spheroid, in the limit $\alpha=0$ becoming a pair of contiguous spheres given by

$$r(\theta)/a=|\cos\theta|. \tag{17b}$$

This is shown in Fig. 3, with origin of coordinates at the point of contact. Using the same parameters as in the first example, Eq. (16) gives

$$T_{00}^{\mathrm{Ray}}=-(1/1.44)(2.3997+1.1996i)\times 10^{-4}. \tag{17c}$$

Again the machine-computed value agrees to 15 figures. Similar results are obtained in the lossless case. This example is of particular interest in giving a closed-form result for the multiple scattering problem involving two penetrable spheres and may provide a useful check on more conventional procedures. Notice that Eqs. (14) and (16) also apply to the limiting case of a torus (doughnut with zero hole diameter) obtained from Eq. (17a) for an inverted *prolate* spheroid as $1/\alpha\to 0$.

It is no surprise, of course, that the far-field scattering is isotropic in these cases. The real value of Eq. (14) lies in enabling one to do near-field computations, via Eq. (7), which should provide useful information, including determi-

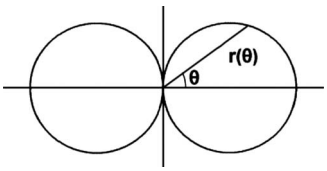


FIG. 3. Two contiguous spheres, here considered as a single obstacle, generated by an inverted oblate spheroid in the limit as aspect ratio in Eq. (17a) goes to zero.

nation of the region where the outgoing wave representation diverges. This will occur when the rate of increase in the Hankel functions is more than enough to overcome the fall-off exhibited in Table I.

Equation (16) contains both δ^3 and δ^6 terms. There is also a δ^5 term, not included because correct to leading order it does not contribute in either the real or imaginary part of T_{00}^{Ray}. Looking back at the derivation of Eq. (14), the higher order terms can be included by using the full binomial series, i.e., taking

$$T = (I + \sum_{p=1}^{\infty} i^p Q_0^p) iR. \tag{18}$$

This is the Rayleigh expansion. Keeping the $p=1$ term in the summation picks up all contributions to the δ^3, δ^5, and δ^6 terms along with some of the higher order terms. The next term in the summation picks up any missing δ^7 contributions, and so on. In our experience truncation at $2N$ terms is generally enough for seven figure accuracy (N is matrix truncation size), provided δ is such that the series is not too close to divergent. As with the cylindrical case, the series of Eq. (18) will converge as long as the spectral radius of Q_0 is less than unity.[6,11] Notice that this gives a means of determining the radius of convergence of the Rayleigh expansion. Based on our results, typically the series converges up to a limit of δ somewhere in the range 1–1.5, and hence above the Rayleigh region.

The coefficient of each higher power of δ involves infinite series sr, s^2r, and so on, and these series must be summed in order to obtain explicit closed-form analytical results. The major exception is the spheroid. For that case it follows from Eq. (12c) that each of the series terminates after a finite number of terms, as one can see by inspection, and thus any desired number of terms of the Rayleigh expansion can be obtained explicitly.

IV. HIGH ASPECT RATIOS

Experience has shown that shapes departing greatly from spherical tend to converge more slowly, if at all, versus truncation size. In many cases this is apparently due to large amplitude oscillatory terms in the integrands of the Q-matrix elements, which arise with high aspect-ratio bodies. One way of dealing with this is to carry out computations with higher precision. This approach was described by Mishchenko and Travis[7] at NASA and Havemann and Baran[8] in the United Kingdom, working in quad precision (31 decimal digits) with application to light scattering. An alternate approach was introduced by Sarkissian *et al.*[9] at the NRL, who considered underwater acoustic scattering by long, thin, rigid circular rods. They recognized that contributions from the oscillatory terms would vanish when integrated over an *infinitely* long circular cylinder, and based on this idea were able to eliminate the troublesome integrations exactly. Here we extend their approach to penetrable objects and investigate some of the shapes for which the technique is applicable, following the terminology used earlier.[6]

To this end, consider the quadric surfaces described by ($a,\ldots,k$ real)

$$ax^2 + by^2 + cz^2 + dxy + eyz + fzx + gx + hy + jz + k = 0. \tag{19a}$$

In spherical coordinates r,θ,ϕ, for simplicity assuming rotational symmetry, this can be rewritten as the quadratic equation

$$(a \sin^2\theta + c\cos^2\theta) r^2(\theta) + j\cos(\theta) r(\theta) + k = 0. \tag{19b}$$

We call the solutions of this equation, including the infinite plane, along with those shapes for which $1/r(\theta)$ is continuous over the range $0 \le \theta \le \pi$, *complete quadric* (cq) surfaces. In addition to the plane, this includes a pair of parallel planes, the infinite circular cylinder, prolate and oblate spheroids, and the paraboloid of revolution, but not the hyperboloid, which does not cover the full angular range. Circular disks and finite wires are also included as limiting cases of spheroids, although they may only have physical applications in electromagnetics.

Now the products of radial functions in the integrands of the S-matrix of Eq. (11) have the form $j_p(qx)n_q(x)$. Whenever $q>p$, expand this product in ascending powers of x, including negative powers. Then split S into the sum of terms $S^+ + S^-$ containing the positive and negative powers, respectively (but keeping the $1/x$ terms in S^+—see Appendix B). Matrix elements of S^- will only be nonzero above the diagonal. The integrands in S^- will become large whenever the surface passes close to the origin, where at the same time the angle variable changes most rapidly, causing large oscillations and resultant integration errors. Fortunately, this problem can be eliminated in many cases using the ansatz:

For integration over a cq surface,

$$S^- = 0. \tag{20}$$

This equation has been verified numerically for all the surfaces in question. In most if not all cases, Eq. (20) is due to the orthogonality of the Legendre functions. For example, for the infinite plane, one has $1/r(\theta) = \mathrm{const} \times \cos\theta$, so that inverse powers of $r(\theta)$ can be expanded in a finite sum of Legendres. At the same time the product of angular functions can be expanded in a second finite sum of Legendres. The one overlapping term vanishes due to the divergence theorem, and the remaining terms because of orthogonality.

Applications of Eq. (20) are best illustrated by example. The simplest case is the spheroid, given by [compare Eq. (17a)]

$$r(\theta)/a = (\cos^2\theta + \alpha^2 \sin^2\theta)^{-1/2}. \tag{21}$$

Because this is a cq surface, we can discard S^-, and the computation proceeds normally from this point, using

$$Q = R + iS^+. \tag{22}$$

For matrix inversion purposes this equation is much better behaved than the original, in that elements of S no longer grow in magnitude as one moves further above the main diagonal, but instead remain roughly constant. Note that the form of Eq. (22) is unaffected by coordinate translations along the symmetry axis, as long as the origin remains in the interior. Magnitudes of the resulting T-matrix elements for a 40:1 prolate spheroid (with centered origin) are shown in

TABLE IV. Magnitudes of T-matrix elements are shown for a 40:1 prolate spheroid with $N=6$, plotted as in Table I (see text). At $N=12$ these entries are unchanged, although the number of valid significant figures has increased.

$$\begin{pmatrix} -4 & — & -5 & — & -7 & — \\ — & -5 & — & -6 & — & -8 \\ -5 & — & -6 & — & -8 & — \\ — & -6 & — & -7 & — & -9 \\ -7 & — & -8 & — & -9 & — \\ — & -8 & — & -9 & — & -11 \end{pmatrix}$$

Table IV for truncation at $N=6$, with $kr_{\max}=1$, $m=0$, index of refraction $q=1+i$, and density ratio $\rho'/\rho=2$. Judged based on the three consistency criteria, far-field results are good to about four significant figures. Similar results are found for the higher azimuthal modes $m=1,2$, and so on, and also for oblate spheroids, in both cases up to aspect ratios of 100:1. To the best of our knowledge, results in the literature are limited to 20:1 aspect ratios.[12]

The advantage of discarding S^- becomes clear when one examines some of the higher-index matrix elements numerically. For example, in Fig. 4, we have plotted the magnitudes of the integrands of the (0, 10) elements of S^+ and S^- versus θ. One sees immediately that the second of these exceeds the first by as much as 22 orders of magnitude. But because S^- itself actually vanishes, this means that the use of Eq. (22) has prevented a loss of about 22 significant figures for the element in question, assuming of course one had that many to start with. The minima seen in both curves correspond to zeros and accompanying sign changes (the resolution of the plots is 1.57×10^{-5} rad). The zeros are effectively those of $G_{0,10}(\theta)$—see Eq. (11).

Significance errors increase as one moves further above the main diagonal of S or as the aspect ratio is increased. Note that significance loss may be a factor for bodies having only moderate aspect ratios as well. For example, for the 2:1

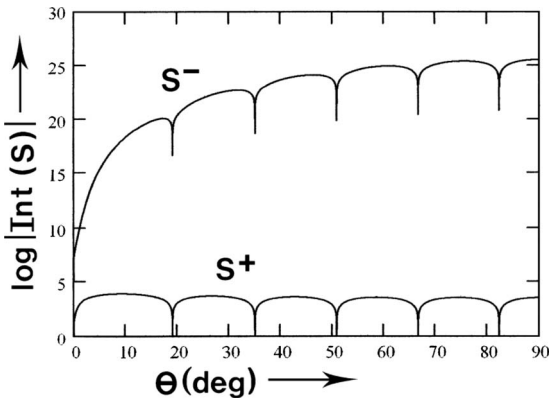


FIG. 4. Magnitudes of the integrands of the (0, 10) elements of S^+ and S^- are shown vs polar angle (with resolution 1.57×10^{-5} rad) for a 40:1 prolate spheroid—see text. Note that the latter dominates by more than 20 orders of magnitude, although S^- itself vanishes.

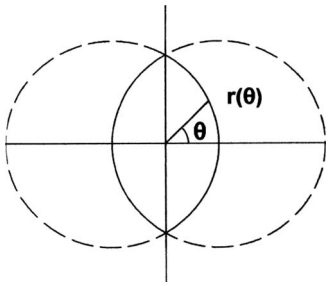


FIG. 5. A convex lens is formed by the intersection of two spheres, with physical and nonphysical parts again indicated by solid and dashed lines.

version of the spheroid discussed above, one finds a potential loss of between four and five figures in the (0, 6) element of S, avoidable using S^+.

Next, consider the finite cylinder of Eq. (15), in particular, the disk case where cylinder radius is greater than half-length (see Fig. 2). Notice that a cq surface (the infinite plane) is present, but the body only occupies a portion of it. Write $S_1(\text{pp})$ to indicate the integral over the physical part of $r_1(\theta)$ (the solid line in Fig. 2), and $S_1(\text{npp})$ for the nonphysical part (the dashed line). Now $S_1(\text{pp})=S_1^+(\text{pp})+S_1^-(\text{pp})$, and by Eq. (20) $S_1^-(\text{pp})+S_1^-(\text{npp})=0$. Combining these equations gives $S_1(\text{pp})=S_1^+(\text{pp})-S_1^-(\text{npp})$, i.e., $S_1(\text{pp})$ has been expressed in terms of integrals involving positive powers of $r(\theta)$ near the origin and inverse powers away from the origin. The final result for Q is

$$Q=R_1(\text{pp})+R_2(\text{pp})+i[S_1^+(\text{pp})-S_1^-(\text{npp})+S_2(\text{pp})]. \tag{23}$$

For the thin rod case, on the other hand, the infinite cylinder is treated as the cq surface, and the roles of S_1 and S_2 are interchanged. This is basically the case considered at the NRL,[9] except that an impenetrable cylinder was employed, which required that the end-caps be rounded off (see Sec. IV).

Next consider the translated sphere, where one would expect to use Eq. (22). Inspection of S^+, however, reveals that elements begin to fall off in magnitude both above and below the main diagonal, the further one goes from the diagonal (although S^- is unchanged). This is contrary to the behavior described above for Eq. (22). There is a simple explanation, however. The translated sphere has the unique property that *its inverse is also a sphere*. Because of this, the orthogonality property enjoyed by elements containing negative powers of $r(\theta)$ will be shared by those with positive powers, leading to the additional cancellations. Calling the resulting matrix S^{++} (see Appendix B), one has

$$Q=R+iS^{++}. \tag{24}$$

Another example using S^{++} is the spherical lens formed by the intersection of two spheres, as shown in Fig. 5. Note that the aspect ratio α can be made as large as one wants (the thin lens limit) by simply shifting each of the spheres by x_0, where

$$x_0=(1-\alpha^2)/(1+\alpha^2). \tag{25a}$$

The leftmost sphere is defined by

$$r(\theta)/a = (1 - x_0^2 \sin^2\theta)^{1/2} + x_0 \cos\theta, \tag{25b}$$

then scaled to the desired size. S is obtained by integrating over the physical part of the leftmost sphere (range 0 to $\pi/2$ because of mirror symmetry), then subtracting off the integral of S^- over the entire sphere (which vanishes). The net result is that

$$Q = R(\mathrm{pp}) + i[S^{++}(\mathrm{pp}) - S^{-}(\mathrm{npp})], \tag{25c}$$

with integrals over the leftmost sphere only. Using the same parameters as for the 40:1 spheroid discussed above but now for a 100:1 lens, for example, there is no problem in obtaining four or five significant figure accuracy in the far field using $N=8$. Note that one can treat parabolic lenses, formed by two intersecting paraboloids, and plano-convex lenses, in similar fashion.

The work of this section plays a significant role with regard to the internal field as well, and this should be commented on at least briefly. The internal field representation was given back in Eq. (3c), for which the expansion coefficients can be expressed as[10]

$$\alpha = -i(Q')^{-1}a. \tag{26}$$

Replacing Q' in Eq. (26) with its modified form, computed as in the above examples, should produce corresponding improvements in precision here also. Similar comments apply for the absorption matrix of Eq. (6a).

V. IMPENETRABLE AND NONREFRACTING OBJECTS

There are two more cases that should be mentioned in order to round out the high aspect-ratio discussion. Notice that there is nothing to prevent us from using the free-space form of the regular wave functions of Eq. (2), or their gradients, to represent fields on the *surface* of an impenetrable object, even though the *spatial* behavior of the wave functions is no longer germane to the physics. This has the advantage that methods of Sec. IV can still be used, in fact, in a much simpler form. Here surfaces must be rounded off where necessary in order to avoid having to introduce edge conditions, which were not needed earlier.

The limiting cases of "soft" and "hard" objects—the Dirichlet (D) and Neumann (N) cases—are characterized by the vanishing of either the pressure or the normal component of particle velocity on the surface, respectively. As shown some time ago, the Q-matrices for these cases are defined by (with a sign change in Q^D)[10]

$$Q^D_{uv} = -(k/4\pi)\int d\boldsymbol{\sigma}\cdot\nabla\chi_u(k\mathbf{r})\psi_v(k\mathbf{r}),$$

$$Q^N_{uv} = (k/4\pi)\int d\boldsymbol{\sigma}\cdot\chi_u(k\mathbf{r})\,\nabla\,\psi_v(k\mathbf{r}), \quad u,v = 1,2,\ldots. \tag{27a}$$

They can alternately be expressed as

$$Q^{D,N}_{uv} = (i/2)\delta_{uv} \mp (k/8\pi)\int d\boldsymbol{\sigma}\cdot\nabla[\chi_u(k\mathbf{r})\psi_v(k\mathbf{r})],$$

$$u,v = 1,2,\ldots. \tag{27b}$$

Matrix elements of $Q^{D,N}$ have a more complex structure in δ than shown in Eq. (12), and the Rayleigh limit formula of Eq. (14) no longer holds in general. It continues to be correct for the isotropic scattering element T_{00}, however, as can be seen by comparison with earlier results.[10]

Now when the Bessel and Neumann functions have the same argument, as in Eq. (27), one can write

$$j_n(z)n_{n'}(z) = j_{n'}(z)n_n(z) - \sum_{s=0}^{\mathrm{int}[(n'-n-1)/2]} c_{nn's}(2/z)^{n'-n-2s+1}, \tag{28a}$$

where

$$c_{nn's} = \frac{(n-n'+s+1)_s\,\Gamma(n-s+1/2)}{4s!\,\Gamma(n+s+3/2)},$$

$$n'-n = 1,2,\ldots. \tag{28b}$$

Here $(a)_0=1$, $(a)_S=(a+s-1)(a)_{S-1}$ otherwise, is the Pochhammer symbol, and the upper limit of summation is the integer part of $(n'-n-1)/2$. As discussed earlier for the cylindrical functions,[6] this equation can be verified by induction, noting that the case $n'-n=1$ is just the Wronskian relation, then making use of the standard interrelations between contiguous Bessel functions.[13]

By inspection of Eqs. (27) and (28), it is clear that both R and S^+ are always symmetric for surface scattering, and because $Q=R+iS^+$ for spheroids, Q is also symmetric in that case. These symmetries have been verified numerically for both spheroids and finite cylinders, and may be very useful in checking computational procedures. Note that S^+ must be taken *over the entire surface* for symmetry to hold, e.g., in the notation of Eq. (24), it is the *sum* $S_1^+(\mathrm{pp})+S_2^+(\mathrm{pp})$ that is symmetric.

In summary, for surface scatterers, high aspect ratios can be dealt with analogously to the volume-scatterer case, and, in fact, more simply, in that splitting into positive and negative powers of the radial variable is by Eq. (28) a fait accompli. This time there is no worry about where to put the $1/z$ term, incidentally—examination of Eq. (28) shows that there is no $1/z$ term on the right-hand side (rhs) (consequently, the $1/z$ terms that appear on the left-hand side (lhs) must cancel each other out). For both hard and soft spheroids, there is no problem in treating aspect ratios up to 1000:1, at least for the moderately small size-to-wavelength ratios considered herein.

The other case of interest for which Eq. (28) applies is the nonrefracting penetrable object, for which the index of refraction is unity, i.e., $q=k'/k=1$, arrived at by letting density and stiffness both vary proportionately so that $\rho'/\rho = \mu'/\mu$. From the geometrical acoustics point of view, a ray enters the object undeflected and with phase velocity unchanged, although boundary conditions require an amplitude change, and of course a reflected ray.

Looking back at Eq. (4a), the Q-matrix for the nonrefracting volume scatterer is seen to be expressible in terms of those for surface scattering, i.e.,

$$Q^{\text{non-refr}} = (\rho'/\rho)Q^N + Q^D, \tag{29a}$$

and hence

$$Q_{uv}^{\text{non-refr}} = (\rho'/\rho + 1)(i/2)\delta_{uv} + (\rho'/\rho - 1)(k/8\pi) \times \int d\boldsymbol{\sigma} \cdot \nabla[\chi_u(k\mathbf{r})\psi_v(k\mathbf{r})]. \tag{29b}$$

From Eq. (29b) the symmetries noted above hold here also, and the scattering is seen to vanish at $\rho'/\rho=1$ (object matched to surroundings). The nonrefracting spheroid is probably the only (nonspherical) volume scatterer for which the Q-matrix is symmetric. Incidentally, it is not clear if there are any cases of *physical* interest for nonrefracting obstacles. We mentioned them here because they may prove useful numerically in providing symmetry relations to check out computational procedures.

VI. DISCUSSION

Beginning at low frequencies, explicit formulas were obtained for both the Rayleigh limit and the Rayleigh expansion of the T-matrix for penetrable scatterers with $\rho'/\rho=1$. It would be of interest to extend these formulas by removing this restriction, if possible. Regarding applications, the Rayleigh expansion may prove useful in determining how closely objects may be situated when multiple scattering occurs before analysis by spherical wave expansions becomes untenable.

Problems connected with high aspect-ratio obstacles were also considered. Large significance losses were seen to occur in the Q-matrix elements in many cases, losses which could be prevented provided the bounding surface of the obstacle conformed at least in part to a "qc," i.e., a planar, spheroidal, or paraboloidal surface. The procedure for doing this applies to both volume- and surface-type scatterers, although it is much simpler in the latter case using the more compact equation (28). Note incidentally that although rigorous proof of some of these ideas is not presently available, numerical computations confirm them in considerable detail.

It would be of value to apply these techniques to vector problems, i.e., electromagnetics and elastodynamics, and in addition remove the restriction to axial symmetry, considering more general shapes. For these purposes the symmetries associated with perfectly conducting and nonrefracting objects may be very useful. Symmetry of Q for perfectly conducting spheroids is well known, and a symmetry proof was sketched for perfectly conducting ellipsoids some time ago, although no numerical verification was given in that case.[14] Public domain programs are available in FORTRAN 77 on the internet for electromagnetic applications—see, for example, the NASA programs documented in literature.[15] It should be relatively simple to use one of these programs to check, for example, the symmetry of Q for nonrefracting dielectric spheroids or perfectly conducting ellipsoids. Such confirmation would help immensely in verifying the existence of cancellations useful in dealing with high aspect ratios in the vector case.

ACKNOWLEDGMENTS

The author is indebted to Geoffrey Doyle and Science Applications International Corp., and Jeff Hale and the U. S. Army Edgewood Chemical and Biological Center, Aberdeen Proving Ground, Maryland, for their encouragement and sponsorship.

APPENDIX A: RAYLEIGH LIMIT FOR *R* AND *S*

The R and S matrices can be expressed more simply when $\rho'/\rho=1$. Beginning with Eq. (4) for R, apply the divergence theorem to get the volume integral

$$R_{nn'} = \varepsilon_m d_{nn'}(k^3/4\pi)(q^2-1) \times \int d\tau j_n(k'\mathbf{r}) j_{n'}(k\mathbf{r}) P_n^m(\cos\theta) P_{n'}^m(\cos\theta)\cos^2 m\phi \tag{A1}$$

with $n,n'=m,m+1,\ldots$. The $d_{n,n'}$ follow from Eqs. (1), (2), and (4). In order to get the Rayleigh limit of this equation, keep only the leading term in the low-frequency expansion of the Bessels, i.e.,[13]

$$j_n(x) \doteq f_n^+ x^n \tag{A2}$$

with $f_n^+=1/1\cdot3\cdots(2n+1)$. Putting this expression in Eq. (A1) and integrating twice, the result can be written as

$$R_{nn'} \doteq d_{nn'}^+(n+n'+3)^{-1}q^n(q^2-1) \times \int_0^\pi d\theta \sin\theta\,[kr(\theta)]^{n+n'+3} P_n^m(\cos\theta) P_{n'}^m(\cos\theta). \tag{A3}$$

Similarly, for the Neumann functions, one has

$$n_n(x) \doteq -f_n^- x^{-(n+1)}, \tag{A4}$$

where $f_{-1}^-=-1$, $f_0^-=1$, and $f_n^-=1\cdot3\cdots(2n-1)$ otherwise. Replacing $j_{n'}$ with $n_{n'}$ and repeating the above process, this time excluding a small spherical volume about the origin, one gets

$$S_{nn'} \doteq I_{nn'} - d_{nn'}^- q^n(q^2-1) \times \int_0^\pi d\theta \sin\theta\, g_{nn'}(\theta) P_n^m(\cos\theta) P_{n'}^m(\cos\theta), \tag{A5a}$$

where

$$g_{nn'}(\theta) = \begin{cases} \ln[kr(\theta)] & \text{for } (n-n'+2)=0 \\ (n-n'+2)^{-1}[kr(\theta)]^{n-n'+2} & \text{otherwise.} \end{cases} \tag{A5b}$$

The normalizing constants in Eqs. (A3) and (A5a) are given by

$$d_{nn'}^{\pm} = \tfrac{1}{4}(\gamma_{mn}\gamma_{mn'})^{1/2} f_n^+ f_{n'}^{\pm}. \tag{A6}$$

Notice that in view of Eq. (12c), the above equation for S

must be further modified for spheroids.

APPENDIX B: EVALUATION OF $S^{\pm}$

The product of spherical Bessel and Neumann functions of different arguments can be expressed as an ascending power series as follows. The standard expansions are[13]

$$j_n(z) = \sum_{k=0}^{\infty} a(n,k)(2^{-1/2}z)^{n+2k}, \tag{B1}$$

$$n_{n'}(z) = \sum_{s=0}^{\infty} b(n',s)(2^{-1/2}z)^{2s-n'-1}, \tag{B2}$$

provided $n' \geq 1$, where

$$a(n,k) = \begin{cases} 2^{n/2}\left[\prod_{w=0}^{n}(2w+1)\right]^{-1} & \text{for } k=0 \\ -[k(2k+2n+1)]^{-1}a(n,k-1) & \text{otherwise,} \end{cases} \tag{B3}$$

$$b(n',s) = \begin{cases} -2^{-(n'+1)/2}\prod_{w=1}^{n'}(2w-1) & \text{for } s=0 \\ -[s(2s-2n'-1\ ]^{-1}b(n',s-1) & \text{otherwise.} \end{cases} \tag{B4}$$

The product is now written as

$$\begin{aligned} jn\mathrm{sum}_{nn'}(q,z) &\equiv j_n(qz)n_{n'}(z) \\ &= \sum_{k=0}^{K}\sum_{s=0}^{K} a(n,k)b(n',s) \\ &\quad \times q^{n+2k}(2^{-1/2}z)^{2k+2s+n-n'-1}, \end{aligned} \tag{B5}$$

where K is an appropriate cut-off value ($K=2N$ appears to work well).

In order to split the summation into positive and negative powers when $n'>n$, consider the step function $h(x)=1$ for $x \geq 0$, zero otherwise. Provided $x \neq 0$, it is clear that $h(x)+h(-x)=1$. Inserting the lhs of this equality in the double sum with x set equal to $2k+2s+n-n'-1$ will separate the positive and negative powers, as can be seen. In addition it will turn out that we should keep the $1/z$ terms with the positive powers, accomplished by adding 1 to x, and also add $\frac{1}{2}$ to exclude the $x=0$ case. Thus, writing $h^{\pm} \equiv h\left[\pm(2k+2s+n-n'+\frac{1}{2})\right]$, and with X as shown in Eq. (B5), the sum of positive powers (including constants and $1/z$ terms) is now

$$\begin{aligned} jn\mathrm{sum}^{+}_{nn'}(q,z) &= \sum_{k=0}^{K}\sum_{s=0}^{K} h^{+}X \\ &= \sum_{k=0}^{s_0}\sum_{s=0}^{s_0} h^{+}X + \sum_{k=s_0+1}^{K}\sum_{s=0}^{s_0} X + \sum_{k=0}^{K}\sum_{s=s_0+1}^{K} X \\ &= \sum_{k=0}^{s_0}\sum_{s=0}^{s_0} h^{+}X + j\mathrm{sum}_{nn'}(qz)n1\mathrm{sum}_{nn'}(z) \\ &\quad + j_n(qz)n2\mathrm{sum}_{nn'}(z). \end{aligned} \tag{B6}$$

In the second line, the sums over s, then k, have been broken into two parts, with break point $s_0=\mathrm{int}[(n'-n)/2]$, the integer part of $(n'-n)/2$. One then sees that h^{+} can be deleted in the second and third terms because its argument is now always positive. But with h^{+} gone, these terms can be factored back into products, and this has been done in the third line. Here the partial sums of radial functions are defined as

$$\begin{aligned} j\mathrm{sum}_{nn'}(qz) &= \sum_{k=s_0+1}^{K} a(n,k)(2^{-1/2}qz)^{2k+n}, \\ n1\mathrm{sum}_{nn'}(z) &= \sum_{s=0}^{s_0} b(n',s)(2^{-1/2}z)^{2s-n'-1}, \\ n2\mathrm{sum}_{nn'}(z) &= \sum_{s=s_0+1}^{K} b(n',s)(2^{-1/2}z)^{2s-n'-1}. \end{aligned} \tag{B7}$$

Because Eq. (B6) will be employed in numerical integration, it is advisable for computational efficiency to further reduce the remaining double sum to single summations. We want to sum over the square array of elements with indices ranging from zero to s_0. Note that elements along the main diagonal and each of the parallel diagonals involve a fixed power of z. One can thus sum over k, making the change in summation variable $s=s_0+r-k$, to obtain a series coefficient for each power of z. Here r denotes the slant row number, $r=0$, giving the $1/z$ terms along the main diagonal, $r=1$, the constants along the next higher parallel, and so on (parallels below the main diagonal are excluded by the step function). The net result is that

$$\sum_{k=0}^{s_0}\sum_{s=0}^{s_0} h^{+}X = \sum_{r=0}^{s_0} c^{+}_{nn'r}(2^{-1/2}z)^{2r+2s_0+n-n'-1}, \tag{B8a}$$

where the coefficients

$$\begin{aligned} c^{+}_{nn'r} &= h(2s_0+2r+n-n'+1/2) \\ &\quad \times \sum_{k=r}^{s_0} a(n,k)b(n',s_0+r-k)q^{2k+n} \end{aligned} \tag{B8b}$$

are independent of z. Here the step function is always unity when $n+n'$ is even. Putting this result in Eq. (B6) gives the final form of $jn\mathrm{sum}^{+}_{nn'}$.

The inverse powers are much simpler to obtain. Going back to Eq. (B5), we have

$$jn\mathrm{sum}^{-}_{nn'}(q,z) = \sum_{k=0}^{s_0}\sum_{s=0}^{s_0} h^{-}X. \tag{B9}$$

The step function has allowed us to reduce both the upper limits of summation in Eq. (17a) to s_0, so that the sum is again over the square array encountered above, but now including only terms *on or below* the main diagonal. The procedure is analogous, this time letting $s=s_0-r-k$, and one obtains

$$jn\mathrm{sum}^{-}_{nn'}(q,z) = \sum_{r=0}^{s_0} c^{-}_{nn'r}(2^{-1/2}z)^{2s_0-2r+n-n'-1}, \qquad \text{(B10a)}$$

where

$$c^{-}_{nn'r} = h[-(2s_0 - 2r + n - n' + 1/2)] \times \sum_{k=0}^{s_0-r} a(n,k)b(n', s_0 - r - k)q^{2k+n}. \qquad \text{(B10b)}$$

Splitting is only necessary above the main diagonal of S, arranged by taking the final quantities of interest to be

$$jn^{\pm}_{nn'}(q,z) = \begin{cases} jn\mathrm{sum}^{\pm}_{nn'}(q,z) & \text{for } n' > n \\ j_n(qz)n_{n'}(z) & \text{otherwise.} \end{cases} \qquad \text{(B11)}$$

The special form S^{++} for translated spheres is obtained by replacing s_0 with $s_1 = n' - n$. One can confirm this by substituting s_1 in Eq. (B8) to find the leading term to be of order $z^{n'-n-1}$, which corresponds to the behavior observed numerically. In both cases, incidentally, the coefficients $c^{-}_{nn'r}$ (including the value of s_0) are unchanged.

[1] *Acoustic, Electromagnetic and Elastic Wave Scattering—Focus on the T-Matrix Approach*, edited by V. K. Varadan and V. V. Varadan (Pergamon, New York, 1980).

[2] V. V. Varadan, A. Lakhtakia, and V. K. Varadan, "Comments on recent criticism of the T-matrix method," J. Acoust. Soc. Am. **84**, 2280–2284 (1988).

[3] M. I. Mishchenko, G. Videen, V. A. Babenko, N. G. Khlebtsov, and T. Wriedt, "*T*-matrix theory of electromagnetic scattering by particles and its applications: A comprehensive reference database," J. Quant. Spectrosc. Radiat. Transf. **88**, 357–406 (2004).

[4] M. I. Mishchenko, G. Videen, V. A. Babenko, N. G. Khlebtsov, and T. Wriedt, "Comprehensive *T*-matrix reference database: A 2004-06 update," J. Quant. Spectrosc. Radiat. Transf. **106**, 304–324 (2007); see also M. I. Mishchenko, G. Videen, N.G. Khlebtsov, T. Wriedt, and N. T. Zakharova, "Comprehensive *T*-matrix reference database: A 2006-07 update," *ibid.* **109**, 1447–1460 (2008).

[5] M. I. Mishchenko, L. D. Travis, and D. W. Mackowski, "T-matrix computations of light scattering by non-spherical particles: A review," J. Quant. Spectrosc. Radiat. Transf. **55**, 535–575 (1996); see also M.I. Mishchenko, L. D. Travis, and A. A. Lacis, *Scattering, Absorption, and Emission of Light by Small Particles* (Cambridge University Press, Cambridge, 2002), Chap. 5.

[6] P. C. Waterman, "The *T*-matrix revisited," J. Opt. Soc. Am. A **24**, 2257–2267 (2007).

[7] M. I. Mishchenko and L. D. Travis, "T-matrix computations of light scattering by large spheroidal particles," Opt. Commun. **109**, 16–21 (1994).

[8] S. Havemann and A. J. Baran, "Calculation of the phase matrix elements of elongated hexagonal ice columns using the *T*-matrix method," J. Quant. Spectrosc. Radiat. Transf. **89**, 87–96 (2004).

[9] A. Sarkissian, C. F. Gaumond, and L. R. Dragonette, "*T*-matrix implementation of forward scattering from rigid structures," J. Acoust. Soc. Am. **94**, 3448–3453 (1993).

[10] P. C. Waterman, "New formulation of acoustic scattering," J. Acoust. Soc. Am. **45**, 1417–1429 (1969).

[11] A. S. Householder, *The Theory of Matrices in Numerical Analysis* (Blaisdell, Waltham, 1964), p. 54.

[12] C. Feuillade and M. F. Werby, "Resonances of deformed gas bubbles in liquids," J. Acoust. Soc. Am. **96**, 3684–3692 (1994).

[13] H. A. Antosiewicz, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (U. S. Government Printing Offices, Washington, DC, 1964), Chap. 10.

[14] P. C. Waterman, "Symmetry, unitarity, and geometry in electromagnetic scattering," Phys. Rev. D **3**, 825–839 (1971).

[15] M. I. Mishchenko and L. D. Travis, "Capabilities and limitations of a current Fortran implementation of the T-matrix method for randomly oriented, rotationally symmetric scatterers," J. Quant. Spectrosc. Radiat. Transf. **60**, 309–324 (1998); the programs themselves are available at www.giss.nasa.gov (Last viewed 7/31/2008).

Three-dimensional propagation and scattering around a conical seamount

Wenyu Luo and Henrik Schmidt
Department of Mechanical Engineering, Massachusetts Institute of Technology, Massachusetts 02139

(Received 2 December 2007; revised 7 July 2008; accepted 21 December 2008)

A three-dimensional (3D) propagation and scattering model is developed for an offset acoustic source in an ocean with axisymmetric bathymetry. Based on the same theoretical foundation as the formulation presented by Taroudakis [M. I. Taroudakis, "A coupled-mode formulation for the solution in the presence of a conical sea-mount," J. Comput. Acoust. **4**, 101–121 (1996)], the present approach combines a spectral decomposition in azimuth with a coupled-mode theory for two-way, range-dependent propagation. However, the earlier formulations were severely limited in terms of frequency, size, and geometry of the seamount, the seabed composition, and the distance between the source and the seamount, and were therefore severely limited in regard to realistic seamount problems. Without changing the fundamental theoretical foundation, the present approach applies a number of modifications to the numerical formulation, leading to orders of magnitude in numerical efficiency for realistic problems. Further, by using a standard normal mode model for determining the fundamental modal solutions and coupling matrices, and by applying a simple superposition principle, the computational requirements are made independent of the distance between the seamount and the source and receivers, and dependent only on the geometry of the seamount and the source frequency. Therefore, realistic propagation and scattering scenarios can be modeled, including effects of seamount roughness and realistic sedimentary structure.
 [DOI: 10.1121/1.3025903]

PACS number(s): 43.30.Bp, 43.30.Gv, 43.20.Fn [RCG] Pages: 52–65

I. INTRODUCTION

Over the past several decades, a large number of numerical models have been developed for acoustic propagation in the ocean. Most of these models provide solutions for two-dimensional (2D) (range and depth) problems, and they provide satisfactory solutions for the majority of propagation problems where the environmental dependence on azimuth is insignificant.

However, there are several classes of propagation problems where three-dimensional (3D) effects cannot be ignored. These include, for example, propagation over and around a seamount or through a strong eddy. For such problems, the 2D models, in which energy does not couple between planes of constant azimuth with respect to the source, often fail to provide accurate solutions, and a 3D model is required to provide accurate solutions for the field in range, depth, and azimuth. An important 3D effect, as shown in Ref. 1 and this paper, is that the azimuthal coupling increases the width of shadows behind islands and seamounts, which can be explained physically by horizontal refraction. The bending of 3D ray paths in basins, troughs, and near seamounts has been investigated theoretically in Ref. 2.

Solving such propagation problems for fully 3D environments is associated with almost prohibitive computational cost. However, in some cases most of the 3D effects can be captured using a simplified ocean model, such as a 2D axisymmetric model of a seamount or a planar 2D model for a continental slope. For such geometries, integral transforms may be used, replacing the solution of the full 3D wave equation by superposition of a finite number of 2D solutions, which may be determined using one of the standard modeling techniques.

The $N\times 2$D approach introduced first by Perkins and Baer[3] assumes out-of-plane scattering to be insignificant, and is therefore applicable only to problems with weak transverse environmental variability. It simply uses a standard 2D model to predict the propagation in vertical "slices" connecting the source and receiver, and is therefore highly efficient in solving 3D propagation problems. From the numerical examples in this paper, it is shown that the $N\times 2$D approach is a good approximation of the true 3D approach under certain conditions, while rather inadequate for other problem configurations, not surprisingly those associated with strong transverse environmental variability, such as propagation across the sides of steep seamounts.

Various 3D approaches have been developed. Some of these 3D models, such as the adiabatic mode parabolic equation approach used in Ref. 1 and FOR3D, a 3D parabolic equation model, presented in Ref. 4 and Ref. 18, can provide approximate solutions for 3D seamount problems.

Athanassoulis and Propathopoulos[5] presented an analytic solution for the 3D problem of acoustic scattering from a nonpenetrable cylindrical island in shallow water. Although that solution is valid only for a rigid or pressure-release cylindrical island, it can serve as a 3D benchmark solution, and has been used as such for validating the present approach.

In 1996, a coupled-mode formulation for the solution of the Helmholtz equation in water in the presence of a conical seamount was developed by Taroudakis.[6] In his work, the conical seamount was divided into a number of rings, in each

of which a series expansion of the acoustic pressure in terms of normal modes and an azimuthal Fourier series was applied.[7] The expansion coefficients were obtained by solving linear systems of equations resulting from the continuity conditions at the vertical interfaces separating the ring-shaped environmental sectors. The formulation introduced was theoretically exact, but the numerical implementation was inefficient, and associated with numerical stability issues, which severely limited the applicability to realistic seamount problems. First, the applied expansion involved Hankel functions of high order, for which the numerical evaluation is inherently unstable for high Fourier orders and small arguments. Second, when the source is far from the seamount, the number of azimuthal modes required for convergence is too large to make this formulation applicable. Finally, the choice of radial Hankel functions in Taroudakis' formulation yields unstable solutions of the mode coupling equations at high azimuthal orders.

To alleviate some of these issues, Eskenazi modified Taroudakis' model by applying the direct global matrix (DGM) approach[8] to obtain stable coupled-mode systems.[9] However, Eskenazi's model was still inefficient for distant source problems, and the use of the global coupled-mode equations still limited the applicability to relatively simple, canonical seamount problems.

Here a 3D spectral coupled-mode model is developed, based on the same theoretical foundation as used by previous investigators, but applying a number of numerical tools which extend the numerical feasibility to problems characterized by large Fourier orders, such as those involving distant sources and realistic seamounts of significant extent. In this work, the seamount propagation problem is reformulated into a scattering problem. As demonstrated in this paper, this reformulation has significant impact in terms of numerical efficiency. The modifications are as follows.

(1) $J_m(\cdot)$ and $H_m^{(1)}(\cdot)$ are used as the two linearly independent solutions of the Bessel equation in the present model, instead of $H_m^{(1)}(\cdot)$ and $H_m^{(2)}(\cdot)$ in Taroudakis' approach. $J_m(\cdot)$ is the Bessel function of order m and the first kind, and $H_m^{(1)}(\cdot)$ and $H_m^{(2)}(\cdot)$ are the Hankel functions of order m and the first and the second kind, respectively. The advantage is that $J_m(\cdot)$ and $H_m^{(1)}(\cdot)$ remain linearly independent numerically for both large and small arguments,[10] whereas the two Hankel functions become numerically indistinguishable at high order m.
(2) Normalized Bessel and Hankel functions are used to avoid overflow and underflow problems; in addition, the asymptotic forms of the normalized Bessel and Hankel functions for small and large arguments are used. As a result, there is no numerical problem in evaluating high-order normalized Bessel and Hankel functions. Moreover, the recurrence relations of Bessel and Hankel functions are used in evaluating these functions of different orders, which improves efficiency.
(3) The single-scatter approximation used in the present model vastly improves numerical efficiency, and as has been shown in the past, is highly accurate for most ocean acoustic propagation problems.[8] Therefore, for each azimuthal mode, instead of solving one linear system of a large dimension as in Eskenazi's model, this model solves multiple linear systems of small dimensions.
(4) In the present model, a standard normal mode model[11] (CSNAP) is used for determining the fundamental modal solutions and their associated coupling matrices. A noteworthy feature of the present approach, which is also true in Taroudakis' approach and Eskenazi's approach,[6,9] is that the coupling matrices are independent of azimuthal orders, so they need to be calculated only once.
(5) The efficiency is improved dramatically by introducing the superposition representation of the external field with respect to the seamount.[12,13] In the present model, the number of azimuthal modes required for convergence depends only on the product of the wavenumber and the radius of the base of the conical seamount. In both Taroudakis' model and Eskenazi's model, this value depends on the product of the wavenumber and the distance between the source and the axis of the conical seamount.
(6) The present model is perfectly scalable and is therefore easily parallelized to run on computer clusters, making it applicable to large-scale 3D problems.

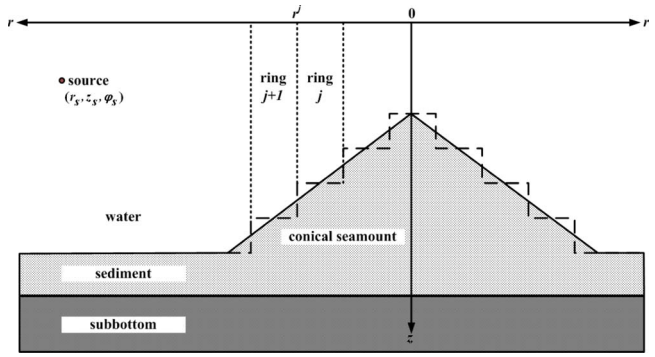


FIG. 1. (Color online) An ocean waveguide with a conical seamount, which is approximated by a number of range-independent ring-shaped sectors.

II. THEORY

A. Conical seamount problem

The geometry considered involves a point source, offset horizontally relative to a conical seamount, as illustrated in Fig. 1. This is a 3D problem in which the acoustic field depends not only on range and depth, but also on azimuth.

In the coupled-mode approach, a number of range-independent ring-shaped sectors are used to approximate such a conical seamount. The notation r^j is used to denote the range at the interface between ring j and ring $j+1$, for $j=1,2,\ldots,J-1$, as illustrated in Fig. 1.

A cylindrical-polar coordinate system is introduced, centered at the axis of the seamount. The location of the point source is denoted by $\boldsymbol{r}_s=(r_s,z_s,\phi_s)$, and the location of a field point is denoted by $\boldsymbol{r}=(r,z,\phi)$. The 3D Helmholtz equation for the field produced by the source, in the outer source ring, is[6,12]

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2 p}{\partial \phi^2}+\rho(z)\frac{\partial}{\partial z}\left(\frac{1}{\rho(z)}\frac{\partial p}{\partial z}\right)+\frac{\omega^2}{c^2(z)}p$$
$$=-\frac{\delta(r-r_s)}{r}\delta(z-z_s)\delta(\phi-\phi_s), \quad (1)$$

where $p=p(r,z,\phi)$ is the acoustic pressure (factoring out the harmonic time dependence $e^{-i\omega t}$, ω being the angular frequency), and $\rho(z)$ and $c(z)$ are the density and sound speed profiles in the outer source ring, respectively.

In ring j, which is a range-independent environment, the density and sound speed profiles are denoted by $\rho^j(z)$ and $c^j(z)$. The 3D Helmholtz equation is[6,12]

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p^j}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2 p^j}{\partial \phi^2}+\rho^j(z)\frac{\partial}{\partial z}\left(\frac{1}{\rho^j(z)}\frac{\partial p^j}{\partial z}\right)+\frac{\omega^2}{c^{j^2}(z)}p^j$$
$$=0, \quad (2)$$

where $p^j=p^j(r,z,\phi)$ is the acoustic pressure in ring j.

B. Spectral coupled-mode approach

In the following, a new, numerically efficient, spectral normal-mode solution is presented for the solution of the Helmholtz equation for the seamount problem. The theoretical foundation for the present approach is identical to that of previous solutions to this problem, such as those presented by Taroudakis[6] and Eskenazi.[9]

The environmental representations used by Eskenazi and Taroudakis are very similar. Both divided the seamount environment into three regions, one being the innermost, cylindrical sector containing the summit of the seamount (I), one containing all the ring-shaped sectors between the central cylinder and the source range (II), and the third being a semi-infinite ring outside the source range (III). The difference between the two lies in the numerical implementation of the spectral coupled-mode representation of the field. Therefore, Taroudakis[6] used unnormalized Hankel functions of the first and the second kind in all sectors for representing the range-dependence of the field. However, as demonstrated by Ricks *et al.*,[14] this choice is numerically unstable at high Fourier orders. Recognizing this, Eskenazi[9] used the stable, normalized Bessel functions of the first kind and the Hankel functions of the first kind. To frame the presentation of the present formulation, the earlier work is first briefly summarized.

1. Taroudakis' formulation

Taroudakis[6] represented the field in ring number j as

$$p^j(r,z,\phi)=\sum_{m=0}^{\infty}\sum_{n=1}^{\infty}R_{mn}(r)\Psi_n^j(z)\Phi_m(\phi), \quad (3)$$

where $\Psi_n^j(z)$ are local, depth-dependent eigenfunctions satisfying the orthogonality relation

$$\int_0^{\infty}\frac{1}{\rho^j(z)}\Psi_n^j(z)\Psi_\nu^j(z)dz=\delta_{n\nu}, \quad n,\nu=1,2,\ldots. \quad (4)$$

$\Phi_m(\phi)$ are azimuthal eigenfunctions

$$\Phi_m(\phi)=e_m\cos m\phi, \quad m=0,1,\ldots, \quad (5)$$

satisfying the orthogonality relation

$$\int_{-\pi}^{\pi}\Phi_m(\phi)\Phi_\mu(\phi)d\phi=\delta_{m\mu}, \quad m,\mu=0,1,\ldots, \quad (6)$$

requiring the coefficients e_m to be

$$e_m=\begin{cases}\dfrac{1}{\sqrt{2\pi}}, & m=0\\ \dfrac{1}{\sqrt{\pi}}, & m\neq 0.\end{cases} \quad (7)$$

To satisfy the initial condition of finite field at the axis of the seamount, and the radiation condition at infinity, the range representation of the field $R_{mn}(r)$ takes on different forms in the three regions, as follows.

(I) Inner cylinder, $r\leq r^1$,

$$R_{mn}(r)=A_{mn}J_m(k_{rn}^1 r). \quad (8)$$

(II) Ring j, $r^1<r\leq r_s$,

$$R_{mn}(r)=B_{mn}^j H_m^{(1)}(k_{rn}^j r)+C_{mn}^j H_m^{(2)}(k_{rn}^j r). \quad (9)$$

(III) Outermost ring, $r>r_s$,

$$R_{mn}(r)=D_{mn}H_m^{(1)}(k_{rn}^J r). \quad (10)$$

Here k_{rn}^j are the horizontal wavenumbers and A_{mn}, B_{mn}^j, C_{mn}^j, and D_{mn} are the coupling coefficients which are determined by applying the source condition and continuity conditions at the vertical sector boundaries. Although theoretically correct, this field representation is associated with serious numerical efficiency and stability issues.

(i) Because $H_m^{(1)}(x)$ and $H_m^{(2)}(x)$ become numerically linearly dependent for $m\gg|x|$, the linear systems for the coupling coefficients are unstable.[10,14]

(ii) When the order of the unnormalized Hankel functions becomes much greater than the argument, the absolute values of the Hankel functions diverge exponentially, leading to numerical overflow problems severely limiting the number of Fourier orders that can be handled.

(iii) The number of azimuthal modes required for convergence is at least $[k_0 r_s]$, where k_0 is the maximum wavenumber in the water column. Therefore, for sources far from the seamount in terms of wavelengths, the number of Fourier orders required for convergence becomes very large, which not only makes the numerical solution inefficient but also exacerbates the numerical overflow and stability issues described above.

Consequently, Taroudakis' approach is only applicable to small-scale 3D problems, characterized by low-frequency and ranges of only a few acoustic wavelengths.

2. Eskenazi's formulation

In Eskenazi's formulation,[9] the field in ring-shaped sector j is represented as

$$p^j(r,z,\phi) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} R_{mn}(r)\Psi_n^j(z)\Phi_m(\phi), \tag{11}$$

where $\Psi_n^j(z)$ and $\Phi_m(\phi)$ are the same vertical and azimuthal modes used by Taroudakis, but a different range solution $R_{mn}(r)$ is defined for the three regions as follows.

(I) Inner cylinder, $r \le r^1$,

$$R_{mn}(r) = b_{mn}^1 \hat{J}_{mn}^1(r). \tag{12}$$

(II) Ring j, $r^1 < r \le r_s$,

$$R_{mn}(r) = a_{mn}^j \hat{H}_{mn}^j(r) + b_{mn}^j \hat{J}_{mn}^j(r). \tag{13}$$

(III) Outermost ring, $r > r_s$,

$$R_{mn}(r) = c_{mn} \frac{H_m^{(1)}(k_{rn}^J r)}{H_m^{(1)}(k_{rn}^J r_s)}. \tag{14}$$

Here, $\hat{J}_{mn}^j$ and $\hat{H}_{mn}^j$ are the unconditionally stable normalized Bessel and Hankel functions:[14]

$$\hat{H}_{mn}^j(r) = \frac{H_m^{(1)}(k_{rn}^j r)}{H_m^{(1)}(k_{rn}^j r^{j-1})}, \tag{15a}$$

$$\hat{J}_{mn}^j(r) = J_m(k_{rn}^j r) H_m^{(1)}(k_{rn}^j r^j). \tag{15b}$$

As shown in Appendix A, the range solution in the outer ring is

$$R_{mn}(r) = \frac{i}{2}\pi \frac{\Psi_n^J(z_s)}{\rho(z_s)} \Phi_m(\phi_s) J_m(k_{rn}^J r_s) H_m^{(1)}(k_{rn}^J r) + a_{mn}^J \frac{H_m^{(1)}(k_{rn}^J r)}{H_m^{(1)}(k_{rn}^J r^{J-1})}. \tag{16}$$

Eskenazi then proceeded to set up the boundary conditions at all sector interfaces, which for each Fourier order m leads to a linear system of equations in the coefficients a_{mn}^j and b_{mn}^j, which is numerically stable for all orders. This formulation therefore successfully eliminates the numerical stability issue. However, the problem of overflow in the generation of the Hankel functions remains, and the dependence of the convergence of the solution on the source range remains as well.

3. Stable and efficient formulation

The present representation of the field differs from Taroudakis'[6] and Eskenazi's[9] in that it eliminates the virtual sector boundary at the source range and instead extends the outermost sector to the base of the seamount. The acoustic field in this sector is then represented as a superposition of the unperturbed field produced by the source in the absence of the seamount and a scattered field produced by the seamount. This basically reformulates the seamount propagation problem into a scattering problem. Scattering problems are traditionally handled using this fundamental superposition principle of decomposing the total field into its incident and scattered components. Although rather trivial, this reformulation has significant impact in terms of numerical efficiency, as demonstrated in the following.

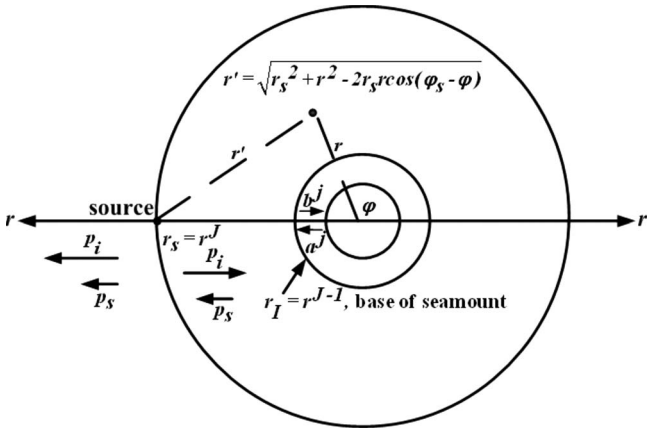


FIG. 2. Superposition principle for representing the field outside the seamount region.

The present approach uses the same stable field representation in each ring-shaped sector as applied by Eskenazi, differing only in the outer region containing the source, due to the use of the superposition principle, as follows.

(I) Inner cylinder $r \le r^1$,

$$p^1(r,z,\phi) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} b_{mn}^1 \hat{J}_{mn}^1(r) \Psi_n^1(z) \Phi_m(\phi). \tag{17}$$

(II) Intermediate sectors $r^1 < r \le r_I$, where r_I is the radius of the base of the seamount. In ring j, i.e., $r^{j-1} < r \le r^j$,

$$p^j(r,z,\phi) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} [a_{mn}^j \hat{H}_{mn}^j(r) + b_{mn}^j \hat{J}_{mn}^j(r)] \times \Psi_n^j(z) \Phi_m(\phi). \tag{18}$$

(III) Outer region containing source $r > r_I$,

$$p(r,z,\phi) = p_i(r',z) + \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} a_{mn}^J \hat{H}_{mn}^J(r) \Psi_n^J(z) \Phi_m(\phi), \tag{19}$$

where r' is the range of a field point with respect to the source (refer to Fig. 2),

$$r'(r,\phi) = \sqrt{r^2 + r_s^2 - 2rr_s\cos(\phi_s - \phi)}, \tag{20}$$

and $p_i(r',z)$ is the 2D normal mode solution,

$$p_i(r',z) = \frac{i}{4} \frac{1}{\rho(z_s)} \sum_{n=1}^{\infty} \Psi_n^J(z_s) \Psi_n^J(z) H_0^{(1)}(k_{rn}^J r'). \tag{21}$$

The depth and azimuthal modes, $\Psi_n^j(z)$ and $\Phi_m(\phi)$, are the same as those used in the previous formulations, and as

was the case for Eskenazi, the numerically stable normalized Bessel and Hankel functions introduced by Ricks *et al.*[14] are used, as defined in Eqs. (15a) and (15b).

In Sec. IV, the field results will be shown in terms of the traditional definition of the transmission loss (TL) in dB re 1 m,

$$\mathrm{TL}(r,z,\phi) = -20\log\left|\frac{p(r,z,\phi)}{p_0(r=1)}\right|, \tag{22}$$

where

$$|p_0(r)| = \frac{1}{4\pi r} \tag{23}$$

is the pressure modulus at a distance r from the source in free space.

III. NUMERICAL IMPLEMENTATION

Compared to the earlier formulations, there are three principal reasons for the improved numerical efficiency of the present formulation. The first is the use of the superposition principle for handling the source field. The second is the use of asymptotic representations of the normalized range functions. The third is the use of a marching solution of the coupled-mode equations, based on the single-scatter approximation.

A. Superposition

Using the earlier formulations of Eskenazi and Taroudakis, convergence for a field point outside the seamount, i.e., $r > r_I$, requires a number of azimuthal orders of at least $[k_0 r_s]$. In contrast, the use of the superposition principle reduces the minimum number of azimuthal orders to $[k_0 r_I]$, where r_I is the radius of the base of the conical seamount. Therefore, for a fixed source frequency, the number of azimuthal orders required depends only on the size of the seamount, not the source distance. For realistic seamount problems, this can lead to orders of magnitude in numerical efficiency.

B. Normalized range functions

As described earlier, both of the earlier formulations suffered from potential overflow and underflow problems in the numerical evaluation of the radial Bessel and Hankel functions. Although providing numerical stability of the solution for the modal coefficients, the use of the normalized Bessel and Hankel functions does not in itself solve this numerical problem for small and large arguments with respect to Fourier order. However, using the asymptotic representations for the Bessel and Hankel functions leads to closed form asymptotics for the normalized functions as well, therefore eliminating the overflow and underflow problems.

Therefore, the following asymptotic forms of Bessel and Hankel functions are here used to derive the asymptotic forms of the normalized Bessel and Hankel functions.[15]

(i) For $m \gg |x|$,

$$H_m^{(1)}(x) \sim -i\sqrt{\frac{2}{\pi m}}\left(\frac{ex}{2m}\right)^{-m}, \tag{24a}$$

$$J_m(x) \sim \frac{1}{\sqrt{2\pi m}}\left(\frac{ex}{2m}\right)^{m}. \tag{24b}$$

(ii) For $m \ll |x|$,

$$H_m^{(1)}(x) \sim \sqrt{\frac{2}{\pi x}}e^{i(x-(m/2)\pi-(\pi/4))}, \tag{25a}$$

$$J_m(x) \sim \sqrt{\frac{2}{\pi x}}\cos\left(x - \frac{m}{2}\pi - \frac{\pi}{4}\right). \tag{25b}$$

These asymptotic forms yield the following asymptotic forms of normalized range functions.[12]

(i) For $m \gg |X|$, $|X| \geq |x|$,

$$\frac{H_m^{(1)}(X)}{H_m^{(1)}(x)} \sim \left(\frac{x}{X}\right)^{m}, \tag{26a}$$

$$\frac{H_m^{(1)\prime}(X)}{H_m^{(1)}(x)} \sim -m\left(\frac{x}{X}\right)^{m}\frac{1}{X}, \tag{26b}$$

$$J_m(x)H_m^{(1)}(X) \sim -\frac{i}{m\pi}\left(\frac{x}{X}\right)^{m}, \tag{26c}$$

$$J_m'(x)H_m^{(1)}(X) \sim -\frac{i}{\pi}\left(\frac{x}{X}\right)^{m-1}\frac{1}{X}. \tag{26d}$$

(ii) For $m \gg |x|$, $|X| \geq |x|$,

$$\frac{H_m^{(1)}(X)}{H_m^{(1)}(x)} \sim \sqrt{\frac{x}{X}}e^{i(X-x)}, \tag{27a}$$

$$\frac{H_m^{(1)\prime}(X)}{H_m^{(1)}(x)} \sim i\sqrt{\frac{x}{X}}e^{i(X-x)}, \tag{27b}$$

$$J_m(x)H_m^{(1)}(X) \sim \frac{1}{\pi\sqrt{xX}}[e^{i(x+X-m\pi-(\pi/2))} + e^{-i(x-X)}], \tag{27c}$$

$$J_m'(x)H_m^{(1)}(X) \sim \frac{i}{\pi\sqrt{xX}}[e^{i(x+X-m\pi-(\pi/2))} - e^{-i(x-X)}]. \tag{27d}$$

C. Single-scatter coupled-mode solution

Rather than using a global matrix representation involving the mutual interaction between the sector interfaces, i.e., including multiple scattering effects, a more efficient single-scatter approximation is applied here, adapted from coupled-mode codes such as CSNAP.[11] We find from computational tests that this highly efficient approximation works well for most ocean acoustic propagation problems, where multiple scattering is in general negligible. A detailed discussion of the single-scatter approximation can be found in Refs. 16 and 8.

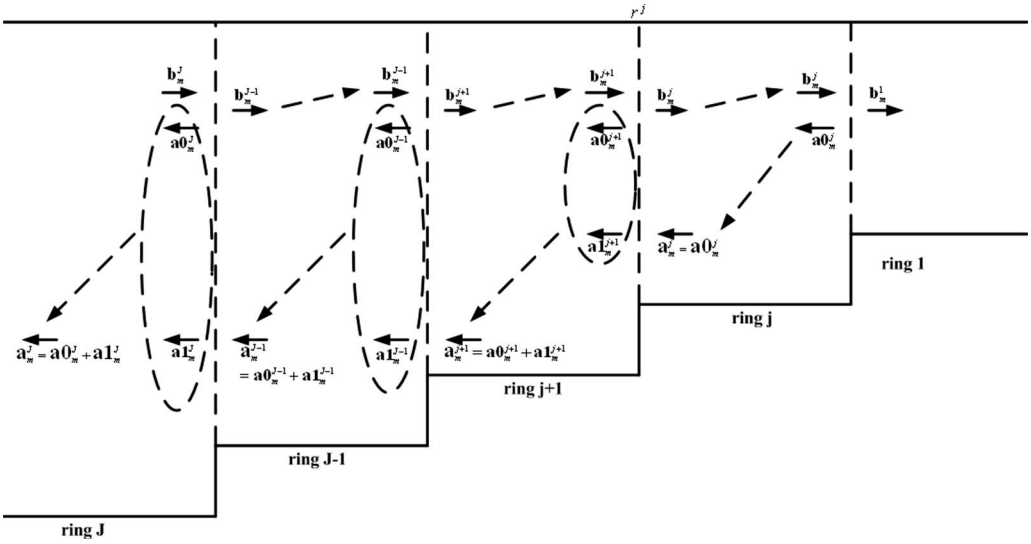


FIG. 3. Schematic representation of the two-way, single-scatter coupled-mode solution. For each azimuthal order, the modal amplitudes $\boldsymbol{b}_m^J$, arising from the decomposition of the source field into its Fourier-Bessel components, are incident on the outermost interface. By ignoring the backscattering from the next sector interface, the reflected $\boldsymbol{a}0_m^J$ and transmitted $\boldsymbol{b}_m^{J-1}$ modal amplitudes can then be found using the local, coupling matrices. The inward propagating modes are then used as incident field on the next sector interface, and the process is repeated until the innermost interface is reached. Here the process is reversed, marching outward, but at each sector interface the transmitted modal amplitudes $\boldsymbol{a}1_m^j$ are added to the reflected modal amplitudes $\boldsymbol{a}0_m^j$ computed during the inward marching.

The marching, single-scatter solution scheme is illustrated schematically in Fig. 3. For each azimuthal order, the modal amplitudes $\boldsymbol{b}_m^J$, arising from the decomposition of the source field into its Fourier–Bessel components, are incident on the outermost interface. By ignoring the backscattering from the next sector interface, the reflected $\boldsymbol{a}0_m^J$ and transmitted $\boldsymbol{b}_m^{J-1}$ modal amplitudes can then be found using the local, coupling matrices. The inward transmitted modes are then propagated across the sector and used as incident field on the next sector interface, and the process is repeated until the innermost interface is reached. Here the process is reversed, marching outward, but at each sector interface the transmitted modal amplitudes $\boldsymbol{a}1_m^j$ are added to the reflected modal amplitudes $\boldsymbol{a}0_m^j$ computed during the inward marching.

The first step in the marching solution technique is the decomposition of the source field, incident on the outermost sector interface at the range of the base of the seamount. Combining the general field expansion in Eq. (18) with the source sector field in Eq. (16), this corresponds to the expansion coefficients

$$b_{mn}^J = \frac{i}{2}\pi\frac{\Psi_n^J(z_s)}{\rho(z_s)}\Phi_m(\phi_s). \tag{28}$$

In principle this two-way marching solution could be used iteratively, converging to the exact results, but it is well established that for the weak range-dependence characteristic of the ocean waveguide, multiple scattering is insignificant, requiring only one two-way marching solution to converge.[8]

The advantage of this two-way coupling approach is that instead of solving a single large linear system arising from the DGM approach,[8] the problem is divided into multiple small linear systems. Because solving each of these small linear systems lowers the computational requirements considerably, this approach is applicable on personal computers, even for realistic seamount problems with many ring-shaped sectors being required.

1. Inward marching

For the inward marching, at sector interface number j, which is between ring $j+1$ and ring j, the modal expansion coefficients in ring j can be expressed in terms of the coefficients in ring $j+1$ as[12] (detailed in Appendix B)

$$\begin{bmatrix} \boldsymbol{b}_m^j \\ \boldsymbol{a}_m^j \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_{m4}^{j+1} & \boldsymbol{R}_{m3}^{j+1} \\ \boldsymbol{R}_{m2}^{j+1} & \boldsymbol{R}_{m1}^{j+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_m^{j+1} \\ \boldsymbol{a}_m^{j+1} \end{bmatrix}, \tag{29}$$

where

$$\boldsymbol{R}_{m1}^{j+1} = -(\boldsymbol{F}_b^j)^{-1}(\hat{\boldsymbol{J}}_m^{j\prime}\boldsymbol{C}_a^{j+1}\hat{\boldsymbol{H}}_m^{j+1} - \hat{\boldsymbol{J}}_m^j\boldsymbol{C}_b^{j+1}\hat{\boldsymbol{H}}_m^{j+1\prime}), \tag{30a}$$

$$\boldsymbol{R}_{m2}^{j+1} = -(\boldsymbol{F}_b^j)^{-1}(\hat{\boldsymbol{J}}_m^{j\prime}\boldsymbol{C}_a^{j+1}\hat{\boldsymbol{J}}_m^{j+1} - \hat{\boldsymbol{J}}_m^j\boldsymbol{C}_b^{j+1}\hat{\boldsymbol{J}}_m^{j+1\prime}), \tag{30b}$$

$$\boldsymbol{R}_{m3}^{j+1} = (\boldsymbol{F}_b^j)^{-1}(\hat{\boldsymbol{H}}_m^{j\prime}\boldsymbol{C}_a^{j+1}\hat{\boldsymbol{H}}_m^{j+1} - \hat{\boldsymbol{H}}_m^j\boldsymbol{C}_b^{j+1}\hat{\boldsymbol{H}}_m^{j+1\prime}), \tag{30c}$$

$$\boldsymbol{R}_{m4}^{j+1} = (\boldsymbol{F}_b^j)^{-1}(\hat{\boldsymbol{H}}_m^{j\prime}\boldsymbol{C}_a^{j+1}\hat{\boldsymbol{J}}_m^{j+1} - \hat{\boldsymbol{H}}_m^j\boldsymbol{C}_b^{j+1}\hat{\boldsymbol{J}}_m^{j+1\prime}), \tag{30d}$$

and $(\boldsymbol{F}_b^j)^{-1}$ is the diagonal matrix with entries

$$(\boldsymbol{F}_b^j)_{\nu\nu}^{-1} = \frac{\pi r^j}{2i}k_{r\nu}^j\frac{H_m^{(1)}(k_{r\nu}^j r^{j-1})}{H_m^{(1)}(k_{r\nu}^j r^j)}, \quad \nu = 1,2,\ldots,N. \tag{31}$$

The coupling matrices are

$$\boldsymbol{C}_{a\ \nu n}^{j+1} = \int_0^\infty \frac{1}{\rho^j(z)}\Psi_\nu^j(z)\Psi_n^{j+1}(z)dz, \quad \nu,n = 1,2,\ldots,N, \tag{32a}$$

$$\boldsymbol{C}_{b\ \nu n}^{j+1} = \frac{k_{rn}^{j+1}}{k_{r\nu}^{j}} \int_0^\infty \frac{1}{\rho^{j+1}(z)} \Psi_\nu^j(z) \Psi_n^{j+1}(z) dz,$$

$$\nu, n = 1, 2, \ldots, N. \tag{32b}$$

Here, $\boldsymbol{a}_m^j$, $\boldsymbol{b}_m^j$, $\boldsymbol{a}_m^{j+1}$, and $\boldsymbol{b}_m^{j+1}$ are the column vectors containing the expansion coefficients. $\hat{\boldsymbol{H}}_m^j$, $\hat{\boldsymbol{J}}_m^j$, $\hat{\boldsymbol{H}}_m^{j+1}$, and $\hat{\boldsymbol{J}}_m^{j+1}$ are diagonal matrices containing the range functions $\hat{H}_{mn}^j(r)$, $\hat{J}_{mn}^j(r)$, $\hat{H}_{mn}^{j+1}(r)$, and $\hat{J}_{mn}^{j+1}(r)$, $n=1,2,\ldots,N$, evaluated at $r=r^j$. Similarly, $\hat{\boldsymbol{H}}_m^{j\prime}$ and $\hat{\boldsymbol{J}}_m^{j\prime}$ are diagonal matrices of the derivatives of the range functions $\hat{H}_{mn}^j(r)$ and $\hat{J}_{mn}^j(r)$, with respect to $k_{rn}^j r$, $n=1,2,\ldots,N$, at $r=r^j$; $\hat{\boldsymbol{H}}_m^{j+1\prime}$ and $\hat{\boldsymbol{J}}_m^{j+1\prime}$ are diagonal matrices with diagonal entries as derivatives of the range functions $\hat{H}_{mn}^{j+1}(r)$ and $\hat{J}_{mn}^{j+1}(r)$, with respect to $k_{rn}^{j+1} r$, $n=1,2,\ldots,N$, at $r=r^j$. N denotes the number of normal modes.

In the single-scatter approximation, the outgoing field from the next sector boundary is ignored, i.e., $\boldsymbol{a}_m^j = \boldsymbol{0}$. Then from Eq. (29), it follows that

$$\boldsymbol{R}_{m2}^{j+1} \boldsymbol{b}_m^{j+1} + \boldsymbol{R}_{m1}^{j+1} \boldsymbol{a}_m^{j+1} = 0, \tag{33}$$

which leads to

$$\boldsymbol{a}_m^{j+1} = -(\boldsymbol{R}_{m1}^{j+1})^{-1} \boldsymbol{R}_{m2}^{j+1} \boldsymbol{b}_m^{j+1}. \tag{34}$$

With $\boldsymbol{a}_m^{j+1}$ computed by Eq. (34), $\boldsymbol{b}_m^j$ can be obtained from Eq. (29),

$$\boldsymbol{b}_m^j = \boldsymbol{R}_{m4}^{j+1} \boldsymbol{b}_m^{j+1} + \boldsymbol{R}_{m3}^{j+1} \boldsymbol{a}_m^{j+1}. \tag{35}$$

2. Outward marching

In the outward marching, the one-way approximation is again applied. By imposing the boundary conditions at the interface between ring $j+1$ and ring j, together with the one-way approximation $\boldsymbol{b}_m^{j+1}=0$ and $\boldsymbol{b}_m^j=0$, as shown in Appendix B, the following relation between the expansion coefficients in ring $j+1$ and those in ring j is obtained,[12]

$$\boldsymbol{a}_m^{j+1} = \boldsymbol{S}_{m1}^j \boldsymbol{a}_m^j, \tag{36}$$

where

$$\boldsymbol{S}_{m1}^j = (\boldsymbol{G}_a^{j+1})^{-1} \times [\hat{\boldsymbol{J}}_m^{j+1\prime} \boldsymbol{C}_c^j \hat{\boldsymbol{H}}_m^j - \hat{\boldsymbol{J}}_m^{j+1} \boldsymbol{C}_d^j \hat{\boldsymbol{H}}_m^{j\prime}], \tag{37}$$

and $(\boldsymbol{G}_a^{j+1})^{-1}$ is the diagonal matrix with entries

$$(\boldsymbol{G}_a^{j+1})_{\nu\nu}^{-1} = i \frac{\pi r^j}{2} k_{r\nu}^{j+1} \frac{H_m^{(1)}(k_{r\nu}^{j+1} r^j)}{H_m^{(1)}(k_{r\nu}^{j+1} r^{j+1})}, \quad \nu = 1, 2, \ldots, N. \tag{38}$$

The coupling matrices are

$$\boldsymbol{C}_{c\nu n}^j = \int_0^\infty \frac{1}{\rho^{j+1}(z)} \Psi_\nu^{j+1}(z) \Psi_n^j(z) dz, \quad \nu, n = 1, 2, \ldots, N, \tag{39a}$$

$$\boldsymbol{C}_{d\nu n}^j = \frac{k_{rn}^j}{k_{r\nu}^{j+1}} \int_0^\infty \frac{1}{\rho^j(z)} \Psi_\nu^{j+1}(z) \Psi_n^j(z) dz, \quad \nu, n = 1, 2, \ldots, N. \tag{39b}$$

It should be noted that the coupling matrices, which are expressed in Eqs. (32a), (32b), (39a), and (39b), are independent of azimuthal orders, so they can be precalculated only once.

D. Numerical stability

Although the theoretical formulation used here is inherently numerically stable, the use of the single-scatter approximation adds a source of instability, equivalent to that encountered in the classical propagator matrix approach to solving the depth-separated wave equation for stratified media. The issue is that the marching solution is carried through regions where the modal solutions are evanescent, and care must be taken not to propagate and magnify truncation errors through such sectors.[8]

1. Inward marching

a. Evaluation of $\boldsymbol{R}_{m3}^{j+1}$ and $\boldsymbol{R}_{m4}^{j+1}$ The evaluation of $(\boldsymbol{F}_b^j)^{-1}$ in Eq. (31) is unstable for $m \gg |k_{r\nu}^j r^j|$ and $m \ll |k_{r\nu}^j r^j|$, because from Eq. (31) it can be seen that $(\boldsymbol{F}_b^j)_{\nu\nu}^{-1}$ depends on $H_m^{(1)}(k_{r\nu}^j r^{j-1}) / H_m^{(1)}(k_{r\nu}^j r^j)$, which is diverging for $m \gg |k_{r\nu}^j r^j|$ and $m \ll |k_{r\nu}^j r^j|$. $(H_m^{(1)}(k_{r\nu}^j r^j) / H_m^{(1)}(k_{r\nu}^j r^{j-1})$ is properly normalized and therefore stable.)

However, the products $(\boldsymbol{F}_b^j)^{-1} \hat{\boldsymbol{H}}_m^{j\prime}$ and $(\boldsymbol{F}_b^j)^{-1} \hat{\boldsymbol{H}}_m^j$ in Eqs. (30c) and (30d) can be evaluated together to avoid evaluating $(\boldsymbol{F}_b^j)^{-1}$, and in this way stable results of $\boldsymbol{R}_{m3}^{j+1}$ and $\boldsymbol{R}_{m4}^{j+1}$ can be obtained. With the following expressions,

$$(\boldsymbol{F}_b^j)_{\nu\nu}^{-1} = \frac{\pi r^j}{2i} k_{r\nu}^j \frac{H_m^{(1)}(k_{r\nu}^j r^{j-1})}{H_m^{(1)}(k_{r\nu}^j r^j)}, \tag{40}$$

$$(\hat{\boldsymbol{H}}_m^j)_{\nu\nu} = \frac{H_m^{(1)}(k_{r\nu}^j r^j)}{H_m^{(1)}(k_{r\nu}^j r^{j-1})}, \tag{41}$$

$$(\hat{\boldsymbol{H}}_m^{j\prime})_{\nu\nu} = \frac{H_m^{(1)\prime}(k_{r\nu}^j r^j)}{H_m^{(1)}(k_{r\nu}^j r^{j-1})}, \tag{42}$$

the product $(\boldsymbol{F}_b^j)^{-1} \hat{\boldsymbol{H}}_m^j$ can be expressed as the diagonal matrix with entries

$$[(\boldsymbol{F}_b^j)^{-1} \hat{\boldsymbol{H}}_m^j]_{\nu\nu} = \frac{\pi r^j}{2i} k_{r\nu}^j, \quad \nu = 1, 2, \ldots, N, \tag{43}$$

and the product $(\boldsymbol{F}_b^j)^{-1} \hat{\boldsymbol{H}}_m^{j\prime}$ can be expressed as the diagonal matrix with entries

$$[(\boldsymbol{F}_b^j)^{-1} \hat{\boldsymbol{H}}_m^{j\prime}]_{\nu\nu} = \frac{\pi r^j}{2i} k_{r\nu}^j \frac{H_m^{(1)\prime}(k_{r\nu}^j r^j)}{H_m^{(1)}(k_{r\nu}^j r^j)}, \quad \nu = 1, 2, \ldots, N. \tag{44}$$

The products $(\boldsymbol{F}_b^j)^{-1} \hat{\boldsymbol{H}}_m^j$ and $(\boldsymbol{F}_b^j)^{-1} \hat{\boldsymbol{H}}_m^{j\prime}$ in Eqs. (43) and (44) are stable, and therefore by substituting Eqs. (43) and (44) back into Eqs. (30c) and (30d), stable $\boldsymbol{R}_{m3}^{j+1}$ and $\boldsymbol{R}_{m4}^{j+1}$ can be obtained.

b. Evaluation of $(\boldsymbol{R}_{m1}^{j+1})^{-1} \boldsymbol{R}_{m2}^{j+1}$ As stated above, the evaluation of $(\boldsymbol{F}_b^j)^{-1}$ in Eq. (31) is unstable for large m. However, from Eq. (34), the product $(\boldsymbol{R}_{m1}^{j+1})^{-1} \boldsymbol{R}_{m2}^{j+1}$ can be evaluated to avoid evaluating $(\boldsymbol{F}_b^j)^{-1}$ in obtaining $\boldsymbol{a}_m^{j+1}$. From Eqs. (30a) and (30b), it can be determined that

$$(\boldsymbol{R}_{m1}^{j+1})^{-1}\boldsymbol{R}_{m2}^{j+1} = (\hat{\boldsymbol{J}}_m^{j\prime}\boldsymbol{C}_a^{j+1}\hat{\boldsymbol{H}}_m^{j+1} - \hat{\boldsymbol{J}}_m^{j}\boldsymbol{C}_b^{j+1}\hat{\boldsymbol{H}}_m^{j+1\prime})^{-1}\boldsymbol{F}_b^{j}$$
$$\times (\boldsymbol{F}_b^{j})^{-1}(\hat{\boldsymbol{J}}_m^{j\prime}\boldsymbol{C}_a^{j+1}\hat{\boldsymbol{J}}_m^{j+1} - \hat{\boldsymbol{J}}_m^{j}\boldsymbol{C}_b^{j+1}\hat{\boldsymbol{J}}_m^{j+1\prime})$$
$$= (\hat{\boldsymbol{J}}_m^{j\prime}\boldsymbol{C}_a^{j+1}\hat{\boldsymbol{H}}_m^{j+1} - \hat{\boldsymbol{J}}_m^{j}\boldsymbol{C}_b^{j+1}\hat{\boldsymbol{H}}_m^{j+1\prime})^{-1}$$
$$\times(\hat{\boldsymbol{J}}_m^{j\prime}\boldsymbol{C}_a^{j+1}\hat{\boldsymbol{J}}_m^{j+1} - \hat{\boldsymbol{J}}_m^{j}\boldsymbol{C}_b^{j+1}\hat{\boldsymbol{J}}_m^{j+1\prime}). \quad (45)$$

The product $(\boldsymbol{R}_{m1}^{j+1})^{-1}\boldsymbol{R}_{m2}^{j+1}$ evaluated by Eq. (45) is stable, so $\boldsymbol{a}_m^{j+1}$ computed by Eq. (34) is also stable.

2. Outward marching

Again, for large m the evaluation of $(\boldsymbol{G}_a^{j+1})^{-1}$ in Eq. (38) is unstable, because $H_m^{(1)}(k_{r\nu}^{j+1}r^j)/H_m^{(1)}(k_{r\nu}^{j+1}r^{j+1})$ is diverging. Therefore, it should be avoided to compute $(\boldsymbol{G}_a^{j+1})^{-1}$ in obtaining $\boldsymbol{a}_m^{j+1}$ by Eqs. (36) and (37). Instead, the bounded products $(\boldsymbol{G}_a^{j+1})^{-1}\hat{\boldsymbol{J}}_m^{j+1}$ and $(\boldsymbol{G}_a^{j+1})^{-1}\hat{\boldsymbol{J}}_m^{j+1\prime}$ can be expressed as

$$(\boldsymbol{G}_a^{j+1})_{\nu\nu}^{-1}(\hat{\boldsymbol{J}}_m^{j+1})_{\nu\nu} = i\frac{\pi r^j}{2}k_{r\nu}^{j+1}J_m(k_{r\nu}^{j+1}r^j)H_m^{(1)}(k_{r\nu}^{j+1}r^j),$$
$$\nu = 1,2,\ldots,N, \quad (46)$$

$$(\boldsymbol{G}_a^{j+1})_{\nu\nu}^{-1}(\hat{\boldsymbol{J}}_m^{j+1\prime})_{\nu\nu} = i\frac{\pi r^j}{2}k_{r\nu}^{j+1}J_m'(k_{r\nu}^{j+1}r^j)H_m^{(1)}(k_{r\nu}^{j+1}r^j),$$
$$\nu = 1,2,\ldots,N. \quad (47)$$

Because the terms $J_m(k_{r\nu}^{j+1}r^j)H_m^{(1)}(k_{r\nu}^{j+1}r^j)$ and $J_m'(k_{r\nu}^{j+1}r^j)H_m^{(1)}\times(k_{r\nu}^{j+1}r^j)$ are bounded and their evaluation is numerically stable, Eqs. (46) and (47) are appropriate in computing $\boldsymbol{S}_{m1}^{j}$ by Eq. (37).

IV. NUMERICAL RESULTS

The new spectral coupled-mode model is applied to two numerical examples. First, the examples are used to illustrate how the azimuthal convergence is improved considerably compared to the earlier models. Second, the examples are used to illustrate how the 3D effects depend on the environment by comparing the results obtained by the new 3D model and the classical $N\times$2D model.

Perkins and Baer[3] introduced the $N\times$2D approach to solve N 2D problems, one for each of N vertical planes passing through the point source. By combining those results, an approximate 3D solution is obtained, ignoring out-of-plane scattering. Therefore, the $N\times$2D approach does not have the ability to treat the azimuthal coupling. In their work,[3] parabolic-equation programs were applied to solve each 2D problem.

CSNAP[11] is an efficient coupled-mode model for 2D range-dependent problems, either in cylindrical geometry with a point source, or in plane geometry with a line source. Here, CSNAP is applied to obtain the $N\times$2D results and for computing the vertical mode functions and coupling matrices for the 3D coupled-mode formulation, eliminating any differences in that regard in comparing the 2 and 3D computations.

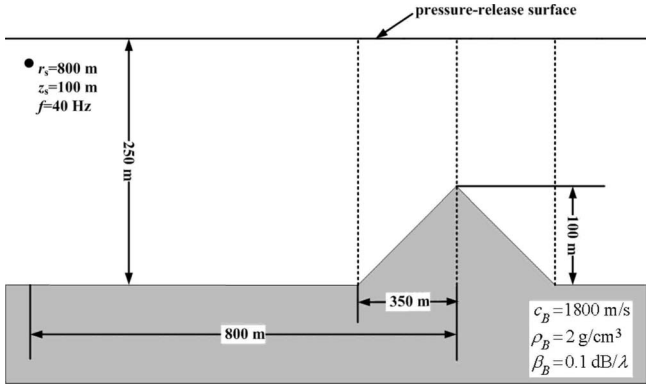


FIG. 4. Schematic of a shallow water waveguide with a conical seamount and a penetrable bottom.

A. Conical seamount in shallow water

The schematic of the problem is shown in Fig. 4. The waveguide consists of a 250 m water column bounded above by a pressure-release surface and below by a penetrable bottom. The water column is isovelocity with c_w=1500 m/s. The seamount is 100 m high, with a radius at the base of 350 m, and has the same acoustic properties as the bottom. A 40 Hz point source is at a depth of 100 m, and at a distance of 800 m from the axis of the seamount. The properties of the bottom are sound speed c_B=1800 m/s, attenuation β_B=0.1 dB/λ, and density ρ_B=2.0 g/cm^3. A false bottom at depth 4000 m is introduced to take into account the contribution from the continuous spectrum.

In this shallow water problem, 179 normal modes are included in mode coupling, and 59 azimuthal modes are used for the purpose of azimuthal convergence.

1. Azimuthal convergence

In Sec. II B 3, it was shown that the number of azimuthal modes required for convergence is $M \geq [k_0 r_I]$, where r_I is the radius of the base of a seamount. In this case, $[k_0 r_I]$=59.

Figure 5 depicts TL versus range along azimuthal angle ϕ=0 with respect to the source, at depth 100 m, with different number of azimuthal modes, showing that M=59 is sufficient to reach convergence in this case. In addition, the 2D result is also included in Fig. 5, which shows that in this example, along this particular azimuthal angle, the 2D result is in great agreement with the 3D result.

2. Significance of 3D effects

To compare the results of the $N\times$2D model and those of the 3D model introduced in this paper, the height of the seamount is set to vary from 25 to 100 m. Results of TL in the horizontal plane at depth 100 m are shown in Fig. 6.

From Figs. 6(a) and 6(b), it can be seen that when the height of the seamount is 25 m, which is relatively small, the 3D effects are insignificant and the $N\times$2D model is a good approximation of the 3D seamount model. As the height of the seamount rises to 100 m, from the results shown in Figs. 6(c) and 6(d), it can be seen that the approximation of the $N\times$2D model to the 3D seamount model deteriorates, and

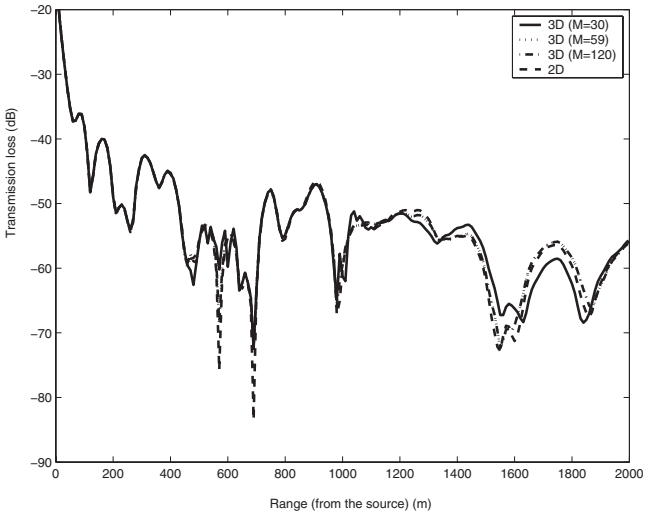


FIG. 5. TL vs range for propagation across a 100 m high seamount, 800 m from the source. The source frequency is 40 Hz. To investigate the convergence of the azimuthal series, M is set to vary from 15 to 120. In this case, $M=[k_0 r_1]=59$ appears sufficient for convergence.

the span of the shadow zone in the 3D result is more pronounced than in the $N\times 2$D result. This example indicates that the azimuthal coupling can be important for seamount problems, and the broadening of shadows behind seamounts by horizontal refraction can be captured by implementing a true 3D model. The phenomenon of increased shadow zone can be explained from a ray point of view.[17] A sloping bottom generally causes the horizontal wavenumber of a normal mode and therefore the index of refraction for that mode to decrease toward shallow water. As a result, the acoustic ray bends away from shallow water toward deep water. In addition, Fig. 6 shows that the 3D and the $N\times 2$D backscattered fields are not significantly different in this shallow water seamount problem.

B. Deep water seamount problem

In this more realistic problem which is taken from Ref. 11, the waveguide consists of a 5000 m inhomogeneous water column limited above by a pressure-release flat sea surface and below by a homogeneous fluid half space with a compressional speed of 2000 m/s, a density of 1 g/cm^3, and an attenuation of 0.1 dB/λ. A false bottom at depth 7000 m is introduced in this deep water problem. The seamount is 100 km from the source, with a radius of the base of 20 km, and has the same acoustic properties as the bottom. Two different heights of the seamount are considered, H

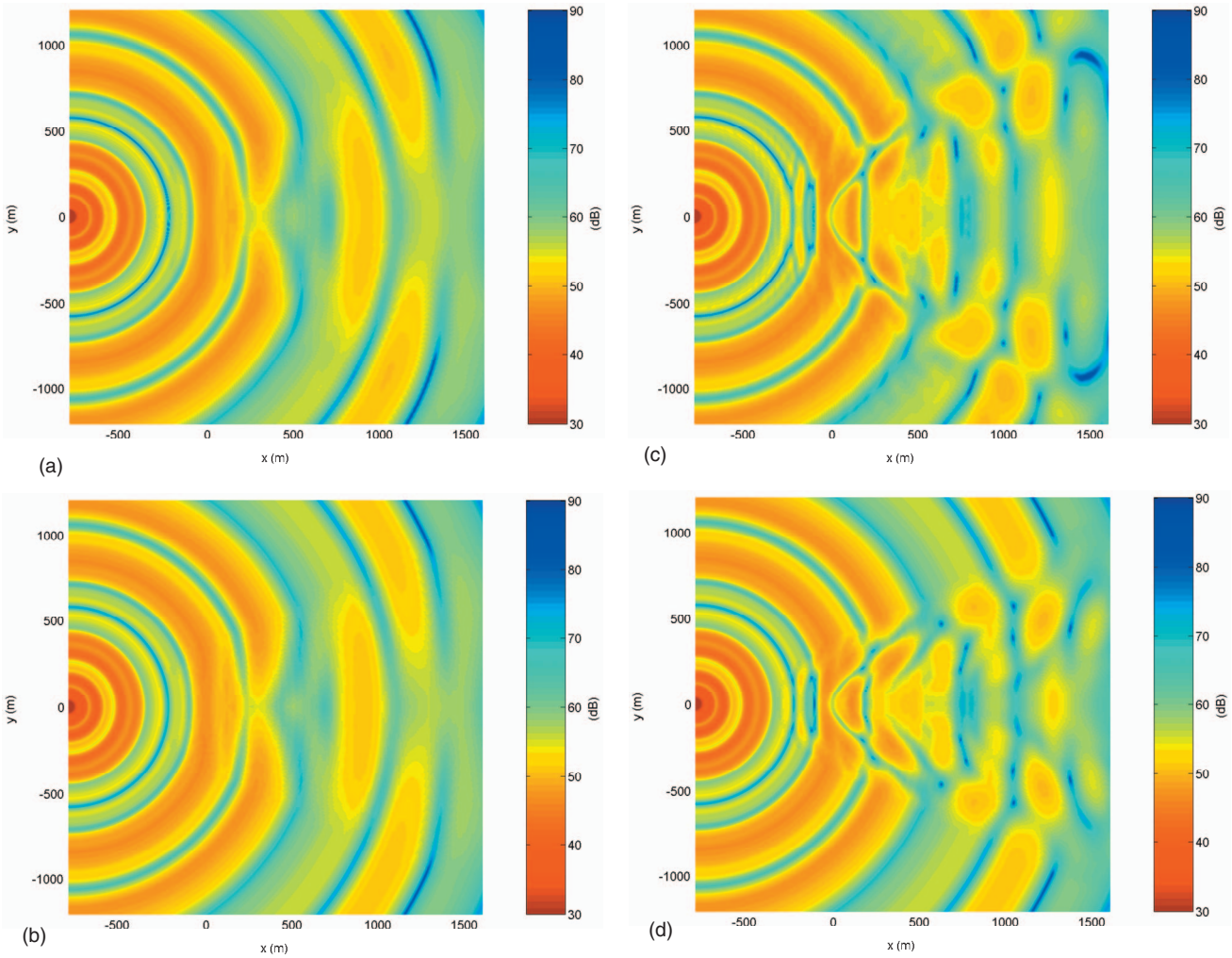


FIG. 6. TL in the horizontal plane at depth 100 m for two different seamount heights, $H=25$ m and $H=100$ m, respectively. The source frequency is 40 Hz. (a) $H=25$ m, 3D result, (b) $H=25$ m, $N\times 2$D result; (c) $H=100$ m, 3D result, (d) $H=100$ m, $N\times 2$D result.

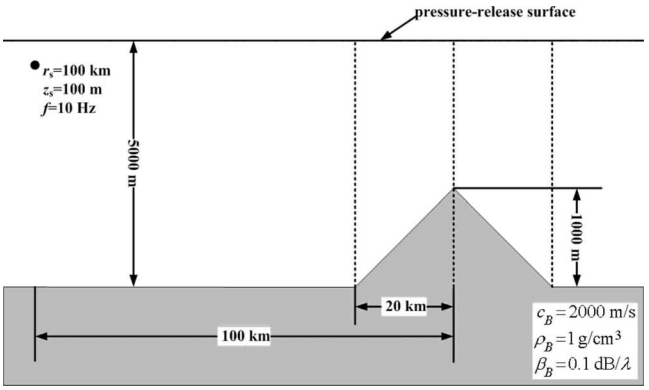


FIG. 7. Schematic of a deep water waveguide with a conical seamount and a penetrable bottom.

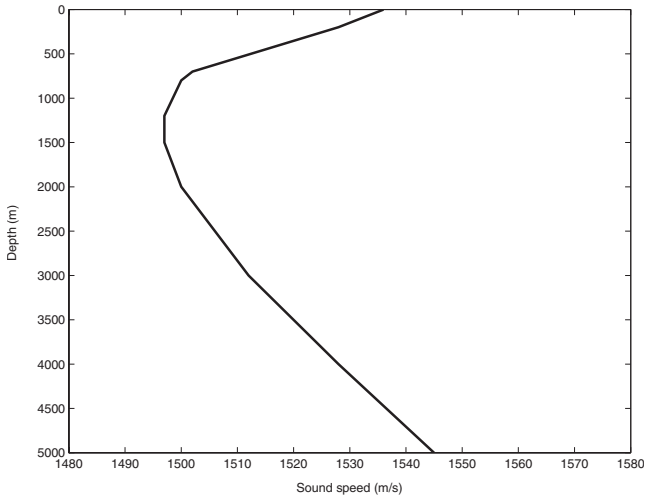


FIG. 8. The sound speed profile in the conical seamount deep water problem.

=1000 m and H=3800 m, respectively. The source depth is 100 m and the source frequency is 10 Hz. The schematic of this problem is shown in Fig. 7. In addition, the sound speed profile is shown in Fig. 8.

In this deep water problem, the modal starting field is limited to contain only the waterborne modes, which are the first 12 modes; otherwise a maximum of 74 modes are used. In addition, for the purpose of azimuthal convergence, the number of azimuthal modes is taken to be 898.

In order to give a clear idea about the ensonification in the waveguide in the presence of a seamount, the 2D TL versus depth and range at azimuth 0° with respect to the source is presented in Fig. 9. This figure shows that a beam of energy cycles up and down in the water column until it hits the seamount and reflects at steeper angles.

Figure 10 compares the 3D and $N\times$2D results for the two different heights of the seamount. Figures 10(a) and 10(b) show the TL, computed by the 3D model and $N\times$2D model, respectively, in the horizontal plane at depth 300 m for a seamount of height 1000 m. These results show that for seamounts with small slope angles and not reaching into the sound channel, the 3D effects are relatively insignificant and the $N\times$2D model provides a good approximation. Figures 10(c) and 10(d) show the corresponding results for a large seamount of height 3800 m. Here the 3D effect is rather obvious. Therefore, it can be seen that for seamounts such as this, with steep slopes, the out-of-plane scattering is significant, providing a much deeper shadow zone than predicted by the $N\times$2D model ignoring such effects. In addition, Fig. 10 shows that the 3D and the $N\times$2D backscattered fields are not significantly different in this deep water seamount problem.

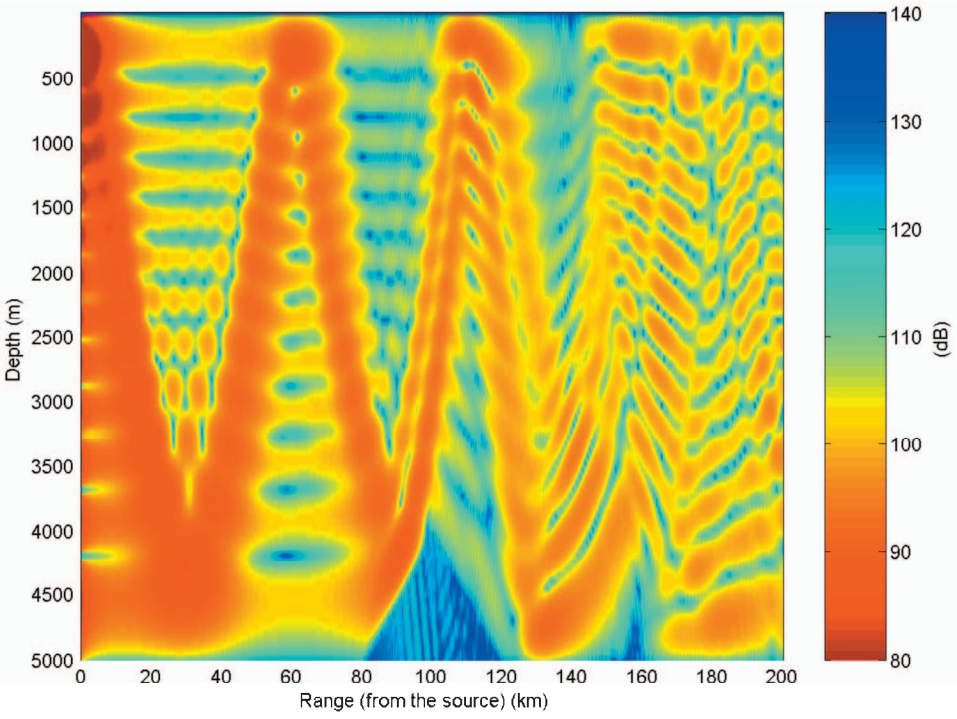


FIG. 9. 2D solution for propagation at 10 Hz in a vertical plane across the peak of the seamount.

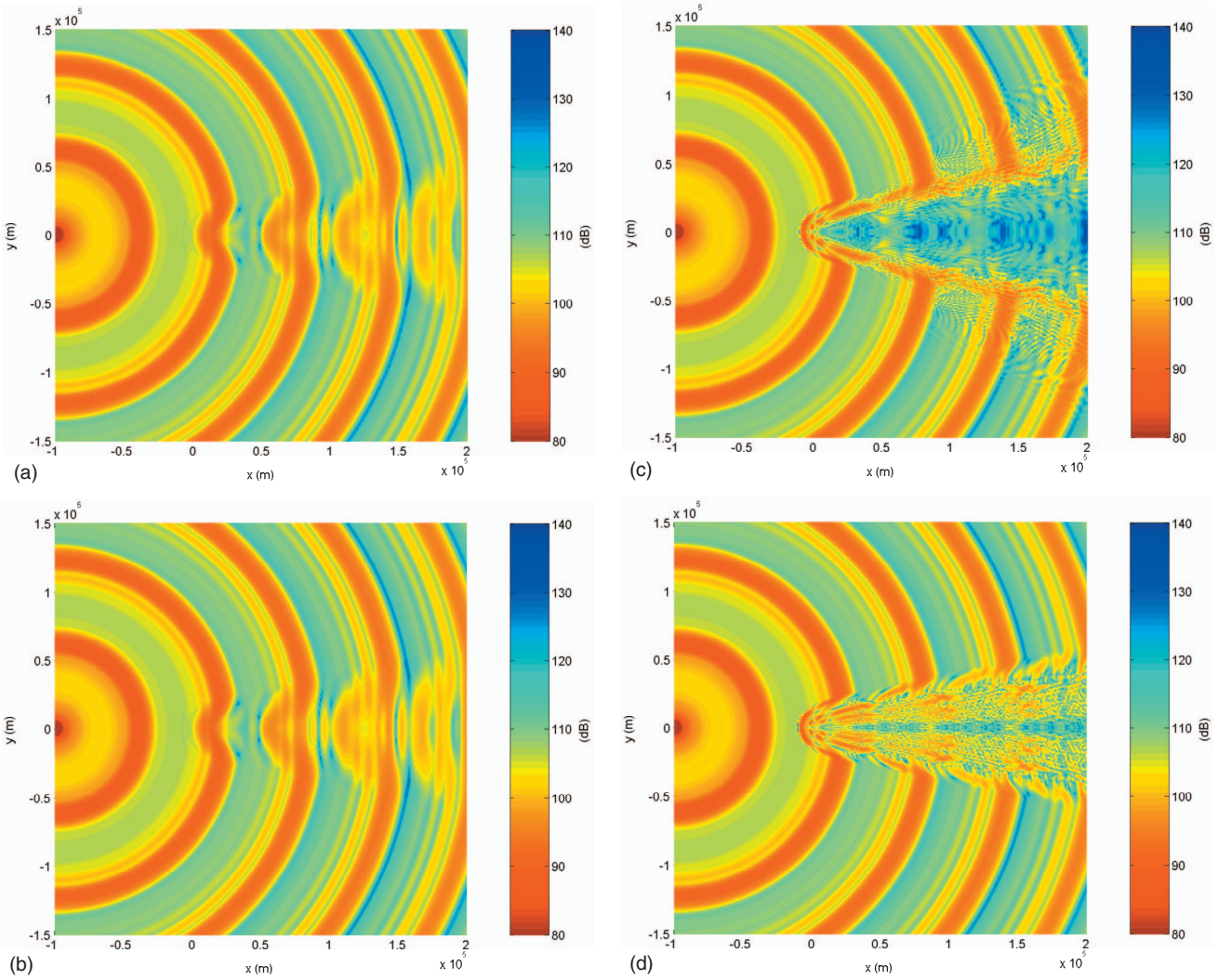


FIG. 10. TL in the horizontal plane at depth 300 m. The source frequency is 10 Hz. (a) $H=1000$ m, 3D result, (b) $H=1000$ m, $N\times 2$D result; (c) $H=3800$ m, 3D result, (d) $H=3800$ m, $N\times 2$D result.

C. Discussion

In the numerical examples presented in this paper, a pair of seamounts are handled at low frequencies. Further investigation shows that for a fixed seamount height, as the source frequency rises, the 3D effects become more significant. In addition, although the source frequencies are quite low in these numerical examples, the present approach is practical for solving 3D seamount problems with the nondimensional wavenumber $[k_0 r_I]$ up to $[k_0 r_I]\approx 10\,000$, on personal computers. For 3D seamount problems at higher frequencies, the present approach should be parallelized and run on supercomputers.

V. CONCLUSIONS

The coupled normal mode method is a very efficient, simple and accurate method to solve range-dependent problems. First, since the normal mode method solves the full-wave Helmholtz equation, it is applicable to cases where the backscattering is not negligible. Second, it is free of angular limitations for propagation angles less than critical, and therefore provides accurate solutions at long ranges. This is achieved by including evanescent modes by introducing a false bottom or by calculating the branch-line integral.[8,19] Third, as an important feature of the spectral coupled-mode approach, the coupling matrices are independent of azimuthal orders, which makes this approach ameliorable to problems of scattering from azimuthal symmetric features.

Here a spectral coupled-mode model has been developed and applied to analyze the propagation and scattering around conical seamounts. Although based on a theoretical foundation identical to that used in earlier related work, the present approach provides more efficient and stable numerical solutions for realistic seamount problems. The keys to its superior performance are the fundamental superposition principle applied to introduce the source, the choice of normalized range solutions which can be evaluated accurately and remain numerically stable even at high azimuthal orders, and the use of a marching, single-scatter approximation for the solution of the coupled-mode equations.

The numerical examples show that azimuthal coupling can be important for seamount problems. When the azimuthal variation is weak, the out-of-plane scattering is negligible and therefore, as shown by the results, the $N\times 2$D

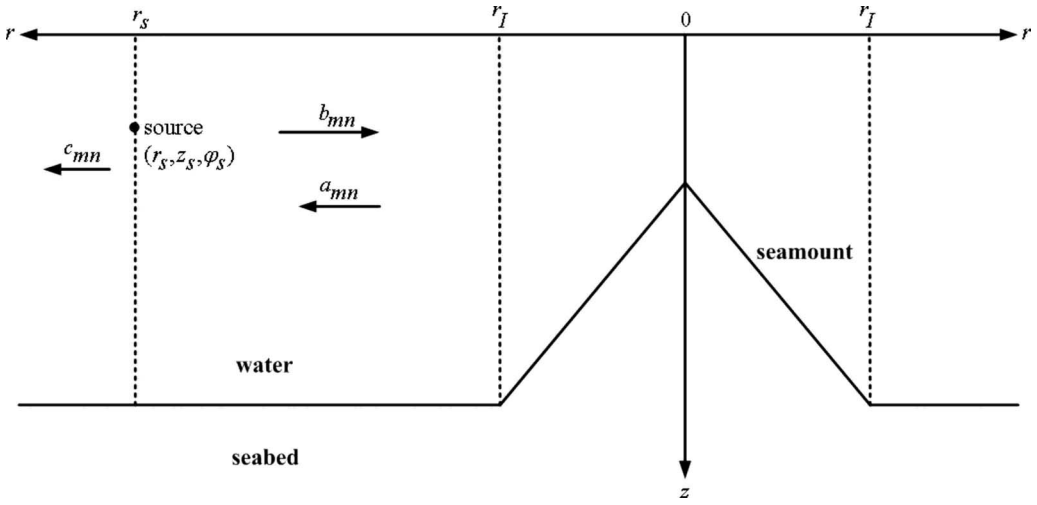


FIG. 11. Schematic of the field in the external region of a seamount.

approach is a good approximation of the true 3D approach. However, when the azimuthal variation is strong, azimuthal coupling cannot be ignored. In this case, the shadows behind seamounts are broadened by horizontal refraction. A true 3D approach, instead of the $N\times 2$D approach, can capture this feature. The dependence of the 3D effects on the environment is also illustrated through the numerical examples, which show that for seamounts with steep slopes and reaching far into the sound channel in terms of wavelengths, the 3D effects are significant. In a situation such as this, true 3D models are required, rather than $N\times 2$D models.

APPENDIX A: SOURCE CONDITION AND RANGE SOLUTION IN THE EXTERNAL REGION OF A SEAMOUNT

In this paper, in the external region of a seamount, the inhomogeneous 3D Helmholtz equation with a point source is

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2 p}{\partial \phi^2}+\rho(z)\frac{\partial}{\partial z}\left(\frac{1}{\rho(z)}\frac{\partial p}{\partial z}\right)+\frac{\omega^2}{c^2(z)}p = -\frac{\delta(r-r_s)}{r}\delta(z-z_s)\delta(\phi-\phi_s), \tag{A1}$$

and the solution is represented as

$$p(r,z,\phi)=\sum_{m=0}^{\infty}\sum_{n=1}^{\infty}R_{mn}(r)\Psi_n(z)\Phi_m(\phi). \tag{A2}$$

1. Equation for $R_{mn}(r)$

Substituting Eq. (A2) into Eq. (A1), and making use of the modal equation

$$\rho(z)\frac{d}{dz}\left(\frac{1}{\rho(z)}\frac{d\Psi_n(z)}{dz}\right)+\left(\frac{\omega^2}{c^2(z)}-k_{rn}^2\right)\Psi_n(z)=0, \tag{A3}$$

and the equation

$$\frac{d^2\Phi_m(\phi)}{d\phi^2}+m^2\Phi_m(\phi)=0,\quad m=0,1,2,\ldots, \tag{A4}$$

yield the following modal form of the Helmholtz equation,

$$\sum_{m=0}^{\infty}\sum_{n=1}^{\infty}\left[\frac{1}{r}\frac{d}{dr}\left(r\frac{dR_{mn}(r)}{dr}\right)+\left(k_{rn}^2-\frac{m^2}{r^2}\right)R_{mn}(r)\right]\Psi_n(z)\Phi_m(\phi) = -\frac{\delta(r-r_s)}{r}\delta(z-z_s)\delta(\phi-\phi_s). \tag{A5}$$

By applying the operator

$$\int_{-\pi}^{\pi}\Phi_\mu(\phi)(\cdot)d\phi \tag{A6}$$

and the operator

$$\int_0^{\infty}\frac{1}{\rho(z)}\Psi_\nu(z)(\cdot)dz \tag{A7}$$

to Eq. (A5), we obtain the equation for $R_{mn}(r)$,

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dR_{mn}(r)}{dr}\right)+\left(k_{rn}^2-\frac{m^2}{r^2}\right)R_{mn}(r) = -\frac{\delta(r-r_s)}{r}\frac{\Psi_n(z_s)}{\rho(z_s)}\Phi_m(\phi_s). \tag{A8}$$

2. Solution for $R_{mn}(r)$

In the external region of a seamount, as shown in Fig. 11, $R_{mn}(r)$ is represented as follows.

(a) $r_I\le r<r_s$,

$$R_{mn}(r)=a_{mn}\frac{H_m^{(1)}(k_{rn}r)}{H_m^{(1)}(k_{rn}r_I)}+b_{mn}J_m(k_{rn}r)H_m^{(1)}(k_{rn}r_s). \tag{A9}$$

(b) $r\ge r_s$,

$$R_{mn}(r)=c_{mn}\frac{H_m^{(1)}(k_{rn}r)}{H_m^{(1)}(k_{rn}r_s)}. \tag{A10}$$

By imposing continuity of pressure at $r=r_s$, we obtain

$$a_{mn}\frac{H_m^{(1)}(k_{rn}r_s)}{H_m^{(1)}(k_{rn}r_I)}+b_{mn}J_m(k_{rn}r_s)H_m^{(1)}(k_{rn}r_s)=c_{mn}. \tag{A11}$$

Integrating Eq. (A8) over $[r_s^-,r_s^+]$ gives

$$\left.\frac{dR_{mn}(r)}{dr}\right|_{r_s^-}^{r_s^+}=-\frac{1}{r_s}\frac{\Psi_n(z_s)}{\rho(z_s)}\Phi_m(\phi_s). \tag{A12}$$

By substituting Eqs. (A9) and (A10) into Eq. (A12), we obtain

$$c_{mn}\frac{H_m^{(1)\prime}(k_{rn}r_s)}{H_m^{(1)}(k_{rn}r_s)}-a_{mn}\frac{H_m^{(1)\prime}(k_{rn}r_s)}{H_m^{(1)}(k_{rn}r_I)}-b_{mn}J_m'(k_{rn}r_s)H_m^{(1)}(k_{rn}r_s)$$
$$=-\frac{1}{r_s}\frac{\Psi_n(z_s)}{\rho(z_s)}\Phi_m(\phi_s). \tag{A13}$$

By solving the linear system of Eqs. (A11) and (A13), we obtain

$$b_{mn}=\frac{i}{2}\pi\frac{\Psi_n(z_s)}{\rho(z_s)}\Phi_m(\phi_s), \tag{A14}$$

and

$$c_{mn}=\frac{i}{2}\pi\frac{\Psi_n(z_s)}{\rho(z_s)}\Phi_m(\phi_s)J_m(k_{rn}r_s)H_m^{(1)}(k_{rn}r_s)$$
$$+a_{mn}\frac{H_m^{(1)}(k_{rn}r_s)}{H_m^{(1)}(k_{rn}r_I)}. \tag{A15}$$

Therefore, the range solution for $r\geq r_s$ is

$$R_{mn}(r)=\frac{i}{2}\pi\frac{\Psi_n(z_s)}{\rho(z_s)}\Phi_m(\phi_s)J_m(k_{rn}r_s)H_m^{(1)}(k_{rn}r)$$
$$+a_{mn}\frac{H_m^{(1)}(k_{rn}r)}{H_m^{(1)}(k_{rn}r_I)}, \tag{A16}$$

and the range solution for $r_I\leq r<r_s$ is

$$R_{mn}(r)=\frac{i}{2}\pi\frac{\Psi_n(z_s)}{\rho(z_s)}\Phi_m(\phi_s)J_m(k_{rn}r)H_m^{(1)}(k_{rn}r_s)$$
$$+a_{mn}\frac{H_m^{(1)}(k_{rn}r)}{H_m^{(1)}(k_{rn}r_I)}. \tag{A17}$$

APPENDIX B: MODE COUPLING BETWEEN TWO NEIGHBORING SECTORS

The coupled-mode equations in this section are derived by following Evans[20] and Jensen *et al.*[8]

1. Mode coupling in inward marching

Consider interface j between two neighboring sectors, sector j and sector $j+1$, by substituting Eq. (18) into the boundary condition

$$p^j(r^j,z,\phi)=p^{j+1}(r^j,z,\phi), \tag{B1}$$

we obtain

$$\sum_{n=1}^{\infty}[a_{mn}^j\hat{H}_{mn}^j(r^j)+b_{mn}^j\hat{J}_{mn}^j(r^j)]\Psi_n^j(z)$$
$$=\sum_{n=1}^{\infty}[a_{mn}^{j+1}\hat{H}_{mn}^{j+1}(r^j)+b_{mn}^{j+1}\hat{J}_{mn}^{j+1}(r^j)]\Psi_n^{j+1}(z). \tag{B2}$$

By applying the operator

$$\int_0^{\infty}\frac{1}{\rho^j(z)}\Psi_\nu^j(z)(\cdot)dz \tag{B3}$$

to Eq. (B2), we have

$$[a_{m\nu}^j\hat{H}_{m\nu}^j(r^j)+b_{m\nu}^j\hat{J}_{m\nu}^j(r^j)]$$
$$=\sum_{n=1}^{\infty}[a_{mn}^{j+1}\hat{H}_{mn}^{j+1}(r^j)+b_{mn}^{j+1}\hat{J}_{mn}^{j+1}(r^j)]C_{a_{\nu n}}^{j+1}, \tag{B4}$$

where

$$C_{a_{\nu n}}^{j+1}=\int_0^{\infty}\frac{1}{\rho^j(z)}\Psi_\nu^j(z)\Psi_n^{j+1}(z)dz. \tag{B5}$$

By substituting Eq. (18) into the boundary condition

$$\frac{1}{\rho^j(z)}\left.\frac{\partial p^j}{\partial r}\right|_{r^j}=\frac{1}{\rho^{j+1}(z)}\left.\frac{\partial p^{j+1}}{\partial r}\right|_{r^j}, \tag{B6}$$

we obtain

$$\frac{1}{\rho^j(z)}\sum_{n=1}^{\infty}k_{rn}^j[a_{mn}^j\hat{H}_{mn}^{j\prime}(r^j)+b_{mn}^j\hat{J}_{mn}^{j\prime}(r^j)]\Psi_n^j(z)$$
$$=\frac{1}{\rho^{j+1}(z)}\sum_{n=1}^{\infty}k_{rn}^{j+1}\times[a_{mn}^{j+1}\hat{H}_{mn}^{j+1\prime}(r^j)$$
$$+b_{mn}^{j+1}\hat{J}_{mn}^{j+1\prime}(r^j)]\Psi_n^{j+1}(z), \tag{B7}$$

where $\hat{H}_{mn}'(r)$ and $\hat{J}_{mn}'(r)$ are derivatives of $\hat{H}_{mn}(r)$ and $\hat{J}_{mn}(r)$ with respect to $k_{rn}r$, respectively. By applying the operator

$$\int_0^{\infty}\Psi_\nu^j(z)(\cdot)dz \tag{B8}$$

to Eq. (B7), we have

$$a_{m\nu}^j\hat{H}_{m\nu}^{j\prime}(r^j)+b_{m\nu}^j\hat{J}_{m\nu}^{j\prime}(r^j)$$
$$=\sum_{n=1}^{\infty}[a_{mn}^{j+1}\hat{H}_{mn}^{j+1\prime}(r^j)+b_{mn}^{j+1}\hat{J}_{mn}^{j+1\prime}(r^j)]C_{b_{\nu n}}^{j+1}, \tag{B9}$$

where

$$C_{b_{\nu n}}^{j+1}=\frac{k_{rn}^{j+1}}{k_{r\nu}^j}\int_0^{\infty}\frac{1}{\rho^{j+1}(z)}\Psi_\nu^j(z)\Psi_n^{j+1}(z)dz. \tag{B10}$$

Equations (B4) and (B9) can be written in the matrix notation,

$$\hat{\boldsymbol{H}}_m^j\boldsymbol{a}_m^j+\hat{\boldsymbol{J}}_m^j\boldsymbol{b}_m^j=\boldsymbol{C}_{\boldsymbol{a}}^{j+1}(\hat{\boldsymbol{H}}_m^{j+1}\boldsymbol{a}_m^{j+1}+\hat{\boldsymbol{J}}_m^{j+1}\boldsymbol{b}_m^{j+1}), \tag{B11a}$$

$$\hat{\boldsymbol{H}}_m^{j\prime}\boldsymbol{a}_m^j + \hat{\boldsymbol{J}}_m^{j\prime}\boldsymbol{b}_m^j = \boldsymbol{C}_b^{j+1}(\hat{\boldsymbol{H}}_m^{j+1\prime}\boldsymbol{a}_m^{j+1} + \hat{\boldsymbol{J}}_m^{j+1\prime}\boldsymbol{b}_m^{j+1}), \tag{B11b}$$

where $\hat{\boldsymbol{H}}_m$, $\hat{\boldsymbol{J}}_m$, $\hat{\boldsymbol{H}}'_m$, and $\hat{\boldsymbol{J}}'_m$ denote the diagonal matrices with entries $\hat{H}_{mn}(r^j)$, $\hat{J}_{mn}(r^j)$, $\hat{H}'_{mn}(r^j)$, and $\hat{J}'_{mn}(r^j)$, respectively. In addition, $\boldsymbol{C}_a$, $\boldsymbol{C}_b$ are the matrices with entries $C_{a\nu n}$, $C_{b\nu n}$, respectively, and $\boldsymbol{a}_m$, $\boldsymbol{b}_m$ the are column vectors with entries a_{mn}, b_{mn}, respectively.

By solving the linear system in Eq. (B11), we obtain an explicit expression for $\boldsymbol{a}_m^j$ and $\boldsymbol{b}_m^j$, as shown in Eqs. (29)–(32).

2. Mode coupling in outward marching

Similar to the derivation of mode coupling in the inward marching, in the outward marching, we impose the boundary conditions in Eqs. (B1) and (B6). However, in the outward marching, we intend to express the expansion coefficients in sector $j+1$, i.e., $\boldsymbol{a}_m^{j+1}$ and $\boldsymbol{b}_m^{j+1}$, in terms of the expansion coefficients in sector j, i.e., $\boldsymbol{a}_m^j$ and $\boldsymbol{b}_m^j$. So, instead of applying the operators in Eqs. (B3) and (B8), we apply the operator

$$\int_0^\infty \frac{1}{\rho^{j+1}(z)}\Psi_\nu^{j+1}(z)(\cdot)dz \tag{B12}$$

to Eq. (B2), and apply the operator

$$\int_0^\infty \Psi_\nu^{j+1}(z)(\cdot)dz \tag{B13}$$

to Eq. (B7), and obtain

$$a_{m\nu}^{j+1}\hat{H}_{m\nu}^{j+1}(r^j) + b_{m\nu}^{j+1}\hat{J}_{m\nu}^{j+1}(r^j) = \sum_{n=1}^{\infty}[a_{mn}^j\hat{H}_{mn}^j(r^j) + b_{mn}^j\hat{J}_{mn}^j(r^j)]C_{c\nu n}^{\ j}, \tag{B14}$$

where

$$C_{c_{\nu n}}^j = \int_0^\infty \frac{1}{\rho^{j+1}(z)}\Psi_\nu^{j+1}(z)\Psi_n^j(z)dz, \tag{B15}$$

and

$$a_{m\nu}^{j+1}\hat{H}_{m\nu}^{j+1\prime}(r^j) + b_{m\nu}^{j+1}\hat{J}_{m\nu}^{j+1\prime}(r^j) = \sum_{n=1}^{\infty}[a_{mn}^j\hat{H}_{mn}^{j\prime}(r^j) + b_{mn}^j\hat{J}_{mn}^{j\prime}(r^j)]C_{d_{\nu n}}^j, \tag{B16}$$

where

$$C_{d_{\nu n}}^j = \frac{k_{rn}^j}{k_{r\nu}^{j+1}}\int_0^\infty \frac{1}{\rho^j(z)}\Psi_\nu^{j+1}(z)\Psi_n^j(z)dz. \tag{B17}$$

Equations (B14) and (B16) can be written in the matrix notation,

$$\hat{\boldsymbol{H}}_m^{j+1}\boldsymbol{a}_m^{j+1} + \hat{\boldsymbol{J}}_m^{j+1}\boldsymbol{b}_m^{j+1} = \boldsymbol{C}_c^j(\hat{\boldsymbol{H}}_m^j\boldsymbol{a}_m^j + \hat{\boldsymbol{J}}_m^j\boldsymbol{b}_m^j), \tag{B18a}$$

$$\hat{\boldsymbol{H}}_m^{j+1\prime}\boldsymbol{a}_m^{j+1} + \hat{\boldsymbol{J}}_m^{j+1\prime}\boldsymbol{b}_m^{j+1} = \boldsymbol{C}_d^j(\hat{\boldsymbol{H}}_m^{j\prime}\boldsymbol{a}_m^j + \hat{\boldsymbol{J}}_m^{j\prime}\boldsymbol{b}_m^j). \tag{B18b}$$

By solving the linear system in Eq. (B18), together with the one-way approximation, $\boldsymbol{b}_m^j = 0$, we obtain an explicit expression for $\boldsymbol{a}_m^{j+1}$ in terms of $\boldsymbol{a}_m^j$, as shown in Eqs. (36)–(39).

[1] M. D. Collins, and B. E. McDonald, "Three-dimensional effects in global acoustics," J. Acoust. Soc. Am. **97**, 1567–1575 (1995).
[2] C. H. Harrison, "Three-dimensional ray paths in basins, troughs, and near seamounts by use of ray invariants," J. Acoust. Soc. Am. **62**, 1382–1388 (1977).
[3] J. S. Perkins and R. N. Baer, "An approximation to the three-dimensional parabolic-equation method for acoustic propagation," J. Acoust. Soc. Am. **72**, 515–522 (1982).
[4] G. Botseas, D. Lee, and D. King, "FOR3D: A computer model for solving the LSS three-dimensional, wide angle wave equation," Naval Underwater Systems Center, New London, CT, Tech. Report No. 7943, 1987.
[5] G. A. Athanassoulis and A. M. Propathopoulos, "Three-dimensional acoustic scattering of a source-generated field from a cylindrical island," J. Acoust. Soc. Am. **100**, 206–218 (1996).
[6] M. I. Taroudakis, "A coupled-mode formulation for the solution in the presence of a conical sea-mount," J. Comput. Acoust. **4**, 101–121 (1996).
[7] E. G. Williams, *Fourier Acoustics* (Academic, London, 1999).
[8] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[9] J. Eskenazi, "A computer model for sound propagation around conical seamounts," MS thesis, Massachusetts Institute of Technology, Cambridge, MA (2001).
[10] D. C. Ricks, "Elastodynamic modeling of fluid-loaded cylindrical shells with multiple layers and internal attachments," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA (1994).
[11] C. Ferla, M. porter, and F. Jensen, "C-SNAP: Coupled SACLANTCEN normal mode propagation model," Report No. SM 274, SACLANT Undersea Research Centre, La Spezia, Italy, 1993.
[12] W. Luo, "Three-dimensional propagation and scattering around a conical seamount," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA (2007).
[13] G. A. Athanassoulis and K. A. Belibassakis, "All-frequency normal-mode solution of the three-dimensional acoustic scattering from a vertical cylinder in a plane-horizontal waveguide," J. Acoust. Soc. Am. **101**, 3371–3384 (1997).
[14] D. C. Ricks and H. Schmidt, "A numerically stable global matrix method for cylinderlly layered shells excited by ring forces," J. Acoust. Soc. Am. **95**, 3339–3349 (1994).
[15] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Applied Mathematic Series No. 55 (National Bureau of Standards, Washington, DC, 1972).
[16] M. B. Porter, F. B. Jensen, and C. M. Feria, "The problem of energy conservation in one-way models," J. Acoust. Soc. Am. **89**, 1058–1067 (1991).
[17] B. E. McDonald, M. D. Collins, W. A. Kuperman, and K. D. Heaney, "Comparison of data and model predictions for Heard Island acoustic transmissions," J. Acoust. Soc. Am. **96**, 2357–2370 (1994).
[18] D. Lee and G. Botseas, "Examination of three-dimensional effects using a propagation model with azimuth-coupling capability (FOR3D)," J. Acoust. Soc. Am. **91**, 3192–3202 (1992).
[19] D. C. Stickler, "Normal-mode program with both the discrete and branch line contributions," J. Acoust. Soc. Am. **57**, 856–861 (1970).
[20] R. B. Evans, "A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom," J. Acoust. Soc. Am. **74**, 188–195 (1983).

Reflection of underwater sound from surface waves

Chris T. Tindle
Physics Department, University of Auckland, Bag 92019, Auckland 1142, New Zealand

Grant B. Deane
Marine Physical Laboratory, Scripps Institution of Oceanography, UCSD, La Jolla, California 92093-0238

James C. Preisig
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 26 February 2008; revised 28 October 2008; accepted 30 October 2008)

A tank experiment has been conducted to measure reflection of underwater sound from surface waves. Reflection from a wave crest leads to focusing and caustics and results in rapid variation in the received waveform as the surface wave moves. Theoretical results from wavefront modeling show that interference of three surface reflected eigenrays for each wave crest produces complicated interference waveforms. There is good agreement between theory and experiment even on the shadow side of caustics where there are two surface reflected arrivals but only one eigenray.
 [DOI: 10.1121/1.3035828]

PACS number(s): 43.30.Cq, 43.30.Dr, 43.30.Gv [JAC] Pages: 66–72

I. INTRODUCTION

The use of sound for underwater communications is made complicated by reflections from the ocean surface.[1] The crests of ocean surface waves act as curved reflectors for underwater sound, which can give focusing and multiple arrivals due to the reflection from different parts of a single crest. The purpose of this paper is to investigate this phenomenon using a time-domain wave front propagation model[2,3] and some scale-model tank experiments.

It is well known that a ray trace for underwater sound propagation in deep water with a sound speed minimum as a function of depth leads to the formation of foci and caustics.[4] The various types of caustics and the consequences for pulse propagation as a result of vertical structure in the sound speed profile are discussed in detail by Brown.[5] There is an extensive literature treating wave scattering from rough surfaces in the electromagnetic and acoustical literature (see Refs. 6 and 7 for reviews). The wavefront model used here was developed to treat reflection of broadband pulses from a deterministic, time-varying surface, and thus differs from stochastic treatments of scattering from random, rough surfaces. The work by Frederickson and Marston[8,9] and Williams *et al.*[10] is particularly relevant to the present study because of their analytical treatment and experimental studies of caustic formation for the case of tone bursts and transient signals reflected by curved surfaces. The experiments reported here are broadband and were specifically designed to test the fidelity of the wavefront model.[2,3]

The present study represents one end of a continuum of surface conditions with a smooth, correlated surface at one extreme and a rough, random surface at the other. The length scale dividing "smooth" from "rough" may be usefully assumed to be proportional to the length of a Fresnel zone and is thus related to the range of wavelengths associated with the acoustic pulse reflected from the surface in addition to the source-receiver geometry and prevailing surface conditions. In practice, sea surface conditions can present well-correlated swell from distant storms, a random, wind-driven wave spectrum, or a mixture of the two. Although it would be valuable to study the surface conditions under which the deterministic modeling approach adopted here fails, this study is restricted to monochromatic, smooth surface waves. When scaled to ocean waves of 10 s period, the experiment reported here is equivalent to transmission at 890 Hz over a distance of 275 m beneath smooth swell with peak to trough heights from 1.6 m up to 10.4 m

A tank experiment was conducted to undertake a detailed reconciliation between the wavefront model described in Refs. 2 and 3 and scale-model propagation measurements. The surface waves in the flume were controllable, uniform, and reproducible. The experiment used very short, high frequency acoustic pulses or "pings" and a short range so that surface reflections arrived before any reflections from the walls or bottom of the tank. It was necessary to allow reverberation to decay away between pings but a repetition rate of 180 pings/s was achieved. The results were recorded and show how the pulses reflected from the surface change in amplitude, phase, and arrival time as the wave progresses. The experimental results have been modeled to very good accuracy with the wavefront model, and this gives confidence that the reflection process beneath smooth surface waves is well understood.

II. THE EXPERIMENT

The experimental arrangement is shown schematically in Fig. 1. The experiment was conducted in a wave tank at Scripps Institution of Oceanography. The tank is 0.6 m wide and 30 m long. The water depth is a nominal 0.6 m. Waves are generated with a computer-controlled paddle at one end of the tank. For the present results, the surface wave frequency was 1.5 Hz with a wavelength of 0.693 m. The

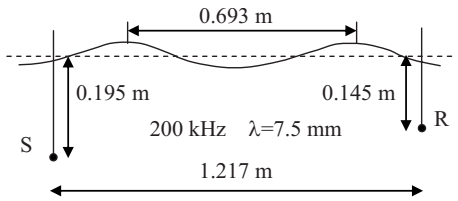


FIG. 1. Experimental arrangement. One cycle pulses at a nominal frequency of 200 kHz were transmitted from source S to receiver R under surface waves moving from left to right.

waves were uniform and repeatable from day to day. The wave height was constant for each run and varied up to a peak to trough height of 46 mm.

The acoustic source S and receiver R were ITC1089D SN transducers. Each transducer is a cylinder about 70 mm long and 10 mm diameter with the active element centered 5 mm from one end. The hydrophones were deployed horizontally, perpendicular to the direction of the waves and were omnidirectional for propagation along the tank. The depths of the nominal centers of the source and receiver transducers were determined by analyzing direct and surface reflected arrivals taken with a flat surface and were found to be 195 and 145 mm, respectively. The horizontal range was 1.217 m. The tank was filled with fresh water with a measured sound speed of 1469 m/s.

The instantaneous wave height near the source was measured to millimeter accuracy with a wave gauge composed of vertical parallel wires in a Wheatstone bridge arrangement. The surface wave speed, which was calculated using the known wave frequency, the measured mean water depth, and the full surface wave dispersion equation, was then used to find the instantaneous surface height as a function of horizontal displacement from the wave gauge. The data to be reported were identified as runs 101–106 and had peak to trough wave heights of 7, 16, 24, 31, 39, and 46 mm, respectively. Movies of experimental waveforms and modeling results for these runs can be viewed on EPAPS.[11]

Pulse transmission and data digitization were synchronized to the surface wave generator. A short pulse was transmitted at 180 pulses/s which gave 120 pulses per surface wave period. The waveform used to excite the source transducer was designed to compensate for the frequency-dependent amplitude and phase response of the ITC1089D ceramic element. The resulting pressure pulse in the water was nominally one cycle at 200 kHz (see the first arrival in Fig. 6). Although it would be possible to include a detailed electromechanical model of the transducers and their associated electronics in the propagation model, we have avoided this complication by using the observed direct arrival pulse at the receiver, averaged over 120 transmissions and scaled to account for geometric spreading, as the source pulse. This averaged pulse was then used directly as the source pulse in the wavefront calculations to model the acoustic field incident on the surface.

Data from the receiver were digitized at 1 538 461.54 samples/s for 6 153 846 samples or about 4 s, providing data for six complete surface waves. The received waveform for each pulse transmission was determined by averaging synchronized pulses over four complete waves, which served to reduce noise from electronic and acoustical sources and reduce any slight variability in reflection between successive wave crests. Initially wave profiles were measured simultaneously with the pulse transmissions. However, the waves were found to be highly repeatable from day to day so that later runs could use wave height data taken from earlier runs.

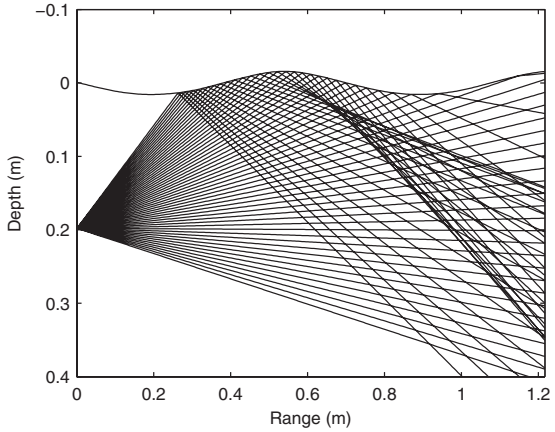


FIG. 2. Ray trace for run 104, ping 36.

III. WAVEFRONT MODELING

When the experimental results are presented, the waveforms will be compared with theoretical waveforms calculated using the wavefront modeling method described in Refs. 2 and 3. The calculations are based on the assumption that the surface can be considered stationary during a single pulse transmission. There are Doppler shifts associated with the movement of the surface, which the current model does not account for but which can be seen in data taken from transmission experiments in the surf zone. Doppler shifts could be included but are not necessary for the present analysis.

Wavefront modeling begins with a ray trace, a representative example of which is shown in Fig. 2. The ray trace is for ping 36 in run 104. The waves had a peak to trough wave height of 31 mm and the waves are traveling from left to right.

In Fig. 2, the wave crest is just before the midrange position, and the rays are focused by the crest to converge about 50 mm below the surface at a range of about 0.7 m. After passing through the focus the rays diverge to form a fan with well-defined boundaries, which are caustics for the ray field. The eigenray paths are shown in Fig. 3 for the same wave position shown in Fig. 2. There are three distinct surface reflected rays reaching the receiver.

As the surface wave moves, the number of eigenrays and the angles of the eigenrays change. When the wave crest is near the middle of the range there are typically three eigenrays, as shown in Fig. 3. When there is a trough near the middle of the range there is only one eigenray. As the wave progresses the position of the focus and the caustics in Fig. 2 changes, and the caustics sweep upward through the receiver position. When the receiver is between the caustics

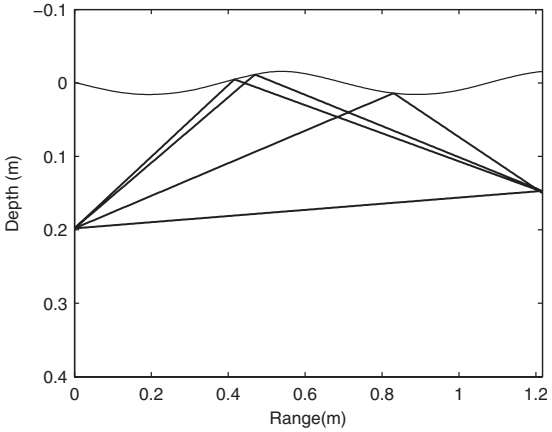


FIG. 3. Eigenrays for the ray trace of Fig. 2.

there are three eigenrays but there is only one eigenray when it is outside. The eigenrays appear or disappear in pairs as a caustic passes the receiver.

The ray trace is used in the wavefront modeling method to determine the depth z^* of a ray at the range of the receiver as a function of its angle θ at the source. A graph of z^* as a function of θ for the same situation shown in Fig. 2 is shown in Fig. 4.

The angles in Fig. 4 are negative because we have adopted the convention of distance measured positive downward from the mean surface. As the ray angle increases the depth of the ray at the receiver range decreases to a minimum at about −25°. This minimum corresponds to the upper caustic in Fig. 2. Further increase in ray angle leads to a maximum depth at about −18° which corresponds to the lower caustic in Fig. 2. Further increase in ray angle shows that the ray meets the surface for a ray angle of about −10°. Further increase in ray angle from −10° to 0° gives an almost linear increase in depth as the ray passes the receiver range without hitting the surface.

The line at a depth of 0.145 m in Fig. 4 corresponds to the depth of the receiver. Intersections of the line and the curve occur when the depth of the ray at the receiver range corresponds to the receiver depth. The intersections determine the angles for the eigenrays of Fig. 3. The direct ray has

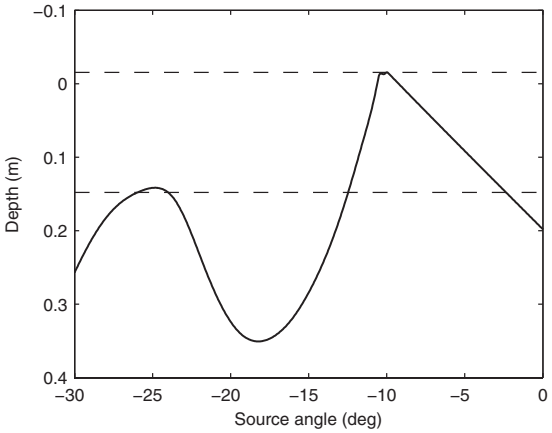


FIG. 4. Ray depth $z^*(\theta)$ at the receiver range as a function of launch angle θ. The upper dashed line is the ocean surface at the range of the receiver. The lower dashed line is the receiver depth.

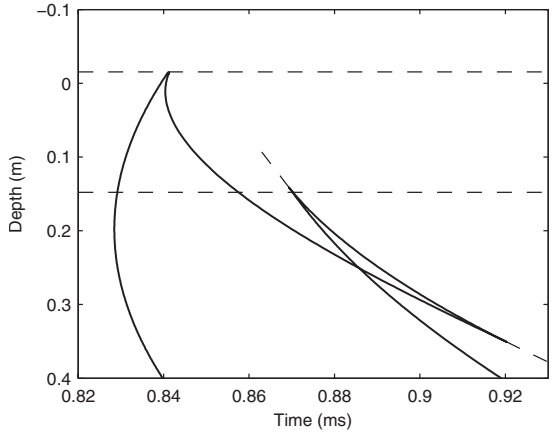


FIG. 5. Ray depth at the receiver range as a function of arrival time. The upper dashed line is the surface at the range of the receiver. The lower dashed line is the receiver depth.

a source angle of about −2° and the surface reflected rays have source angles of about −26°, −24°, and −12°.

The ray trace of the wavefront model is also used to find the ray travel time to the range of the receiver along each ray path. The travel time as a function of depth is shown in Fig. 5 for the ray trace of Fig. 2.

The curve in Fig. 5 is effectively the wavefront as it passes the range of the receiver. The part of the curve up to 0.84 ms is the spherical wavefront representing the direct ray path from the source to the range of the receiver. The later sections of the curve correspond to the surface reflection of the wavefront. The reflection from the concave surface beneath the wave crest has caused the wavefront to fold over on itself. The sharp reversals of the wavefront occur at the caustics. The dashed curves extrapolate the wavefront into the caustic shadow.

Intersections of the wavefront with the line at the receiver depth (0.145 m) show the arrival times of the pulses which follow the ray paths of Fig. 3. In wavefront modeling, the arrival times of the pulses are used to construct the expected waveform, which is shown as the thicker line in Fig. 6. The thinner line waveform in Fig. 6 is the experimental result for run 104, ping 36.

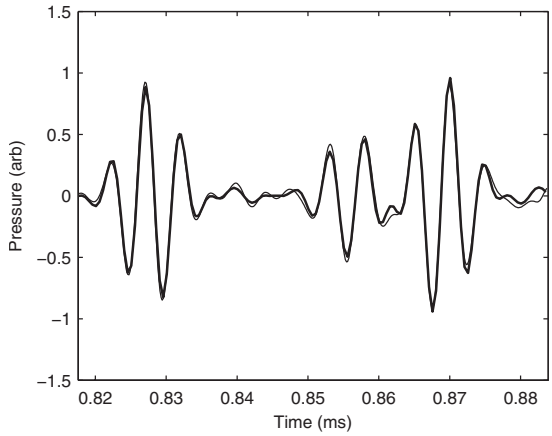


FIG. 6. Receiver waveforms for model (thick line) and data (thin line) for run 104, ping 36.

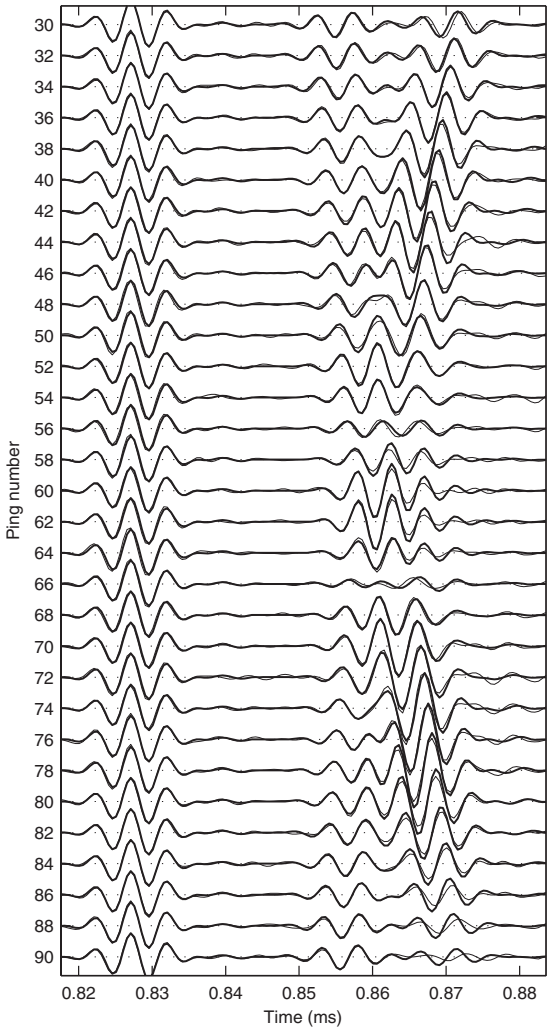


FIG. 7. Waveforms for run 104 and the ping numbers shown. The model is shown as thick lines and the data are shown as thin lines.

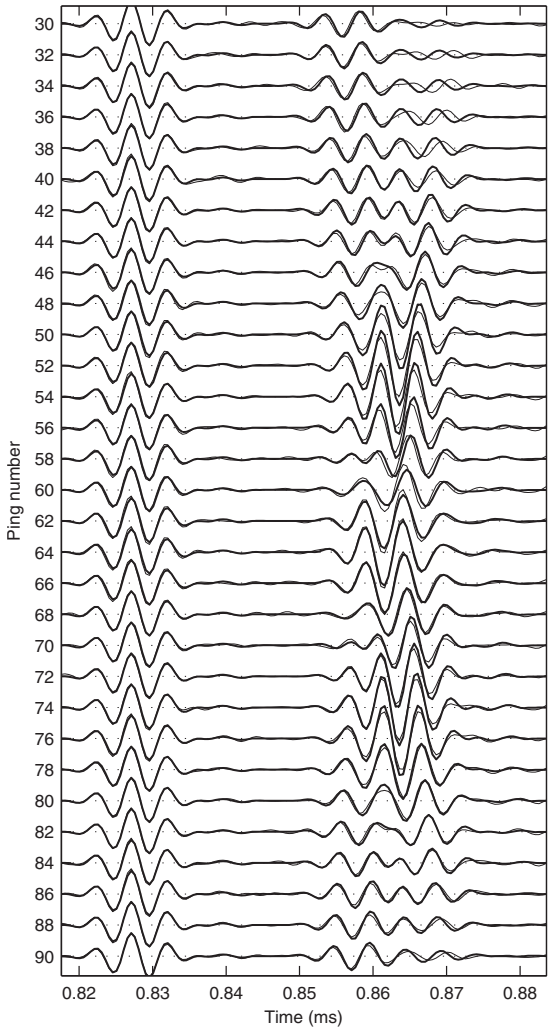


FIG. 8. Waveforms for run 103 and the ping numbers shown. The model is shown as thick lines and the data are shown as thin lines.

The first pulse in Fig. 6, centered at 0.828 ms, is the direct ray arrival. As described earlier, the model waveform for the direct arrival was found by averaging 120 experimental direct pulse waveforms. The slight differences evident between model and experiment for the direct pulse are due to the residual reverberation in the tank.

The second arrival in Fig. 6, centered at 0.855 ms, is from the ray which is reflected near the trough in Fig. 3. The third arrival, centered at 0.869 ms, is the combined pulse due to the two rays reflected near the wave crest in Fig. 3. These two rays have nearby ray angles as shown in Fig. 4 and very similar arrival times as shown in Fig. 5. The two rays are close to the upper caustic and are treated as a pair in the wavefront model.[3] The model waveform for the second and third arrivals is found by placing scaled and phase-shifted copies of the source pulse at the arrival times determined numerically from Fig. 5. The phase shift is achieved by taking the Fourier transform of the pulse, phase shifting the complex signal, and transforming back.

The theoretical expressions used to calculate the waveforms in Fig. 6 and the results to be shown in Figs. 7–10 were derived in Refs. 2 and 3. The amplitude and phase of isolated ray arrivals are given in Eq. (21) of Ref. 3. Pairs of rays which pass near a caustic and arrive with a travel time difference less than a quarter of a period (at the center frequency) give a single pulse described by an Airy function. The phase and amplitude of such pairs are given by Eq. (23) of Ref. 3. The pulse arrivals due to the caustic shadow to be shown in Figs. 7–10 are also described by an Airy function and have phase and amplitude determined using Eq. (23) of Ref. 3. Near a focus there are either three arrivals or one arrival and a caustic shadow. If the travel time differences are less than a quarter period the phase and amplitude of the combined pulse are found using the Pearcey function formula in Eq. (58) of Ref. 2 modified to allow for range dependence as described in Ref. 3. The formulas are summarized in Ref. 12.

IV. RESULTS

As the surface wave progresses all the above diagrams change smoothly. The positions of the focus and caustics move upward. The angles of the eigenrays change. The fold in the wavefront shown in Fig. 5 moves upward and the arrival times of the rays change. The amplitudes of the arrivals also change because they are related to the slope of $z^*(\theta)$. The amplitude of isolated ray arrivals is proportional to

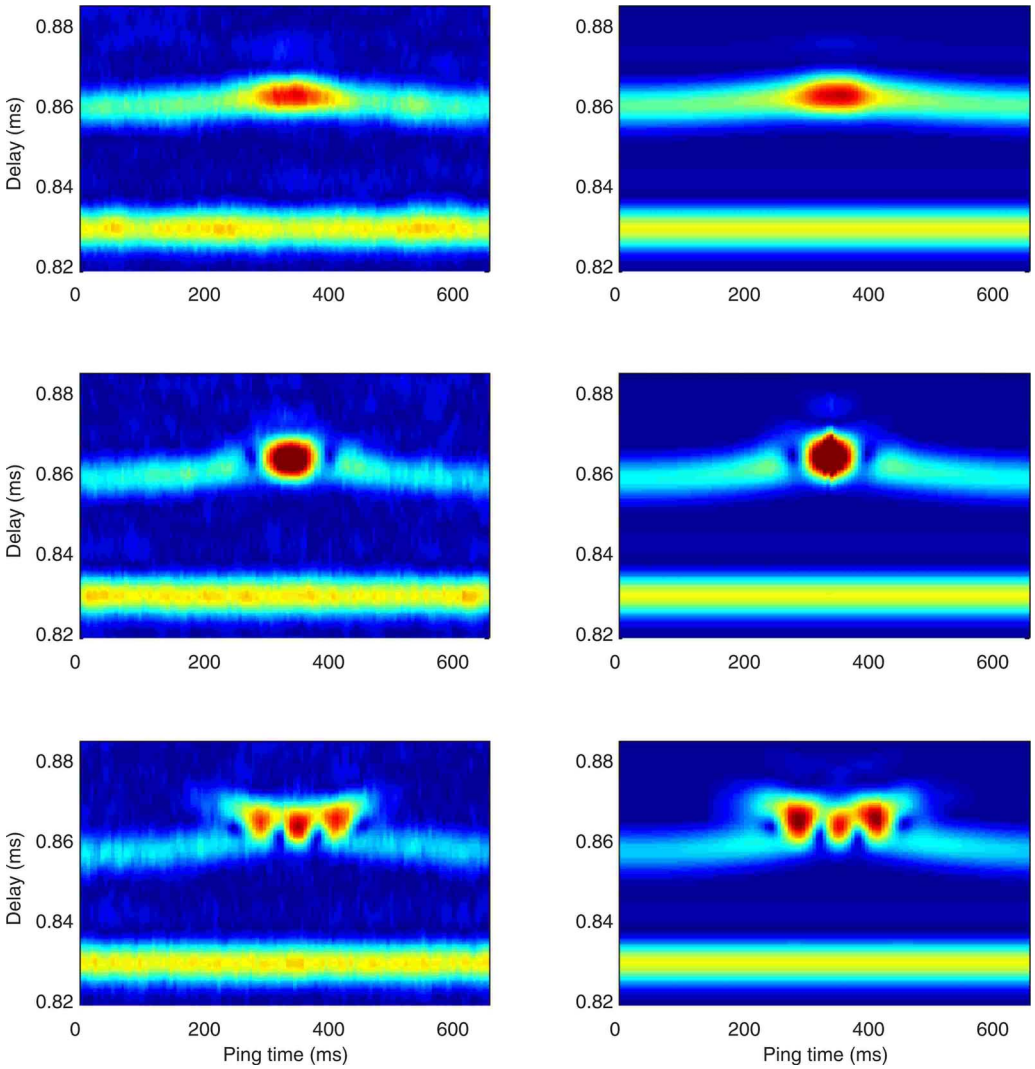


FIG. 9. (Color online) Signal amplitude contours. Individual pings are plotted vertically as a function of delay time. The horizontal axis is the transmission time of each ping. There are 120 pings in the 667 ms wave period. The left panels are the measured signals. The right panels show the wavefront model results. Top panels: run 101. Peak to trough wave height of 7 mm. Middle panels: run 102. Wave height of 16 mm. Bottom panel: run 103. Wave height of 24 mm.

$|dz^*/d\theta|^{-1/2}$ so steeper slope means smaller amplitude. In Fig. 4 the slope at −2° is smaller than the slope at −12° so, as expected, the direct arrival in Fig. 6 has greater amplitude than the isolated surface reflected arrival. Arrivals near a zero of $dz^*/d\theta$ occur in pairs and are treated as a pair using Airy functions to give a pulse of finite amplitude.[2,3]

The waveforms for a subset of pings for run 104 for a peak to trough wave height of 31 mm are shown in Fig. 7. The waveform of Fig. 6 is repeated four lines down in Fig. 7. For all pings in Fig. 7, the first pulse is the constant direct arrival, unaffected by the surface waves.

Waveforms for pings 30–34 show two surface reflected arrivals. The first surface arrival has a ray path similar to the isolated surface reflection in Fig. 3. The second surface arrival has no ray path and is due to the caustic shadow. If Figs. 2–5 were shown for pings 30–34, the upper caustic in Fig. 2 would be below the source depth. Correspondingly, the extremum of the ray depth curve in Fig. 4 would not intersect the receiver depth. The folded wavefront of Fig. 5 due to the surface reflection would intersect the receiver depth only once. However, the dashed extrapolation of the wavefront leads to a field in the shadow zone of the caustic and so pings 30–34 show a third arrival due to the caustic shadow.

Ping 36 has been discussed above. It has three surface reflections with the second and third arrivals very close together. Pings 36–82 all have three distinct surface reflections. The ray angles and arrival times of the individual eigenrays change smoothly as the surface wave moves. In pings 36–42 the second and third eigenrays arrive with a path difference less than a quarter wavelength which is equivalent to a travel time difference of less than a quarter of a period. As noted above the two surface reflections arriving close together are treated as a pair and the resulting field is given by an Airy function.[2,3]

Pings 44–74 all have three surface reflected eigenrays which interfere to produce maxima at pings 52 and 62 and minima at pings 56 and 66. The details of the interference is well reproduced by the wavefront model. For pings 76–82 the second and third surface reflections arrive close together and are treated as a pair as described above. Pings 84–90

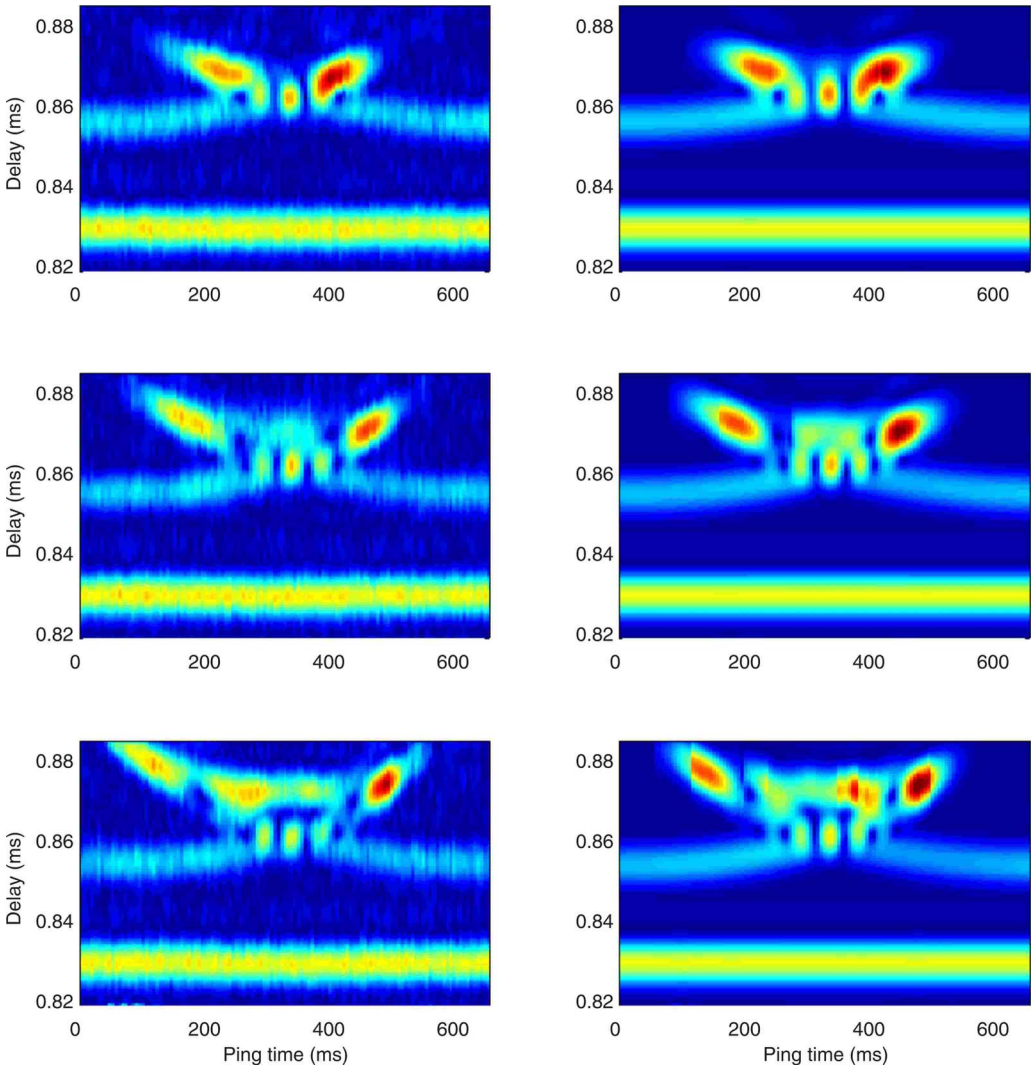


FIG. 10. (Color online) As for Fig. 9. Top panels: run 104. Peak to trough wave height of 32 mm. Middle panels: run 105. Wave height of 39 mm. Bottom panel: run 106. Wave height of 46 mm.

have only one eigenray but the second pulse observed in the waveforms is due to the shadow of the lower caustic in Figs. 4 and 5.

Figure 8 shows results for run 103. The wave height was 24 mm peak to trough. This wave height is smaller than for run 104 so the time spread of the surface reflections is less and no separated surface reflected pulses are observed. Nevertheless the same processes occur, as shown in Fig. 7. Pings 30–44 and 82–90 have one eigenray and a caustic shadow. Pings 46–54 and 72–80 have three eigenrays with the second and third arrivals treated as a pair. Pings 56–70 have three separate arrivals which interfere to give minima at pings 60 and 68 and a maximum at ping 64. The waveform for the surface reflected signal for ping 64 is almost exactly an inverted copy of the source signal and shows that the three surface reflected rays arrive almost simultaneously and reinforce each other.

The full set of data and model results is shown in Figs. 9 and 10. The waves in the tank experiment have a wavelength of 0.693 m and the peak to trough wave heights vary from 7 mm for run 101 up to 46 mm for run 106. To provide an oceanic context, when scaled by a factor of 225, the results correspond to transmission of 890 Hz pulses a distance of 275 m under ocean swells of period 10 s with peak to trough heights from 1.6 m (run 101) to 10.4 m (run 106).

Each panel in Figs. 9 and 10 shows contours of signal amplitude made by plotting the absolute value of the Hilbert transform of the waveform. Each ping is plotted vertically as a function of time, and the time scale is the same as in Figs. 7 and 8. The horizontal axis is the start time of each ping. There are 120 pings in each panel covering the 667 ms period of the surface waves. In all cases the central pings coincide approximately with the wave crest and the first and last pings occur at the wave trough. The waveforms in Fig. 8 are a subset of the data in Fig. 9, bottom panels. The waveforms in Fig. 7 are a subset of the data in Fig. 10, top panels.

The top panels in Fig. 9 show data and model for run 101 with a peak to trough wave height of 7 mm. The corresponding ray diagram would show weak focusing to a focus beyond the range of the receiver with no caustic present. The surface reflected signal is slightly delayed compared to the first and last pings because of the extra distance to the surface. The signal is increased in amplitude due to the converging effect of reflection from the curved surface.

The middle panels in Fig. 9 show data for run 102 with a peak to trough wave height of 16 mm. The corresponding ray diagram would show that the focus is formed just before the range of the receiver and the two caustics are very close together. This gives very strong enhancement of the received signal and a rapid and intense focus as the wave passes. It is this sort of strong signal enhancement that is a major complication for underwater acoustics communications systems.

The bottom panels in Fig. 9 show data for run 103 with a peak to trough wave height of 24 mm. A subset of the waveforms for this figure has been shown in Fig. 8. There is strong reinforcement in the center of the pattern when the surface wave is midway between source and receiver and the three surface reflected rays similar to those of Fig. 3 arrive almost simultaneously as noted above.

The strong signals at ping times of 280 and 420 ms in the bottom panel of Fig. 9 come from near the caustic. Two rays arrive very close together and give strong maxima near the caustics. At times before 280 ms and after 420 ms the surface reflected signal is due to the sound in the shadow zone of the caustics and the sound energy decays steadily with distance from the caustic.

The top panel in Fig. 10 is for run 104 and a wave height of 31 mm. Some of the waveforms have been shown in Fig. 7. At the center of the pattern two of the surface reflected rays arrive simultaneously and they reinforce to give the local maximum in Fig. 10 and pings 60–62 of Fig. 7. The third arrival from the crest of the wave is sufficiently delayed compared to that for run 103 that it does not reinforce the other two rays and there is smaller overall amplitude than for run 103.

The middle and lower panels in Fig. 10 are for runs 105 and 106 with wave heights of 39 and 46 mm, respectively. There are single arrivals for the earlier and later pings in each panel. In the center of each pattern two of the surface reflected rays arrive simultaneously and reinforce. The third surface arrival comes later and is somewhat separated from the other two. For the pings near but not at the center there are three distinct surface arrivals corresponding to the three arrivals of Fig. 5. As the wave progresses the surface reflected pulses partly overlap and give rise to the complicated interference patterns seen in both data and model results.

V. DISCUSSION AND CONCLUSIONS

The experimental results show that the presence of surface waves significantly alters propagation conditions for transmission and reception of underwater acoustics signals. It is clear that communications systems need to be designed to allow for rapid changes in signal level and rapid changes in interference from multiple paths.

The good agreement between theory and experiment shown in Figs. 6–10 is confirmation that the wavefront modeling method gives an accurate description of the propagation. The model is based on a ray interpretation of the propagation and correctly predicts the amplitude, phase, and travel time of pulses along each ray path.

Agreement between theory and experiment is not perfect. There is some noise in the data, as can be seen from the left panels of Figs. 9 and 10. The main source of difference between theory and experiment is probably noise in the measurement of the surface wave height. The focusing effect of a concave reflector is very sensitive to the details of the surface. This is because the ray trace uses both height and slope of the water surface and relies on smooth variation as the launch angle changes. It was necessary to smooth the surface waves by sampling every 3 cm in range and fitting a smooth curve to the surface wave data points.

Even though the experiment was conducted at a nominal frequency of 200 kHz and a wavelength of 7.5 mm, scaling the waves to a nominal swell period of 10 s shows that it is actually a low frequency situation. The good agreement of the wavefront modeling ray based approach is surprising given that the wave heights of up to 46 mm lead to path differences of only one or two wavelengths.

The experiment has shown the variations in signal travel time, amplitude, and phase associated with reflections from surface waves. The close agreement between theory and experiment shows that the interpretation in terms of pulse propagation along ray paths is physically correct.

ACKNOWLEDGMENT

The support of the Office of Naval Research, Grant No. N00014-04-1-0728, is gratefully acknowledged.

[1]J. C. Preisig and G. B. Deane, "Surface wave focusing and acoustic communications in the surf zone," J. Acoust. Soc. Am. **146**, 2067–2080 (2004).
[2]C. T. Tindle, "Wavefronts and waveforms in deep-water sound propagation," J. Acoust. Soc. Am. **142**, 464–475 (2002).
[3]C. T. Tindle and G. B. Deane, "Shallow water sound propagation with surface waves," J. Acoust. Soc. Am. **117**, 2783–2794 (2005).
[4]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer-Verlag, New York, 1993).
[5]M. G. Brown, "The transient wave fields in the vicinity of cuspoid caustics," J. Acoust. Soc. Am. **79**, 1367–1384 (1986).
[6]J. A. Ogilvy, "Wave scattering from rough surfaces," Rep. Prog. Phys., **50**, 1553–1608 (1987).
[7]J. A. Ogilvy and H. M. Merklinger, *Theory of Wave Scattering From Random Rough Surfaces* (Hilger, New York, 1991).
[8]C. K. Frederickson and P. L. Marston, "Transverse cusp diffraction catastrophes produced by the reflection of ultrasonic tone bursts from a curved surface in water: Observations," J. Acoust. Soc. Am. **92**, 2869–2876 (1992).
[9]C. K. Frederickson and P. L. Marston, "Travel time surface of a transverse cusp caustic produced by reflection of acoustical transients from a curved metal surface in water," J. Acoust. Soc. Am. **95**, 650–660 (1994).
[10]K. L. Williams, J. S. Stroud, and P. L. Marston, "High-frequency forward scattering from Gaussian spectrum, pressure release, corrugated surfaces. I. Catastrophe theory modeling," J. Acoust. Soc. Am. **96**, 1687–1702 (1994).
[11]See EPAPS Document No. E-JASMAN-125-041901 for movies of experimental waveforms and modeling results. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.
[12]C. T. Tindle, G. Deane, and J. Preisig, "Multipath reflection from surface waves," Acoustics 08 Paris Proceedings, Societe Francaise d'Acoustique.

Use of the distorted wave Born approximation to predict scattering by inhomogeneous objects: Application to squid

Benjamin A. Jones,[a] Andone C. Lavery, and Timothy K. Stanton
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543-1053

(Received 8 August 2007; revised 7 August 2008; accepted 10 October 2008)

A new method has been developed to predict acoustic scattering by weakly scattering objects with three-dimensional variability in sound speed and density. This variability can take the form of inhomogeneities within the body of the scatterer and/or geometries where the acoustic wave passes through part of the scattering body, into the surrounding medium, and back into the body. This method applies the distorted wave Born approximation (DWBA) using a numerical approach that rigorously accounts for the phase changes within a scattering volume. Ranges of validity with respect to material properties and numerical considerations are first explored through comparisons with modal-series-based predictions of scattering by fluid-filled spherical and cylindrical fluid shells. The method is then applied to squid and incorporates high resolution spiral computerized tomography (SCT) scans of the complex morphology of the organism. Target strength predictions based on the SCT scans are compared with published backscattering data from live, freely swimming and tethered squid. The new method shows significant improvement for both single-orientation and orientation-averaged scattering predictions over the DWBA-homogeneous-prolate-spheroid model. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3021298]

PACS number(s): 43.30.Ft, 43.30.Sf, 43.30.Zk [KGF] Pages: 73–88

I. INTRODUCTION

Extracting biologically important information such as organism type and abundance from acoustic scattering data relies heavily on the availability of experimentally validated scattering models [see reviews by Horne (2000) and Simmonds and MacLennan (2005)]. The application of a model often dictates the level of complexity used in representing the scatterer. For example, predictions of volume scattering, which may use an ensemble of discrete predictions averaged over a range of parameters such as size and orientation of the scatterer, do not need the level of complexity required to predict discrete echoes for unique identification of class, size, or species of the scattering organism.

Much research has focused on organisms, principally zooplankton, that are classified as weakly scattering with fluidlike material properties because of their high numerical abundance and importance in the ocean food web. Weakly scattering implies a density and sound speed similar to that of the surrounding water, whereas, fluidlike material properties refers to tissue that does not support shear waves. Significant advances have been made in developing acoustic scattering models for this group of organisms involving shapes of varying complexity. The representation of shape in early studies was based on models of simple geometric fluid volumes such as spheres, finite-length cylinders, and prolate spheroids (Anderson, 1950; Yeh, 1967; Johnson, 1977; Stanton, 1988). More recent models have incorporated two-dimensional (Stanton *et al.*, 1988; McGehee *et al.*, 1988; Amakasu and Furusawa, 2006; Stanton and Chu, 2000) and three-dimensional (Lavery *et al.*, 2002) details of zooplankton shape.

One general approach that can account for anatomical complexities of weak scatterers is the distorted wave Born approximation (DWBA) (Chu *et al.*, 1993; Stanton *et al.*, 1993). This volume integral formulation for weakly scattering bodies can account for three-dimensional shape as well as inhomogeneous material properties. Accurately applying this formulation, however, requires digitizing an organism's complex inner and outer anatomy at sufficiently high resolution (Stanton and Chu, 2000), an issue addressed in this paper. The DWBA formulation has been successfully applied to model the scattering of sound from various types of zooplankton, including euphausiids, copepods, and decapod shrimp using high resolution outer shapes, although in these cases the changes in phase that occur when the acoustic wave passes through part of the scattering body into the surrounding medium, and back into the body were not taken into consideration. These model predictions have been validated through laboratory and, to some extent, field measurements (Simmonds and MacLennan, 2005; Lawson *et al.*, 2006; Lavery *et al.*, 2007; Stanton and Chu, 2000).

Accounting for internal anatomical structure in the DWBA and other models remains a challenge. One modeling restriction is the limited information available concerning sound speed and density of different tissues within most marine organisms. Even when these variations are known, implementation into the model can be a challenge. For both of these reasons, application of the DWBA has been prima-

[a]Present address: USS Bonhomme Richard (LHD 6), FPO AP 96617-1656. Electronic mail: bnyjones@gmail.com

rily used to model marine organisms as homogeneous scatterers, as in most of the aforementioned studies. Application of the DWBA to inhomogeneous scatterers has been limited to a subset of those studies. In those exploratory cases the material properties were varied, somewhat heuristically, along the length of an organism's body, with the variations corresponding to segmentation of the exoskeleton (Stanton *et al.*, 1988; Stanton and Chu, 2000; Lavery *et al.*, 2002). Furthermore, those studies did not account for the influence of inhomogeneities on the phase of the incident acoustic wave.

Given the demonstrated usefulness of the DWBA when applied to homogeneous objects, the formulation is further explored in this paper to account for three-dimensional variations of material properties within the objects. Although the DWBA formulation does, in principle, account for three-dimensional inhomogeneous material properties, evaluating the integral in this case requires careful consideration of the phase of the incident signal within the object. In the original Born approximation (undistorted), the amplitude and phase of the incident wave are only dependent on the position of the wavefront with respect to some arbitrary origin and the material properties of the surrounding medium. This is due to the general assumption that the incident wave is unmodified by the weakly scattering body. In contrast, the DWBA is a modification to the Born approximation in which the wavenumber inside the scattering volume is determined by the material properties within the body [e.g., see Stanton *et al.* (1993)].

The undistorted Born approximation has, in fact, been used to successfully describe scattering by inhomogeneous continuous random media, such as atmospheric and oceanic turbulence (Goodman, 1990; Lavery *et al.*, 2003; Tatarski, 1961). However, when the Born approximation has been applied to finite bounded scatterers, such as spheres, cylinders, and zooplankton, it has been found that the frequency responses of the scattering, most notably the positions of nulls in these spectra, are not accurately predicted. It is this result, which originally motivated use of the DWBA (although not shown explicitly in the original papers), that more accurately predicts the position of the nulls in frequency spectra of the scatterers (Stanton *et al.*, 1993; Chu *et al.*, 1993). In the DWBA, the phase of the wavefront, at any point in the volume, is dependent both on the distance traveled by the incident wave and any sound speed variations encountered along the path traveled. Accounting for this dependence presents a significant challenge in applying the DWBA to inhomogeneous objects and is the subject of this paper.

The DWBA is applied to inhomogeneous bodies in this paper by accurately accounting for both the amplitude and phase change of the incident wave as it travels through a scattering volume. Sound speed and density can vary arbitrarily in three dimensions, and the only restriction is that the variations must be small compared with the average values of the material properties of both the scattering volume and the surrounding medium. The numerical implementation of this method uses a two-part algorithm that calculates phase and amplitude separately for every discretization before integrating over the entire volume. This method of solving the DWBA is referred to as the phase-tracking DWBA method throughout this article. This approach to implementing the DWBA is first tested against analytical solutions to the simple cases of fluid-filled spherical and cylindrical fluid shells and then applied to the more complex case of squid.

Although relatively little work has been published on acoustic scattering models of squid, there is considerable interest in this research in two different areas. The first is commercial fisheries. Acoustic stock assessments of squid can be used to complement more traditional techniques such as trawl surveys by rapidly surveying large volumes of water and providing real time population assessments (Starr and Thorne, 1988). Understanding sound scattering by squid is essential for a quantitative interpretation of these acoustic surveys. Secondly, there is increasing interest in the predator-prey relationship between echolocating marine mammals and squid. Beaked whales, for instance, hunt squid using broadband ultrasonic sonar (Johnson *et al.*, 2004; Madsen *et al.*, 2005) and may use characteristics of the frequency responses from prey to help discriminate between targets in the water column (Jones *et al.*, 2008). Scattering models that help define the dominant scattering mechanisms of squid may elucidate factors that are exploited by the whales in discriminating between prey and nonprey.

Measurements of density and sound speed of squid suggest that these invertebrates are well suited to being modeled as weak scatterers (Mukai *et al.*, 2000; Kang *et al.*, 2006; Iida *et al.*, 2006). Accordingly, previous models of squid use the exact liquid prolate spheroid model and the DWBA formulation, also using a prolate spheroidal geometry (Arnaya and Sano, 1990; Mukai *et al.*, 2000). However, these models assume homogeneous material properties within the scattering volume of simple shape. This assumption is clearly a simplification of the complex anatomy of squid. For example, the main section of the squid body includes the mantle, an outer shell of tissue that partially encloses a seawater-filled cavity called the mantle chamber.

In this study high resolution, spiral computerized tomography (SCT) scans of squid have been taken. The results of these scans enable the incorporation of the complex interior and exterior shape of the squid's body into the phase-tracking DWBA method by differentiating between seawater-filled cavities and the squid's body. Appendages (i.e., arms, tentacles, and fin) as well as internal structure (i.e., the mantle cavity) are treated as inhomogeneities within the scattering volume. Material property variation due to internal organs, however, is not included in this study. Scattering predictions based on SCT scans of a locally caught squid are compared with published measurements of both anesthetized, tethered squid and live, freely swimming squid.

This paper is organized as follows: the phase-tracking DWBA for inhomogeneous objects is first developed. In order to assess the accuracy of this method, it is then applied to simple geometric shapes of both homogeneous and inhomogeneous material compositions for which there are analytical solutions. Two scattering formulations are used for comparisons. Modal-series-based scattering models (spherical shell and finite-length cylindrical shell) are used to compare predictions of scattering by objects with various-sized inhomo-

geneities and a range of material properties. Since the modal-series-based solution for scattering from cylindrical shells applies only near-normal incidence, a DWBA-based finite-length cylinder model is used to compare predictions of scattering by homogeneous elongated bodies over all orientations. Finally, the phase-tracking DWBA method is applied to the more complex case of squid. Spiral CT scans provided high resolution three-dimensional measurements of a locally available species of squid, *Loligo pealeii*. These measurements were then scaled, and in one case altered, so that predictions of acoustic scattering could be compared with published scattering data involving squid of the species *Todarodes pacificus*. Conclusions are then made about the benefits of this new approach.

II. THEORY

A. Definitions of fundamental quantities

Acoustic scattering from an object in the far field can be described in terms of the amplitude of the incident sound wave, P_0, and the scattering amplitude, f,

$$P_{\text{scat}} = P_0 \frac{e^{ik_1 r}}{r} f, \tag{1}$$

where r is the distance from the object to the receiver. The acoustic wavenumber of the surrounding medium, k_1, is defined as $2\pi/\lambda_1$, where λ_1 is the acoustic wavelength in the surrounding medium.

Target strength (TS) is the logarithmic measure of the backscattered signal, expressed in decibels, relative to m^2,

$$\text{TS} = 10 \log \sigma_{\text{bs}}, \tag{2}$$

where $\sigma_{\text{bs}}=|f_{\text{bs}}|^2$ is the differential backscattering cross section and f_{bs}, or backscattering amplitude, is the scattering amplitude evaluated in the backscattering direction. Mean TS is determined by averaging σ_{bs} prior to logarithmic conversion and is given as $\langle \text{TS} \rangle = 10 \log \langle \sigma_{\text{bs}} \rangle$.

In order to compare scattering from objects that are similar in proportion but of different overall sizes, the reduced target strength (RTS) is often used. Throughout this article scattering predictions will refer to either RTS or TS, which only differ by a constant. For a sphere or spherical shell of outer radius a (Fig. 1), RTS is given by

$$\text{RTS} = 10 \log \frac{\sigma_{\text{bs}}}{\pi a^2}. \tag{3}$$

In the case of elongated objects such as cylinders or cylindrical shells of length L, the RTS is given by

$$\text{RTS} = 10 \log \frac{\sigma_{\text{bs}}}{L^2}. \tag{4}$$

For the case of scattering from squid, the length L is replaced by dorsal mantle length, L_{ml}. Additionally, due to the lack of published data on mean mantle widths, from which mean cylindrical radii for squid could be derived, the nondimensional term $k_1 L_{\text{ml}}$ is used instead of the more commonly used $k_1 a$ for scattering predictions.

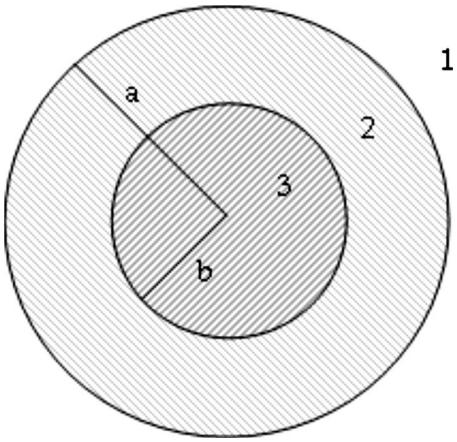


FIG. 1. Bisection of a spherical shell or cross section of a cylindrical shell. Indices 1–3 correspond to fluids of different material properties (i.e., sound speed, c_i, and density, ρ_i, $i=1,2,3$). The radii a and b are the outer and inner shell radii, respectively, such that $(a-b)/a$ corresponds to fractional shell thickness, τ.

B. The DWBA formulation

The Born approximation was derived for weakly scattering bodies in which the total pressure field inside a scattering volume is approximated by the unperturbed incident field. The approximation is valid when the amplitude of the scattered wave is much smaller than the incident wave (Morse and Ingard, 1968). Due to its volume integral form, this approximation may be used for arbitrarily shaped scatterers with any orientation with respect to the incident sound wave. Furthermore, it is valid for all acoustic frequencies. The DWBA formulation involves replacing the incident wavenumber vector inside the integral with the wavenumber vector resulting from the local sound speed and the frequency of the incident sound wave. It is this modification that accounts for the phase change of the sound wave due to material property inhomogeneities within the scattering volume. Within this formulation, Eq. (8) of Stanton *et al.* (1993), the backscattering amplitude is given by

$$f_{\text{bs}} = \frac{k_1^2}{4\pi} \int\int\int_v (\gamma_\kappa - \gamma_\rho) e^{2i\boldsymbol{k}_v \cdot \boldsymbol{r}_v} dv. \tag{5}$$

The terms γ_κ and γ_ρ are defined in terms of compressibility, κ, and density, ρ, or can also be expressed in terms of the commonly used ratios of density, $g_v \equiv \rho_v/\rho_1$, and sound speed, $h_v \equiv c_v/c_1$, given here by

$$\gamma_\kappa \equiv \frac{\kappa_v - \kappa_1}{\kappa_1} = \frac{1 - g_v h_v^2}{g_v h_v^2}, \tag{6}$$

$$\gamma_\rho \equiv \frac{\rho_v - \rho_1}{\rho_v} = \frac{g_v - 1}{g_v}. \tag{7}$$

In all cases the subscript "1" indicates parameters of the surrounding medium and the subscript "v" indicates parameters within the scattering volume. Accordingly, the wavenumber in the external medium is the constant k_1, while the wavenumber vector and the position vector within the volume are $\boldsymbol{k}_v$ and $\boldsymbol{r}_v$, respectively. Throughout this paper, when a model formulation is applied to a scattering problem, the

subscript "v" takes on the integers "2,3,…" corresponding to given material property parameters of that region within the volume. For example, in the region designated by "3" in Fig. 1, the density ratio, g_v, becomes $g_3=\rho_3/\rho_1$.

In Stanton *et al.* (1998), this formulation was simplified to a line integral for objects that are axisymmetric at any point along their lengthwise (deformed) axis. For straight, finite-length cylinders, the line integral can be further reduced to the following closed form equation:

$$f_{\mathrm{bs}} = \frac{k_1 k_v a^2 L}{2}(\gamma_\kappa - \gamma_\rho) e^{-\mathrm{i}Lk_v \sin\theta} \times \frac{J_1(2k_v a\cos\theta)}{2k_v a\cos\theta}\frac{\sin(k_v L\sin\theta)}{k_v L\sin\theta}, \tag{8}$$

where J_1 is the Bessel function of the first kind and θ is the angle of incidence measured from the incident wave vector, $\boldsymbol{k}_1$, to a plane orthogonal to the longitudinal axis of the cylinder (e.g., $\theta=0$ is normal incidence). It should be noted that the exponential phase term is dependent on the zero-phase origin used in the calculation, although this does not affect the calculations of TS or RTS, which use the absolute value of f_{bs}. This model is termed the *DWBA finite-length cylinder model (closed form)* throughout this paper and is valid over all angles of orientation. This result is used to test the phase-tracking DWBA approach over a range of orientations of the scattering volume with respect to the incident acoustic wave.

C. Numerical implementation of the phase-tracking DWBA method

The DWBA volume integral formulation can, in principle, account for cases where the scattering body has inhomogeneous material properties or where the wave passes through part of the body, into the surrounding medium, and back into the body. However, evaluating the integral for these conditions is a challenge as the phase must be calculated at each point in the volume by piecewise integration of the exponential phase term. In other words, the phase at any given point depends not only on the local sound speed but also on the variation in sound speed over the path the sound wave has already traveled through the scattering body. Given the complexity of the problem, a numerical approach has been developed to implement the DWBA for inhomogeneous objects.

The model presented in this article, which numerically computes the DWBA volume integral, is based on discretizing the scattering volume. The volume is composed of l cross-sectional matrices, M^l. Each element of each matrix, M^l_{ij}, represents one volume element, or *voxel*, and is assigned an integer value depending on the material properties of that voxel. Although only two fluid types are used to represent scattering volumes in this study, this model can, in general, be applied to volumes containing any arbitrary variation of material properties. Both internal inhomogeneities and geometries where the sound wave has multiple points of entry into and exit from the scattering body are treated as inhomogeneous media (see, for example, Fig. 2).

In order to accurately account for the phase of the sound wave as it propagates through an inhomogeneous medium, it is necessary to incorporate a phase-tracking algorithm. This algorithm, analogous to a ray-tracing routine, calculates the phase change associated with the sound speed of each differential element along a path traversed by the sound wave. For backscattering, the one-way phase change is simply doubled to calculate the full phase term.

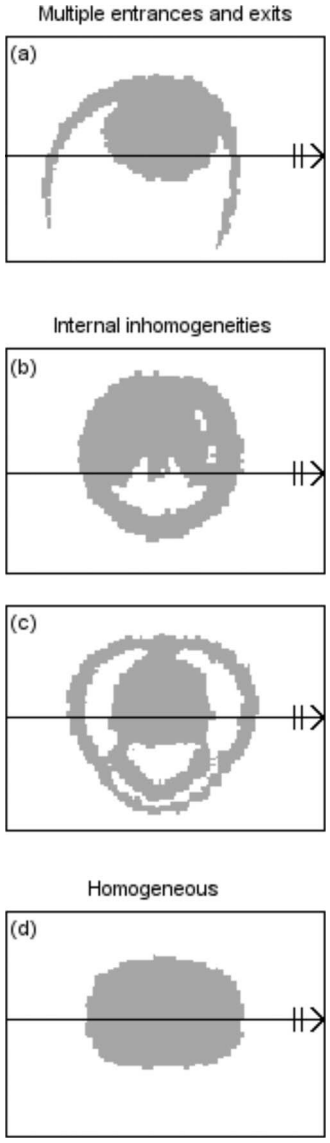


FIG. 2. Ray diagrams, using actual SCT scans of *L. pealeii*, that illustrate various types of inhomogeneous and homogeneous scattering volumes including (a) squid body and fin: an irregularly shaped object in which the ray path is traced alternately into squid tissue and seawater (treated as an inhomogeneous medium), [(b) and (c)] squid mantle and mantle cavity: a continuous external boundary with internal inhomogeneities, and (d) squid neck: a homogeneous volume.

The cumulative phase change is calculated by first digitally rotating the volume matrix using "nearest neighbor" interpolation, such that the incident wavenumber vector, $\boldsymbol{k}_1$, is orthogonal to the new cross sections $\tilde{M}^l$. The rotation limits the phase-tracking to a one-dimensional problem. The differential phase change is calculated for each voxel from the locally assigned sound speed and differential length over which the sound wave travels. The total one-way phase change is then found as the cumulative summation of these differential phase changes from the zero-phase reference plane to this voxel.

Finally, a search routine creates a linear index of all voxels within the rotated volume assigned a particular material property and matches the appropriate amplitude term with the round-trip-phase term. Using a Cartesian coordinate system (x,y,z), the analytical expression given in Eq. (5) becomes

$$f_{\mathrm{bs}} = \frac{k_1^2}{4\pi} \sum_{l=1}^{N_z} \sum_{j=1}^{N_y} \sum_{i=1}^{N_x} (\gamma_\kappa - \gamma_\rho)_{ij}^l e^{2\mathrm{i}\varphi_{ij}^l} dv, \tag{9}$$

where dv is the product of the differential lengths, dx, dy, and dz, and the phase term, φ_{ij}^l, is a cumulative summation of the phase change, in the z direction, and is given by

$$\varphi_{ij}^l = \sum_{q=1}^{l} k_{ij}^q dz. \tag{10}$$

D. Modal-series-based scattering models

In order to validate the phase-tracking DWBA method for inhomogeneous objects, comparisons are made with analytical modal-series-based solutions for scattering from simple geometric, shelled, and hence inhomogeneous objects. The wave equation can be separated and solved exactly for a limited number of simple shapes (e.g., spheres, infinite cylinders, and prolate spheroids), resulting in modal series solutions. The modal series solutions for these simple geometric volumes are derived, using the far-field condition, after separating the wave equation in the respective coordinate system and applying two boundary conditions at each interface: continuity of pressure and continuity of radial velocity. In this section scattering models are presented for fluid-filled spherical and cylindrical (finite-length) fluid shells that are based on these modal series solutions.

The solutions to the spherical and infinite cylindrical shells are exact. The model for the finite-length cylindrical shell, however, is adapted from the latter case and is an approximation. This approximation neglects end effects and is, therefore, accurate only for near-normal incidence. Accordingly, this formulation is only applied at normal incidence to test the phase-tracking DWBA model over a range of frequencies, shell thicknesses, and material properties.

1. Fluid-filled spherical fluid shell

The far-field modal-series-based solution to the wave equation for a fluid-filled spherical fluid shell in the backscattered direction is expressed as

$$f_{\mathrm{bs}} = \frac{\mathrm{i}}{k_1} \sum_{n=0}^{\infty} (2n+1)(-1)^n A_n, \tag{11}$$

where A_n is the nth modal series coefficient and is given in the Appendix. This solution is valid for all sound speeds and densities, shell thicknesses, and acoustic frequencies.

2. Fluid-filled cylindrical (finite-length) fluid shell

Stanton (1988) related the scattering amplitude of a finite-length cylinder to that of an infinite-length cylinder using an approximation in which the volume flow per unit length of the scattered field of the cylinder is approximated by that of the infinite cylinder. These equations can, likewise, be applied to a finite cylindrical shell. Using the same approximation, the backscattering by a fluid-filled cylindrical (finite-length) fluid shell at normal incidence is

$$f_{\mathrm{bs}} = \frac{\mathrm{i}L}{\pi} \sum_{n=0}^{\infty} -\epsilon_n (-1)^n B_n, \tag{12}$$

where B_n is the nth modal series coefficient for the *infinite-length* fluid-filled cylindrical fluid shell and is given in the Appendix. The term ϵ_n is the Neumann number ($\epsilon_0=1$, $\epsilon_n=2$ for $n=1,2,3,\ldots$).

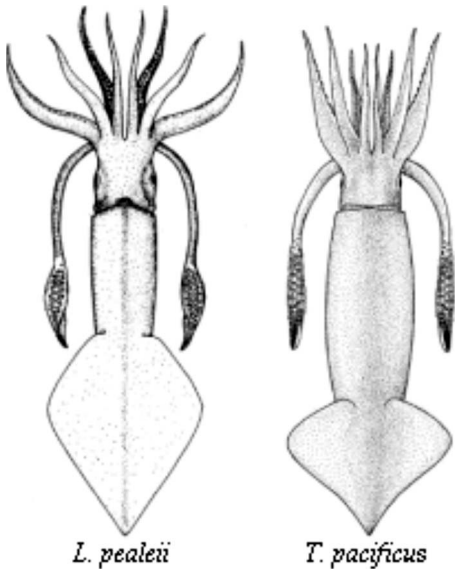


FIG. 3. Sketches of the species of squid discussed in this article (Roper *et al.*, 1984) The *L. pealeii*, shown on the left, is the littoral species found near Woods Hole, MA, from which high resolution morphometry was obtained using SCT scans. The *T. pacificus* shown on the right, a pelagic species found in the western Pacific Ocean, has a smaller aspect ratio (length-to-width ratio). Published acoustic scattering data from experiments involving *T. pacificus* are used to compare with phase-tracking DWBA predictions.

III. MATERIALS AND METHODS

A. Animals studied

Two species of squid are used in this research (Fig. 3). The morphometry of a locally available long-finned squid, *L. pealeii*, is measured at high resolution, while published target strength data are available for a second species, *T. pacificus*, or the Japanese common squid. Scattering predictions are made based on the three-dimensional morphology of *L. pealeii*, digitally scaled to match the aspect ratio (i.e., length-to-width ratio) of a *T. pacificus* and compared with published data. In one case, partial scans of two different *L. pealeii* were combined so that the arms and tentacles of the free-swimming *T. pacificus* could best be modeled.

1. Long-finned squid, Loligo pealeii

L. pealeii specimens, obtained from the Marine Biological Laboratory, Woods Hole, MA, were selected due to ease of availability and well documented details of their anatomy.

TABLE I. Material properties of squid from published sources. Values given as a ratio of squid material properties to those of seawater in the same publication.

Reference	Material	g	h
Hashimoto and Maniwa (1952)[a]	Whole squid	⋯	1.007[b]
Mukai *et al.* (2000)	Whole squid	1.025	⋯
Kang *et al.* (2006)	Whole squid	1.029	1.041
Iida *et al.* (2006)	Mantle tissue	1.043	1.053

[a]In Japanese, cited by Kang *et al.* (2006).
[b]Based on assumed sound speed in seawater of 1500 m/s.

This coastal species of squid has a long, slender body and a large fin in proportion to its mantle length. These squid reach maturity at a mantle length of approximately 13–15 cm and approach a maximum mantle length of 40–50 cm (Roper *et al.*, 1984).

2. Japanese common squid, Todarodes pacificus

T. pacificus is an oceanic squid with a muscular, moderately slender body and a relatively short fin. Members of this species reach maturity at a mantle length of 17–25 cm and a maximum size of 50 cm (Roper *et al.*, 1984). Although these squid have a smaller aspect ratio than *L. pealeii*, the two species are broadly similar (Fig. 3).

3. Squid material properties

Acoustic scattering predictions require accurate values of sound speed and density of the scatterer. This is especially true of weakly scattering organisms whose material properties may vary from the surrounding water by only several percent (Chu *et al.*, 2000). Very little has been published on squid material properties (Table I). However, some recent measurements provide useful information for this study. For example, Kang *et al.* (2006) measured average density and sound speed of whole anesthetized squid in seawater. Their estimates of sound speed as compared to previously reported values show better agreement with measured TS when used as parameters in a Kirchhoff ray mode model. Iida *et al.* (2006) recently measured values of sound speed and density of squid mantle tissue. These values are higher than those reported for whole squid. This is expected, however, as these measurements are for a single tissue type, whereas the whole squid measurements are average values of sound speed and density through both the squid tissue and the seawater-filled mantle cavity. In light of these results, the values given by Iida *et al.* (2006) were used in the phase-tracking DWBA method ($g_v=1.043$ and $h_v=1.053$) to represent the various parts of the squid's body (i.e., mantle, internal organs, etc.). Seawater-filled cavities were given values of 1 for g_v and h_v.

B. High resolution morphometry of squid: SCT scans

SCT scans were taken of a live, anesthetized squid (Fig. 4) and of a dead, previously frozen squid at the Marine Research Facility at the Woods Hole Oceanographic Institution. A spiral CT instrument conducts measurements in a helical pattern by taking a circular scan while the body is continuously advanced through the scanning aperture. Cross sections are then reconstructed by interpolation into two-dimensional images, with each pixel containing the x-ray attenuation in Hounsfield units (Hofer, 2000).

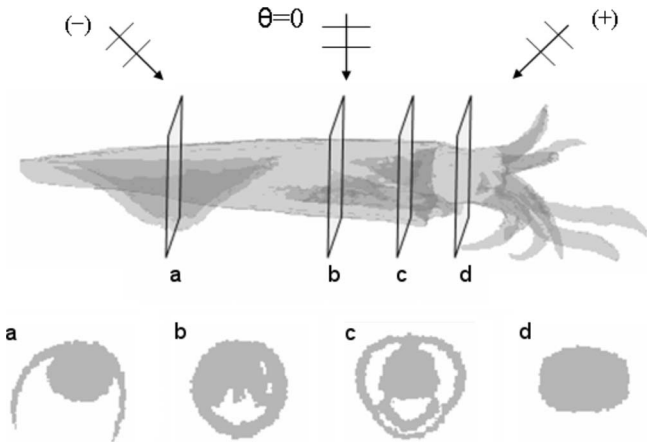


FIG. 4. Volume rendering composed of SCT scan images of *L. pealeii* (in semitransparent gray-scale) and four binary cross sections, M_l, showing the (a) fin, [(b) and (c)] mantle cavity, and (d) neck. In the cross sections, gray indicates squid body and white indicates surrounding seawater and seawater-filled cavities. Arrows show the orientation of the squid with respect to an incident sound wave (−/+ values of θ correspond to the head being down/up, respectively, and $\theta=0$ corresponds to normal incidence).

The live squid measured in this study was anesthetized in a 1% ethanol seawater solution. During scanning, the animal was suspended vertically in seawater to minimize distortion of body shape caused by laying the animal, unsupported, on a hard surface. A contrast agent was added to the seawater to help distinguish between x-ray attenuation of the solution and the squid body. The animal did not recover from the anesthesia and died either during the scanning process or soon after completion of the scan. The second specimen, a previously frozen dead squid was measured while suspended vertically in air. Scans of both specimens were taken by a Siemens Volume Zoom four-slice SCT scanner using pitch and collimator settings of 2 and 1 mm, respectively. Images of 512×512 pixels were reconstructed from the data using 0.5 mm slice thickness. Individual pixel size was 0.5078×0.5078 mm^2. Images were later resized on separate mathematical software using bicubic interpolation such that all pixels represent one cubic volume element, or voxel, with dimensions, l_{vs}, equal to $0.5\times0.5\times0.5$ mm^3. In this way dimensional ratios are preserved in the image rotations discussed in the next section.

C. Numerical implementation of SCT scans

The SCT images of squid were cropped to the size of the largest cross section and then converted to binary matrices, M^l, using a simple threshold technique (Fig. 4). Scattering predictions were made by applying Eq. (9) to the resultant three-dimensional matrix. For incorporation into the phase-tracking DWBA method (inhomogeneous case), each voxel, M^l_{ij}, identified as seawater, both inside and outside the body, was assigned a value of 0, while voxels representing the squid's body were assigned a value of 1. Tissue sound speed and density could not be directly inferred from SCT data. Thus, for modeling purposes all tissues not identified as sea-

water were assumed to have material properties of mantle tissue. The binary conversion was insensitive to the threshold level in the cases of both squid scanned for this study. In the live, anesthetized case the contrasting agent made the seawater significantly higher in x-ray attenuation than the squid's body. In the case of the dead specimen scanned in air, the reverse was true, with the squid's body exhibiting much higher x-ray attenuation than the surrounding air.

1. Scaled volume

In order to model scattering by *T. pacificus*, the morphometry derived from the SCT scans of *L. pealeii* was scaled so that the aspect ratio (i.e., the ratio of mantle length to mantle width) was approximately equal to that of the *T. pacificus*. The aspect ratio of *T. pacificus* is approximately 10% smaller than that of *L. pealeii*.

2. Hybridized volume

One focus of this study was to incorporate a realistic swimming position of squid by using SCT scans of the live, anesthetized specimen. However, the anesthetized squid splayed its arms outward (Fig. 4) in a position that is not representative of the squid's natural swimming position (i.e., arms and tentacles folded into an anvil shape). In order to obtain a more realistic swimming morphometry, SCT scans of the dead squid, of similar size to the anesthetized squid, were taken with arms and tentacles folded together. The measurements of the more naturally positioned arms and tentacles from this animal were used to create a hybrid squid with the anesthetized squid's body. This hybrid squid was scaled as discussed in the previous section and used in the phase-tracking DWBA model. These predictions are compared with phase-tracking DWBA predictions for the nonhybridized live squid and with published measurements.

D. Numerical issues related to phase-tracking approach

The accuracy of digitization of a volume limits the range of frequencies for which the phase-tracking DWBA method can be applied. In general the ratio of acoustic wavelength to maximum voxel dimension should be about 20:1 although this requirement depends on the modeling approach used (Stanton and Chu, 2000). A closely related limitation is the computational time. Calculating the frequency-dependent backscattering amplitude, f_{bs}, for broadband sound, where many finely incremented discrete frequencies are considered, can take a significant amount of computing time. Similarly, calculating the scattering as a function of orientation of the object at a discrete frequency is computationally intensive. In this model calculating f_{bs} as a function of orientation required significantly more processing time than calculating f_{bs} as a function of frequency, per discretization, due to the digital rotation of the volume.

IV. RESULTS

A. Scattering from fluid-filled spherical and cylindrical (finite-length) fluid shells

The phase-tracking DWBA method is applied to computer-generated volumes of two simple shapes, spherical and finite-length cylindrical shells. In both cases the shell itself and interior volume are fluid materials. Phase-tracking DWBA predictions for spherical shells are compared with the exact modal series solution for fluid-filled spherical fluid shells [Eq. (11)]. Scattering predictions of cylindrical shells are compared, at normal incidence, with the modal-series-based solution for fluid-filled cylindrical (finite-length) fluid shells [Eq. (12)] and at various orientations for homogeneous finite-length cylinders with the DWBA finite-length cylinder model (closed form) [Eq. (8)].

1. Shell thickness

Scattering predictions for fluid shells of various thicknesses are compared with modal-series-based results to verify that the phase-tracking DWBA method accurately accounts for inhomogeneities (Figs. 5 and 6). Shell thickness is used as a proxy for various-sized inhomogeneities. Cylindrical and spherical shells, with fractional shell thickness, τ (see definition in Fig. 1 caption) varying from 100% (i.e., homogeneous) to 10%, are used for comparison. In all cases, the density and sound speed of the shells are set to 1% greater than both the interior fluid and the surrounding fluid. All predictions show that scattering from spherical shells and cylindrical shells at normal incidence is strongly affected by the presence of inhomogeneities. Small scale structure was observed due to interference between waves scattered from the front and back interfaces of the object. While these interference patterns, for differing shell thicknesses, are nearly identical in terms of null spacing, the large scale structure varies significantly.

Comparison between the phase-tracking DWBA predictions and the modal series solutions shows excellent agreement. In the spherical case (Fig. 5), the numerical model predictions were generally within ± 1 dB of the modal series solution for all values of shell thicknesses up to values of k_1a of 12.5 ($\lambda/l_{vs}=50.3$) and within ± 2 dB up to values of k_1a of 17.5 s ($\lambda/l_{vs}=35.9$). In the cylindrical case at normal incidence (Fig. 6) the phase-tracking DWBA predictions matched the modal-series-based solution generally within ± 0.5 dB for k_1a up to 5 ($\lambda/l_{vs}=50.3$) and ± 2 dB for $k_1a \leqslant 12.5$ ($\lambda/l_{vs}=20.2$). In both cases some large discrepancies (>10 dB) existed; however, these were isolated to nulls of very low scattering values. As discussed in a following section, the disagreement between the phase-tracking DWBA predictions and modal-series-based solutions at higher values of k_1a is a numerical issue related to voxel resolution. Further comparisons using homogeneous spheres of increased resolution confirm that the disagreement seen at higher k_1a is primarily due to resolution issues (results not shown).

2. Material properties

As discussed above, the DWBA formulation is only valid for weak scatterers. In order to investigate the range of

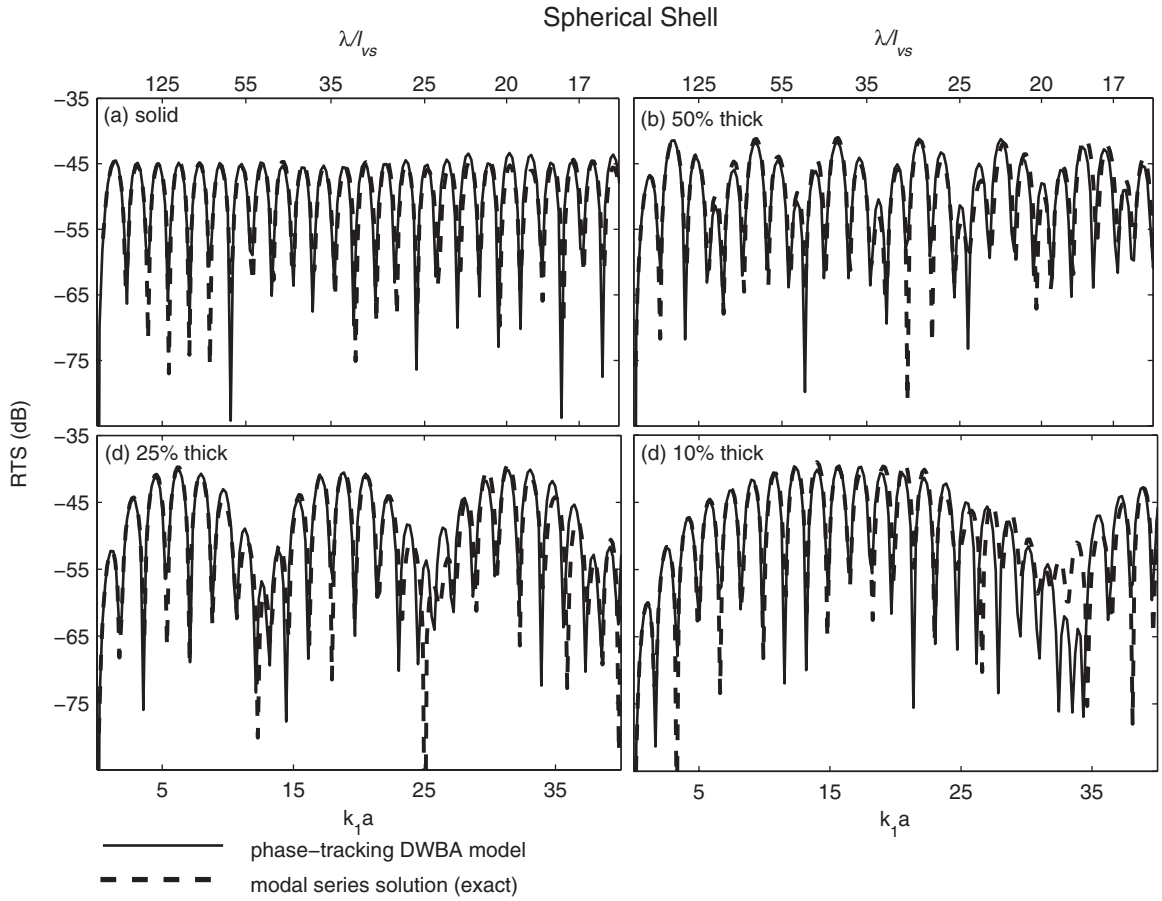


FIG. 5. Effect of varying shell thickness on backscattering by a fluid-filled spherical fluid shell. Phase-tracking DWBA predictions (solid lines) are compared with exact modal series solutions [Eq. (11)] (dashed lines). [Parameters: material properties: $g_2=1.01$, $h_2=1.01$, $g_1=g_3=1$, $h_1=h_3=1$, and $l_{vs}=0.01a$]

accuracy for the phase-tracking DWBA method with respect to material properties, finite-length cylindrical shells with various sound speed and density contrasts are considered. Model predictions at normal incidence are compared with the modal-series-based fluid-filled cylindrical (finite-length) fluid shell solution [Eq. (12)] that can describe a wide range of material profiles.

Scattering predictions were made using this model for finite-length cylindrical shells with fractional shell thickness of 25% and g and h values ranging from 1.02 to 1.06 (results not shown). Frequency was varied such that the maximum k_1a ranged from 0 to 10, which equates to a minimum resolution of $\lambda/l_{vs} \simeq 25$. As a result of increasing sound speed and density contrast, the values of RTS increased correspondingly. There was excellent agreement outside the deepest portion of the nulls (± 1 dB) for $k_1a \leqslant 5$ ($\lambda/l_{vs} \simeq 50$). Errors of up to 2 dB were seen at higher values of k_1a and are largely due to resolution issues discussed in a following section.

3. Orientation of scattering object

In order to test the phase-tracking DWBA method for sound waves incident on an object at various angles, scattering from a homogeneous finite-length cylinder at a fixed value of k_1a and varying tilt angles, θ, was considered. Since the modal-series-based solution for finite cylinders is only accurate near broadside incidence, the phase-tracking DWBA method was compared with the DWBA finite-length cylinder model (closed form) [Eq. (8)], which is valid for all orientations. When evaluated at the highest resolution, the phase-tracking DWBA method showed very good agreement with the analytic solution (± 1 dB) except in the deepest nulls [Fig. 7(a)].

It should be noted that the value chosen for k_1a (5.03) in this portion of the study was neither in the deepest part of the null nor at a peak in the frequency response curve [Fig. 6(a)]. As seen in all the modeling results [Figs. 5–7 and 7(a)], the accuracy of the phase-tracking DWBA predictions at very low scattering values (i.e., near or in very deep nulls) is worst. This discrepancy is magnified because of the logarithmic scale. It is expected that this error for low values of predicted scattering is somewhat mitigated by the fact that very deep nulls are not as common in real organisms as in ideal shapes. The presence of these nulls in scattering from ideal shapes comes from nearly complete destructive interference of coherent scattering from the front and back interfaces.

4. Effects of varying voxel resolution

In order to test the accuracy of the phase-tracking DWBA method at different voxel resolutions, scattering predictions for homogeneous, cylindrical volumes of the same dimensions and varying voxel size, l_{vs}, are compared to the DWBA finite-length cylinder model (closed form) [Eq. (8)].

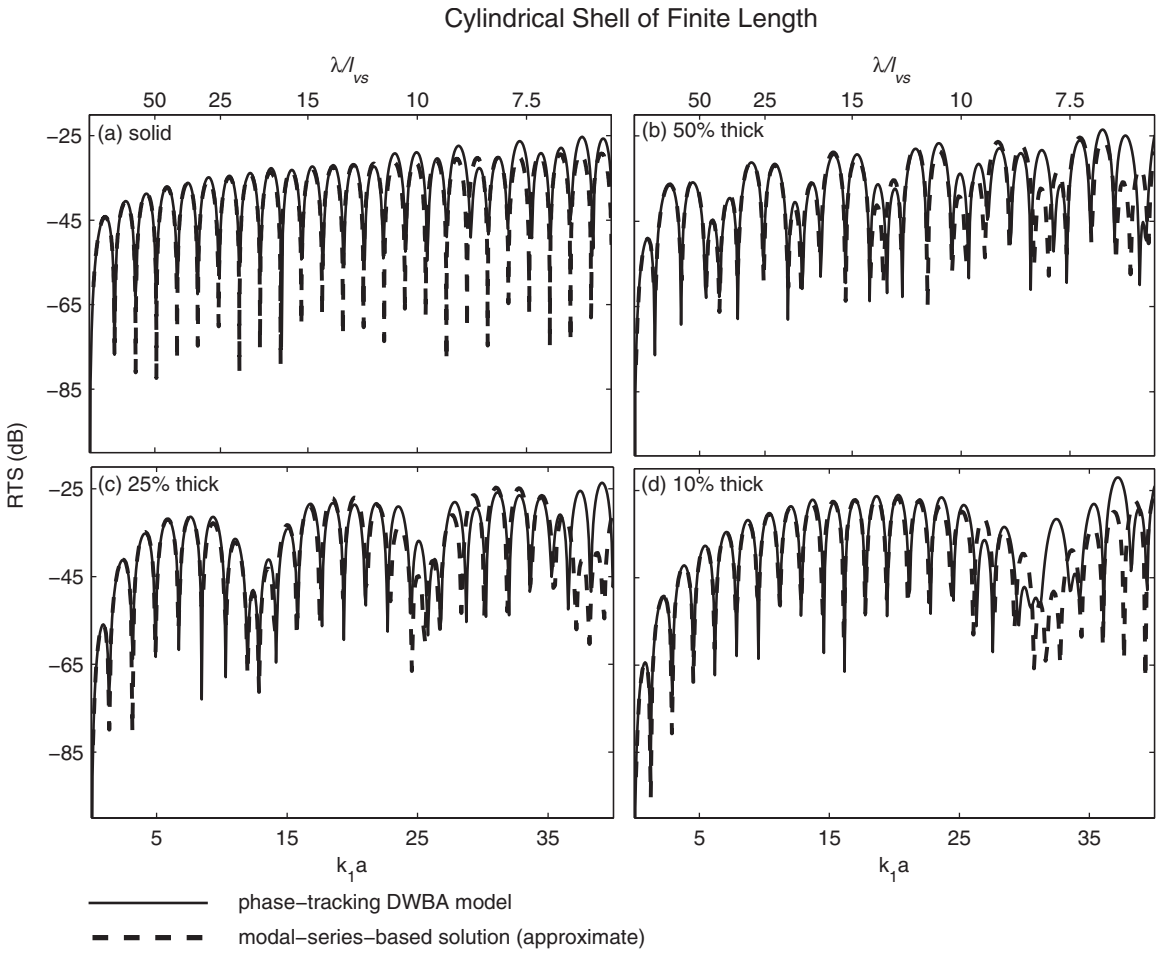


FIG. 6. Effect of varying shell thickness on backscattering by a fluid-filled cylindrical (finite-length) fluid shell. Phase-tracking DWBA predictions (solid lines) are compared with an approximate modal-series-based solution [Eq. (12)] (dashed lines). [Parameters: aspect ratio: $L/2a=2.5$; angle of incident wave: $\theta=0$ (normal incidence); material properties: $g_2=1.01$, $h_2=1.01$, $g_1=g_3=1$, $h_1=h_3=1$, and $l_{vs}=0.025a$.]

Resolution, expressed as wavelength with respect to voxel size (λ/l_{vs}), is varied from 75 to 12.5 for the RTS versus orientation predictions given in Figs. 7(a)–7(d). The model predictions matched the analytical solution fairly well for the large scale structure at a minimum resolution of 12.5 and were considerably better at a resolution of 25. The small scale structure was generally within ±2 dB at a resolution of 50.

The effects of resolution can also be seen in the results of broadside scattering from simple geometrical shapes. Figures 5 and 6 show how increasing k_1a, thus decreasing λ/l_{vs}, affects the predictions. In the case of a spherical shell (Fig. 5), an accuracy of ±2 dB requires a resolution of approximately 35. In the cylinder case a similar accuracy can be obtained by a resolution of approximately 20.

Although no clear pattern emerged, it has been shown that the resolution at which an object needs to be digitized for this phase-tracking DWBA approach depends on the scattering geometry. For an object with two-dimensional curvature, as with a cylinder, a minimum resolution of 20 provides an accuracy at normal incidence of ±2 dB, whereas an object with three-dimensional curvature, as in the case of a sphere, requires a somewhat higher level of resolution ($\lambda/l_{vs} \geqslant 35$). Finally, predictions of scattering by a cylinder over all angles of incidence require a significantly higher resolution for similar accuracy ($\lambda/l_{vs} \geqslant 50$).

B. Application to squid

The phase-tracking DWBA method with high resolution morphology derived from SCT scans of squid is first used to compare homogeneous and inhomogeneous model predictions for squid. Then, inhomogeneous model predictions are compared with measurements of scattering by live, freely swimming squid (Arnaya *et al.*, 1989b, 1989c; Kang *et al.*, 2005), as well as by tethered, anesthetized squid (Kang *et al.*, 2005). These comparisons are made using both single-orientation and tilt-averaged results, where tilt is defined as rotation about the lateral axis of the squid (e.g., positive tilt is head up). All measurements of scattering by live squid concern only the species *T. pacificus*. Material properties for squid tissue and swimming tilt-angle distributions come from published sources, and, therefore, there were no floating parameters used in these predictions.

1. Effect of inhomogeneities on scattering predictions

In order to analyze the effects of including internal inhomogeneities (i.e., the seawater-filled mantle cavity) in modeling predictions, phase-tracking DWBA predictions using the high resolution morphometry of the live, anesthetized squid are compared for the homogeneous and inhomogeneous cases. Scattering predictions were made versus k_1L_{ml} for dorsal aspect backscattering at various angles of tilt. The

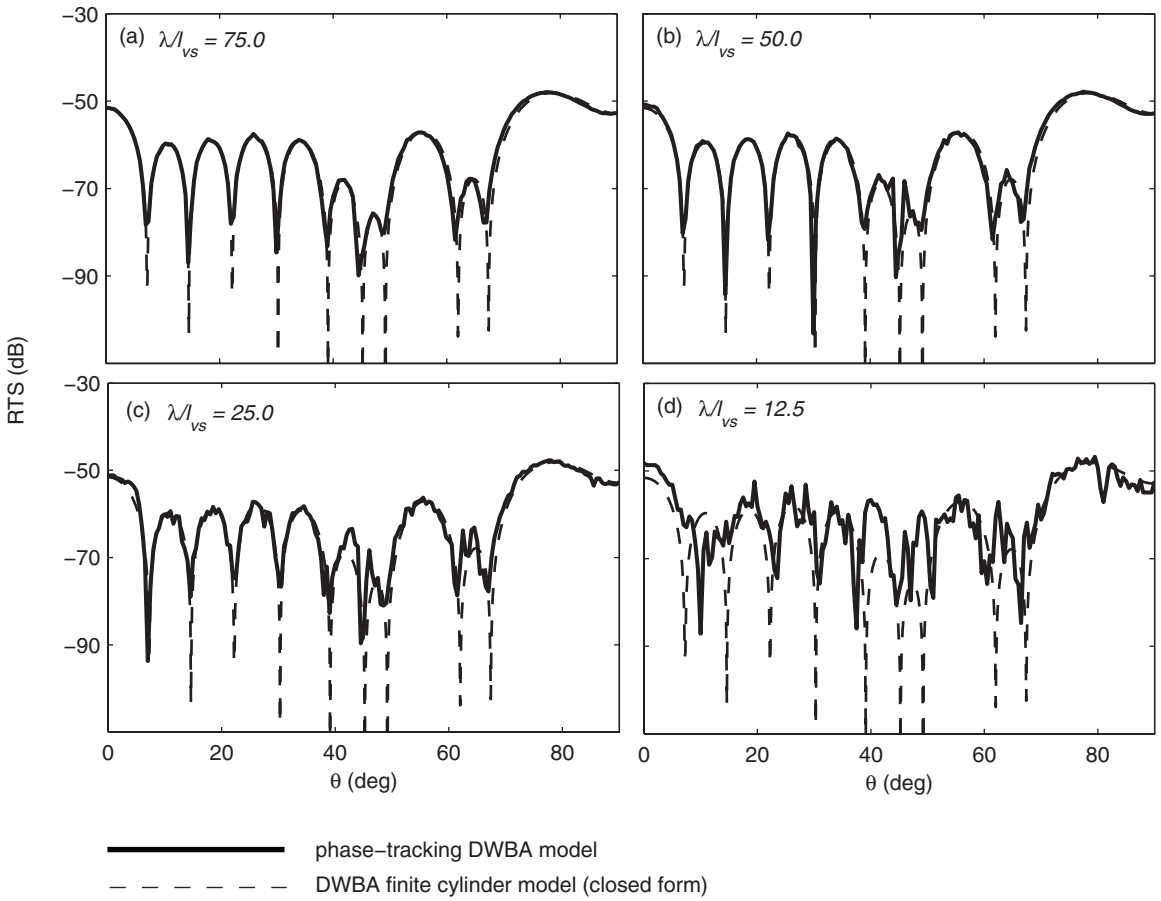


FIG. 7. Predictions of backscattering vs orientation for a homogeneous finite-length cylinder. Phase-tracking DWBA predictions (solid lines) are compared with DWBA finite-length cylinder model (closed form) [Eq. (8)] (dashed lines). [Parameters: $g=1.01$, $h=1.01$, $k_1a=5.03$, and aspect ratio, $L/2a=2.5$.]

mantle chamber represents 5.6% of the squid's body by volume. The effect of including this seawater-filled cavity is generally a $\pm 1-3$ dB difference outside the deepest nulls as illustrated by predictions at four discrete angles ranging from $-30°$ to $+30°$ [Figs. 8(a)–8(d)], with the inhomogeneous model tending to predict higher TSs.

The effect of varying the position of the appendages, also treated by the phase-tracking DWBA method as volume inhomogeneities, can be seen by comparing the model predictions for the splayed-arms case and the straight-arms, hybrid-construction case with published measurements at various tilt angles (Fig. 9). By incorporating the more natural swimming position of the squid arms and tentacles (straight-arms case), scattering predictions were improved for a range of angles off-normal incidence ($-20° < \theta < -10°$) by up to 5 dB. It is clear, however, that predictions of far off-normal angles of incidence for models using both positions of the appendages underestimate the actual backscattered TS by 5–10 dB or more. Furthermore, the scattering predictions show considerably more variability than the data associated with both tethered squid and freely swimming squid, although at least one set of measured data involved averages (tethered squid) that could contribute to the differences.

Additionally, comparisons were made between the phase-tracking DWBA predictions (high resolution morphometry, inhomogeneous case) and DWBA predictions using a homogeneous prolate spheroid. When these models are compared at normal, dorsal incidence there are significant differences noticed in the structure of the frequency response [Fig. 8(a)]. While the prolate spheroid has a very regularly spaced interference pattern, the frequency response predicted by the phase-tracking DWBA method for the complex morphology of the squid tends to vary in both null spacing and peak amplitude.

2. Effect of tilt angle on scattering predictions

Scattering predictions versus tilt angle for dorsal aspect backscattering by *T. pacificus* are compared to predictions using the phase-tracking DWBA method (high resolution morphometry, inhomogeneous material properties, and straight- and splayed-arms cases) and DWBA predictions for a homogeneous prolate spheroid of equivalent volume and similar radius (Fig. 9). The frequency of the incident wave, as a model parameter, was chosen to correspond to a value of k_1L_{ml} that matched the frequency and mantle length of squid used in the published study of scattering by squid. The published results of live squid ensonified at 38 kHz (Kang *et al.*, 2005) include freely swimming squid, ensonified by a split-beam transducer with swimming angle determined from perpendicular side-view cameras, and anesthetized squid tethered at various angles of tilt. The phase-tracking DWBA

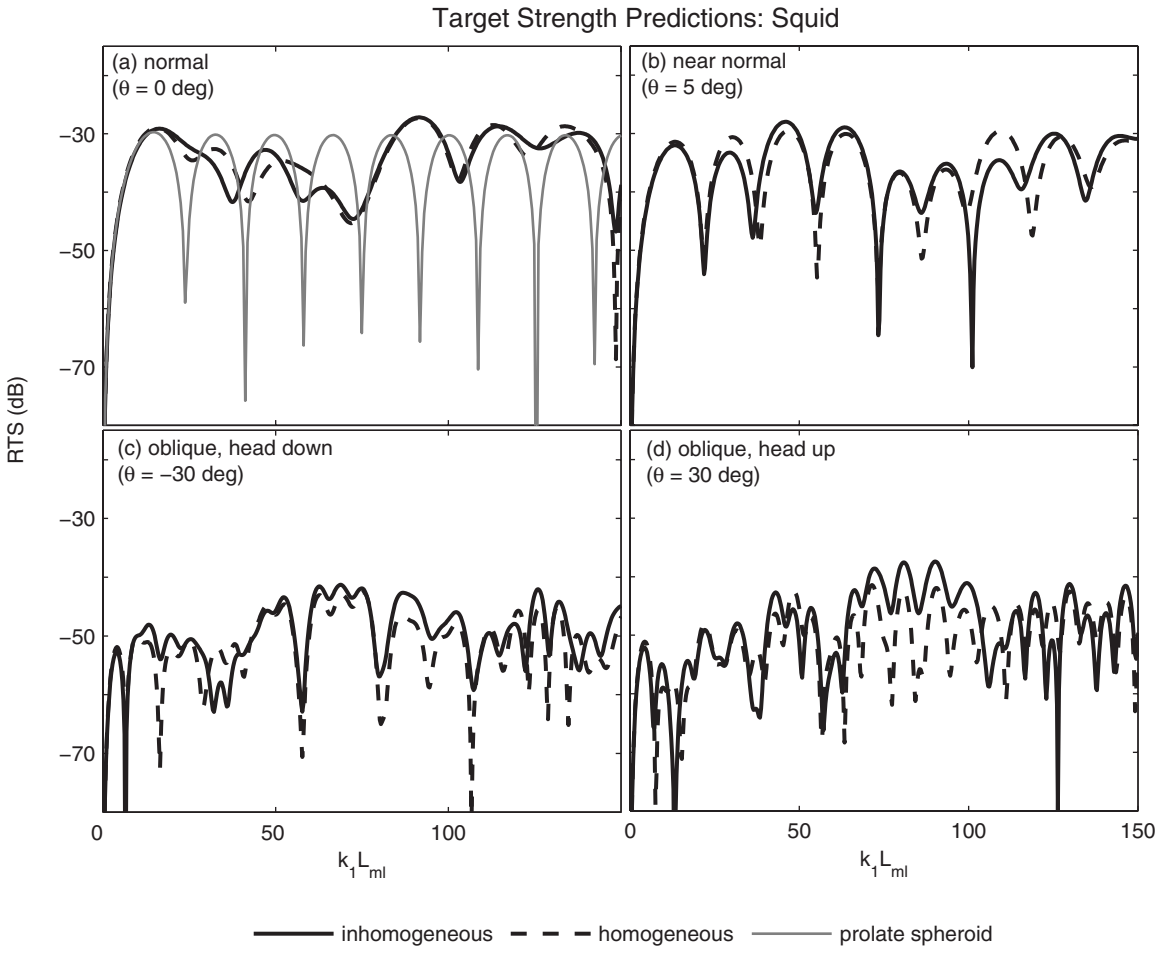


FIG. 8. Predictions of scattering by squid (*T. pacificus*) for various fixed orientations using scaled morphometry of *L. pealeii* (splayed arms) with homogeneous material properties (dashed lines) and inhomogeneous material properties, which includes interior structure of mantle cavity (solid lines). Predictions using the DWBA-based prolate spheroid model are shown for normal incidence only (solid gray line).

predictions (straight-arms case) agree well between −20° and +30° tilt in the anesthetized case. This model is significantly more accurate than the prolate spheroid model which strongly underpredicts the observed scattering levels at angles off-normal incidence ($|\theta| \geqslant 5°$).

3. Comparison of model predictions with averaged target strength measurements

Scattering predictions with respect to tilt angle are made for squid by compiling phase-tracking DWBA predictions for two degree angle increments from −50° to +50°. Results were weighted using two different tilt distributions of freely swimming squid from published sources (Kang *et al.*, 2004; Arnaya *et al.*, 1989a). Published measurements from three studies on TSs of *T. pacificus* are compared with these predictions (Fig. 10). The model predictions are cut off at a value of $k_1 L_{\mathrm{ml}}$ of 150 corresponding to a minimum resolution, $\lambda / l_{\mathrm{vs}}$, of 12.5. The acoustic measurements come from three sources. Measurements from Arnaya *et al.* (1989b, 1989c) were estimated from mean volume backscattering strength. The data from Kang *et al.* (2005) were measured from individual squid using split-beam transducers at 38 and 120 kHz. Coincident tilt distribution of squid measured in these studies was not available. Thus, the predictions presented here use both tilt distributions to compare with measured data.

Phase-tracking DWBA predictions and published data were compared using differential backscattering amplitude, where the mean absolute error is defined as $\hat{E} = \langle |\sigma_{\mathrm{bs}}^{\mathrm{data}} - \sigma_{\mathrm{bs}}^{\mathrm{pred}}| \rangle$ and the root mean squared error is defined as $E_{\mathrm{rms}} = \sqrt{\langle |\sigma_{\mathrm{bs}}^{\mathrm{data}} - \sigma_{\mathrm{bs}}^{\mathrm{pred}}|^2 \rangle}$. The phase-tracking DWBA predictions (splayed-arms case) using a tilt distribution of $[-4, 11.1]$ showed better agreement ($\hat{E} = 1.23 \times 10^{-4}$, $E_{\mathrm{rms}} = 1.52 \times 10^{-4}$) than the corresponding predictions using the tilt distribution of $[-17.7, 12.7]$ ($\hat{E} = 2.26 \times 10^{-4}$, $E_{\mathrm{rms}} = 2.46 \times 10^{-4}$), as seen in Fig. 10(a). Furthermore, the model predictions using the straight-arms hybrid morphometry of squid and a tilt distribution of $[-4, 11.1]$ [Fig. 10(b)] agreed most closely with the data ($\hat{E} = 9.69 \times 10^{-5}$, $E_{\mathrm{rms}} = 1.25 \times 10^{-4}$).

Finally, the phase-tracking DWBA predictions are compared with DWBA predictions for a homogeneous prolate spheroid of an equivalent volume and similar radius. Both predictions use a tilt distribution of $[-4, 11.1]$ due to the better fit with observed scattering data. By incorporating the high resolution digitizing of the outer shape only (i.e., using a homogeneous splayed-arms squid scaled to match the aspect ratio of the measured species), an improvement of

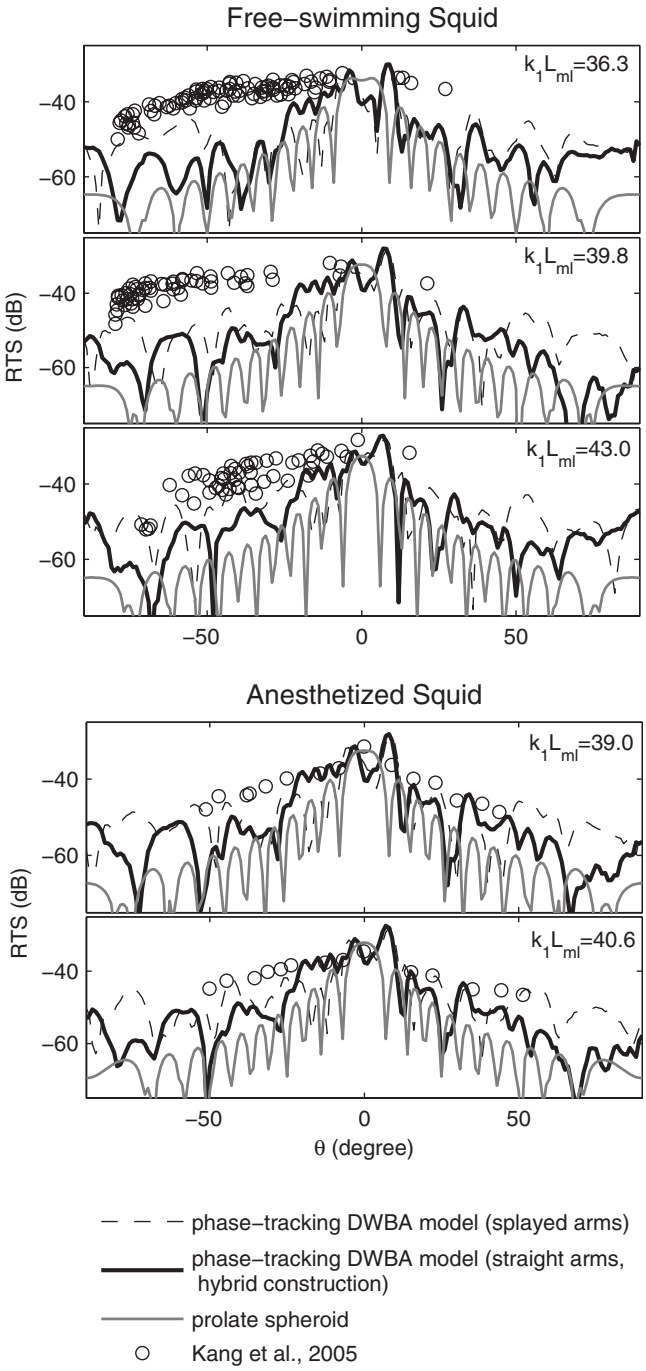


FIG. 9. Predictions of scattering by squid (*T. pacificus*) vs tilt angle using scaled morphometry of *L. pealeii* compared with experimental data. Black lines (solid and dashed) show phase-tracking DWBA predictions, with no averaging, using high resolution morphometry of *L. pealeii* and inhomogeneous material properties scaled to match the aspect ratio of *T. pacificus*. Dashed black lines show predictions using morphometry of the anesthetized squid (splayed-arms case). Solid black lines show predictions using hybrid morphometry of the two specimens of *L. pealeii* (straight-arms case). Solid gray lines show DWBA model predictions using an equivalent volume, homogeneous prolate spheroid with a similar radius. Predictions are compared with published measurements of *T. pacificus* (data points). Top three plots show data from live, freely swimming squid, ensonified at 38 kHz, while bottom two plots show data from anesthetized tethered squid also ensonified at 38 kHz (Kang *et al.*, 2005)

2–6 dB over the homogeneous prolate spheroid model was obtained for a wide range of *ka* (results not shown). Replacing the homogeneous splayed-arms squid with an inhomogeneous straight-arms squid resulted in predictions with further improvements of generally 0–2 dB. These improvements were through increases in level. In some narrow ranges of *kL*, the increases were 1–3 dB. Overall, the phase-tracking DWBA method, incorporating inhomogeneities and more realistic swimming positions (i.e., straight arms and tentacles), generally produced a 3–7 dB improvement in predictions over the homogeneous prolate spheroid model [Fig. 10(b)].

V. DISCUSSION AND CONCLUSION

A method that accurately accounts for scattering by an inhomogeneous weakly scattering object has been developed. This new method takes advantage of high resolution interior and exterior measurements of morphology and rigorously accounts for both the amplitude and phase of the scattered wave by using a phase-tracking approach to the DWBA. The inhomogeneous object can take the form of (1) variations of material properties within the body of the scatterer and/or (2) appendages or a part of the body that may be oriented such that the incident sound wave first passes through part of the body, into the surrounding fluid, and then back into another part of the body. This numerical method is applicable for all frequencies and all angles of scatterer orientation and can be applied to arbitrarily shaped three-dimensional objects. The only constraint in the weakly scattering formulation is that the material properties (density and sound speed) may not deviate by more than several percent of the surrounding fluid.

The phase-tracking DWBA method has been tested against modal-series-based solutions to spherical and cylindrical shells and the DWBA finite-length cylinder model (closed form). There was excellent agreement in the comparisons over the entire range of shell thickness and tilt angle both with respect to overall scattering levels and structure of the scattering (i.e., position of the nulls). Furthermore, the model predictions agreed well with the analytic solutions to scattering over the range of material properties considered in this study. Additionally, an important result emerged from the modal-series-based solutions for scattering by objects of varying shell thicknesses. It has been shown here that the nonaveraged broadband frequency response of a weakly scattering inhomogeneous object cannot be accurately modeled by a homogeneous object of the same outer form.

The phase-tracking approach, in part, addresses a long-standing issue associated with predicting backscattering by weakly scattering, inhomogeneous, and/or irregularly shaped elongated bodies at angles well away from normal incidence. It has been shown in previous studies that predictions of scattering by elongated zooplankton, averaged over orientations near normal incidence, are relatively insensitive to small variations in shape and material properties (Stanton and Chu, 2000). This has been attributed to the dominance of scattering from the front and back interfaces of the elongated scatterer. It has also been demonstrated that once the organism is oriented well-off broadside, these interfaces contribute less, making variations of shape and material property significant. This issue is common to a wide range of species of marine organisms, including commonly occurring zooplankton such as copepods and euphausiids. There have been at-

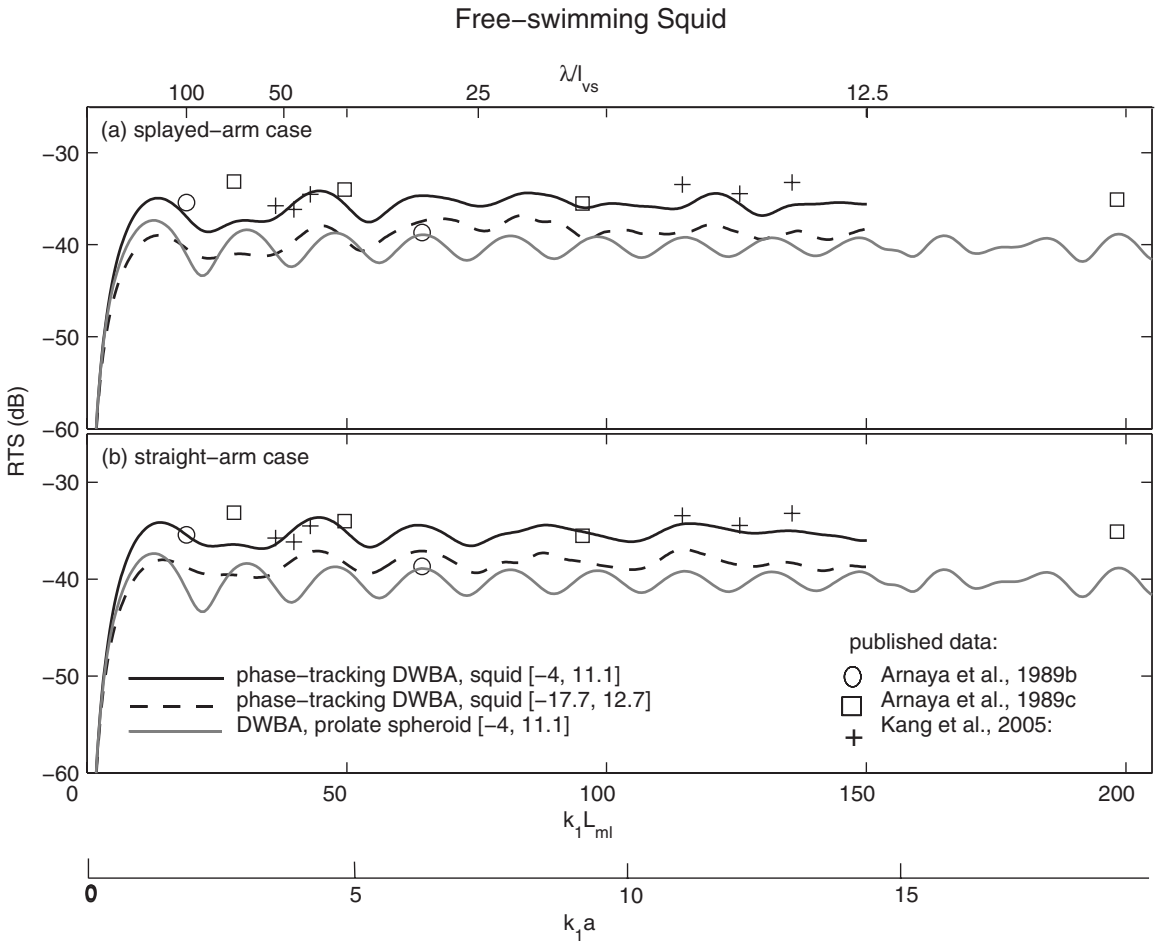


FIG. 10. Average scattering predictions for squid (*T. pacificus*) compared with experimental data. Predictions of scattering by squid use inhomogeneous material properties and (a) scaled morphometry of the anesthetized *L. pealeii* and (b) hybrid morphometry of the two specimens of *L. pealeii* to incorporate more natural swimming positions of the squid's arms and tentacles, also scaled. All predictions are averaged over the tilt angle using normal distributions limited to ± 2 standard deviations of the mean. The two tilt-angle distributions presented use mean angle and standard deviation, $[\bar{\theta}, s]$, from published sources of swimming squid: $[-4, 11.1]$ from Arnaya *et al.* (1989a), solid black lines, and $[-17.7, 12.7]$ from Kang *et al.* (2004), dashed black lines. Solid gray lines show DWBA model predictions using an equivalent volume, homogeneous prolate spheroid with a similar radius. Predictions are compared with published measurements of live, freely swimming squid, *T. pacificus*: circles (Arnaya *et al.*, 1989b) squares (Arnaya *et al.*, 1989c), and plus signs (Kang *et al.*, 2005).

tempts to address this issue in earlier studies by modeling the shape of the organisms at high resolution (i.e., resolution cells much smaller than an acoustic wavelength). For example, in Stanton *et al.* (1998), the body of a euphausiid was assumed to be randomly rough at scales of $\lambda/10$, with circular symmetry at each point of the body; in Stanton and Chu (2000), the outer boundaries of both euphausiids and copepods were digitized at $\lambda/10$, again with circular symmetry at each point; in Lavery *et al.* (2002), CT scans were used to digitize the outer boundary in three dimensions at a minimum resolution of $\lambda/10$; in Amakasu and Furusawa (2006) appendages were separately digitized and incorporated into the modeling of the scattering by euphausiids. Furthermore, some of these studies included, quite heuristically, inhomogeneities within the body by applying various profiles of material properties to the scattering object. In each of the studies, the backscattering levels for angles well-off-normal incidence were shown to increase significantly due to incorporation of the fine features of the body (surficial roughness, inhomogeneities within the body, appendages). However, none of the studies accounted for the fact that the phase change of the acoustic wave is altered as a result of passing through an inhomogeneity of differing sound speed. An important consequence of the phase-tracking approach is the inherent improvement in predictions of scattering at off-normal angles of incidence where inhomogeneities contribute significantly to the scattering.

The phase-tracking DWBA method also directly addresses the issues associated with modeling the scattering by squid. Laboratory data show that the scattering by squid is strongly dependent on tilt angle. Since free-swimming squid are known to swim at tilt angles roughly 4°–20° away from horizontal, quantitative interpretation of data from vertically aimed echosounders is one case requiring a scattering model that is accurate over a range of orientation angles. Prior to this study, the most advanced model of squid assumed it to be a homogeneous prolate spheroid that did not account for details of the shape of the body or inhomogeneities. As shown in our studies, the prolate spheroid model produces reasonable predictions (when compared with laboratory data) for the scattering by squid at near-normal dorsal incidence. However, our predictions also show that this model significantly underpredicts the scattering for angles away from normal incidence. When measured data and predictions are av-

eraged over the observed distribution of swimming angles, the prolate spheroid again significantly underpredicts the data. By incorporating high resolution measurements of the outer shape and internal structure of squid obtained from spiral CT scans, the phase-tracking DWBA method produces significantly improved predictions over the DWBA, homogeneous prolate spheroid model. Improvements were seen in the TS levels for both single-orientation and tilt-averaged predictions and were most notable for single-orientation predictions at a range of off-normal angles of incidence.

Improvements in the scattering predictions for squid over earlier methods are a result of three aspects of the modeling of shape. These included high resolution depiction of the outer form of the squid body, internal inhomogeneities, and natural swimming position of squid appendages (e.g., arms and tentacles). When combined, these additional levels of complexity yielded a 3–7 dB improvement in the tilt-averaged predictions. Note that an improvement of 3 dB corresponds to a factor of 2 improvement in estimates of numerical density of the squid. As a result of these improvements, the tilt-averaged predictions with the new method are generally consistent with measured data over a wide range of ka. Also, for predictions at single angles of orientation, at or near-normal incidence, there were differences of 1–3 dB between the homogeneous and inhomogeneous cases.

Although broadband measurements of scattering by squid were not available for this study, another implication of rigorously accounting for inhomogeneities in a DWBA scattering model is that more accurate predictions may be made of frequency responses of weakly scattering marine organisms at discrete angles of orientation. Information about the frequency-dependent structure of individual echoes from organisms such as squid may aid in studies of prey discrimination by echolocating toothed whales and in the use of broadband scientific echo sounders for spectral classification of organisms.

For angles of incidence well away from normal incidence (and also well away from expected swimming angles), the new approach still significantly underpredicts the experimental data. There could be several sources of this discrepancy. For example, in this method it is assumed that the entire body is a weak scatterer where the material properties deviated from the surrounding fluid by only several percent. However, it is possible that there are small portions of the body that have much higher density and sound speed contrasts and can contribute significantly to the scattering under some conditions. These portions include a long thin gladius (or pen), a parrotlike beak, statoliths (small, bony inner ear organs), a hard cartilaginous cranium, and hard lenses in the eye (Williams, 1909). Iida *et al.* (2006) also identified the liver, a low density organ surrounded by higher density tissue, as a likely scatterer of sound. There are also uncertainties in the degree to which the tissue is inhomogeneous and there are uncertainties associated with the exact positioning of the appendages of the squid. Since the scattering by the main interfaces of the body is small at large angles of incidence, any of these other sources (or some combination) might contribute significantly to the scattering. These other sources are not a particular concern in this study since they do not appear to contribute significantly to the scattering at tilt angles associated with the swimming position of squid. However, in applications where the angles are much larger, those other sources need to be studied such as through the methods described in Nash *et al.* (1987) or Iida *et al.* (2006).

In conclusion, a new method for predicting acoustic scattering by weakly scattering bodies has been developed through applying a phase-tracking approach to the DWBA. This approach rigorously accounts for phase shifts due to inhomogeneities and appendages of the body. Furthermore, it makes use of high resolution three-dimensional digitizing of the body through spiral CT scans of the body. When applied to squid, the approach represents a significant improvement over the previous model of a low resolution homogeneous prolate spheroid. The improvements were based entirely on measured quantities and did not rely on any "floating" parameters. Although the application in this paper involved squid, this approach has the capability of making accurate predictions of scattering by a wide variety of important marine organisms in the fluidlike anatomical group, including euphausiids and copepods. The main challenges in all of the modeling will be the determination of material properties (sound speed and density) and orientation distribution, accounting for parts of the bodies that are not weak scatterers and obtaining sufficiently high resolution measurements of shape for small organisms acoustically sampled at high frequencies.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Darlene Ketten at Woods Hole Oceanographic Institution and the colleagues in her group, particularly Julie Arruda, R.T.(R) who conducted the CT scans of squid presented here and Sr. Research Asst. Scott Cramer who aided in the handling, transport, and preservation of the specimens. The squid specimens used in this study were identified and provided by Ed Enos, Superintendent of the Aquatic Resources Division at Marine Biological Laboratory. We would also like to mention Dr. Roger Hanlon who provided great perspective concerning squid. Finally, the guidance and literature provided by Dr. Kohji Iida were most helpful in the comparisons with measured scattering by squid conducted in this research. Funding of this research was provided, in part, by the Office of the Oceanographer of the Navy and the Academic Programs Office at the Woods Hole Oceanographic Institution.

APPENDIX: MODAL SERIES COEFFICIENTS FOR FLUID-FILLED FLUID SHELLS

Modal series solutions for spherical and cylindrical scatterers have been derived for a wide range of material property profiles such as solid elastic, fluid, and fluid-filled elastic shells (Anderson, 1950; Goodman and Stern, 1962; Stanton, 1988; Faran, 1951). To the best of our knowledge, explicit expressions for the modal series coefficients for fluid-filled fluid shells have not been published; therefore, the following details are provided for spherical and cylindrical shells. In the following equations, the properties of the fluid surround-

ing a scatterer are denoted by the subscript "1," the fluid shell of the scatterer is denoted by subscript "2," and the fluid inside the shell is denoted by the subscript "3" (Fig. 1).

1. Spherical shells

The nth modal series coefficient for a fluid-filled spherical fluid shell expressed as the variable, A_n, in Eq. (11) is given below as the ratio of the determinants of two matrices. These matrices contain the spherical Bessel functions of the first kind, j_n, the second kind, η_n, and the third kind (also known as the spherical Hankel functions of the first kind), $h_n^{(1)}$. Primes on the Bessel functions [e.g., $h_n^{(1)\prime}(k_1a)$] indicate derivatives with respect to their argument

$$A_n = \frac{\begin{vmatrix} j_n(k_1a) & 0 & -j_n(k_2a) & -\eta_n(k_2a) \\ j_n'(k_1a) & 0 & -\frac{\rho_1c_1}{\rho_2c_2}j_n'(k_2a) & -\frac{\rho_1c_1}{\rho_2c_2}\eta_n'(k_2a) \\ 0 & -j_n(k_3b) & j_n(k_2b) & \eta_n(k_2b) \\ 0 & -\frac{\rho_2c_2}{\rho_3c_3}j_n'(k_3b) & j_n'(k_2b) & \eta_n'(k_2b) \end{vmatrix}}{\begin{vmatrix} h_n^{(1)}(k_1a) & 0 & -j_n(k_2a) & -\eta_n(k_2a) \\ h_n^{(1)\prime}(k_1a) & 0 & -\frac{\rho_1c_1}{\rho_2c_2}j_n'(k_2a) & -\frac{\rho_1c_1}{\rho_2c_2}\eta_n'(k_2a) \\ 0 & -j_n(k_3b) & j_n(k_2b) & \eta_n(k_2b) \\ 0 & -\frac{\rho_2c_2}{\rho_3c_3}j_n'(k_3b) & j_n'(k_2b) & \eta_n'(k_2b) \end{vmatrix}}. \tag{A1}$$

2. Cylindrical shells

The nth modal series coefficient for a fluid-filled cylindrical (infinite-length) fluid shell, expressed as the variable, B_n, in Eq. (12), is given below as the ratio of the determinants of two matrices. These matrices contain the cylindrical Bessel functions of the first kind, J_n, the second kind (also known as the Neumann functions), N_n, and the third kind (also known as the Hankel functions of the first kind), $H_n^{(1)}$. Primes on the Bessel functions [e.g. $H_n^{(1)\prime}(k_1a)$] indicate derivatives with respect to their argument,

$$B_n = \frac{\begin{vmatrix} J_n(k_1a) & 0 & -J_n(k_2a) & -N_n(k_2a) \\ J_n'(k_1a) & 0 & -\frac{\rho_1c_1}{\rho_2c_2}J_n'(k_2a) & -\frac{\rho_1c_1}{\rho_2c_2}N_n'(k_2a) \\ 0 & -J_n(k_3b) & J_n(k_2b) & N_n(k_2b) \\ 0 & -\frac{\rho_2c_2}{\rho_3c_3}J_n'(k_3b) & J_n'(k_2b) & N_n'(k_2b) \end{vmatrix}}{\begin{vmatrix} H_n^{(1)}(k_1a) & 0 & -J_n(k_2a) & -N_n(k_2a) \\ H_n^{(1)\prime}(k_1a) & 0 & -\frac{\rho_1c_1}{\rho_2c_2}J_n'(k_2a) & -\frac{\rho_1c_1}{\rho_2c_2}N_n'(k_2a) \\ 0 & -J_n(k_3b) & J_n(k_2b) & N_n(k_2b) \\ 0 & -\frac{\rho_2c_2}{\rho_3c_3}J_n'(k_3b) & J_n'(k_2b) & N_n'(k_2b) \end{vmatrix}}. \tag{A2}$$

Amakasu, K., and Furusawa, M. (**2006**). "The target strength of Antarctic krill (Euphausia superba) measured by the split-beam method in a small tank at 70 kHz," ICES J. Mar. Sci. **63**, 36–45.
Anderson, V. C. (**1950**). "Sound scattering from a fluid sphere," J. Acoust. Soc. Am. **22**, 426–431.
Arnaya, I. N., and Sano, N. (**1990**). "Studies on acoustic target strength of squid VI. Simulation of squid target strength by prolate spheroidal model," Bull. Fac. Fish., Hokkaido Univ. **41**, 32–42.
Arnaya, I. N., Sano, N., and Iida, K. (**1989a**). "Studies on acoustic target strength of squid II. Effect of behaviour on averaged dorsal aspect target strength," Bull. Fac. Fish., Hokkaido Univ. **40**, 83–99.
Arnaya, I. N., Sano, N., and Iida, K. (**1989b**). "Studies on acoustic target strength of squid III. Measurement of the mean target strength of small live squid," Bull. Fac. Fish., Hokkaido Univ. **40**, 100–115.
Arnaya, I. N., Sano, N., and Iida, K. (**1989c**). "Studies on acoustic target strength of squid IV. Measurement of the mean target strength of relatively large-sized live squid," Bull. Fac. Fish., Hokkaido Univ. **40**, 169–181.
Chu, D. Z., 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. Z., Wiebe, P., and Copley, N. (**2000**). "Inference of material properties of zooplankton from acoustic and resistivity measurements," ICES J. Mar. Sci. **57**, 1128–1142.
Faran, J. (**1951**). "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418.
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**, 338–344.
Hashimoto, T., and Maniwa, Y. (**1952**). "Studies on fish finder III. Studies on sound speed in fish body," Tech. Rep. Fish. Boat **3**, 208–215.
Hofer, M. (**2000**). *CT Teaching Manual* (Thieme, New York).
Horne, J. K. (**2000**). "Acoustic approaches to remote species identification: A review," Fish. Oceanogr. **9**, 356–371.
Iida, K., Takahashi, R., Tang, Y., Mukai, T., and Sato, M. (**2006**). "Observation of marine animals using underwater acoustic camera," Jpn. J. Appl. Phys., Part 1 **45**, 4875–4881.
Johnson, M., Madsen, P. T., Zimmer, W. M. X., de Soto, N. A., and Tyack, P. L. (**2004**). "Beaked whales echolocate on prey," Proc. R. Soc. London, Ser. B **271**, S383–S386.
Johnson, R. (**1977**). "Sound scattering from a fluid sphere revisited," J. Acoust. Soc. Am. **61**, 375–377.
Jones, B. A., Stanton, T. K., Lavery, A. C., Johnson, M. P., Madsen, P. T., and Tyack, P. L. (**2008**). "Classification of broadband echoes from prey of a foraging Blainville's beaked whale," J. Acoust. Soc. Am. **123**, 1753–1762.
Kang, D., Iida, K., Mukai, T., and Hwang, D. (**2004**). "Acoustic target strength of Japanese common squid, Todarodes pacificus, and important parameters influencing its TS: Swimming angle and material properties" in OCEANS'04 MTS/IEEE/TECHNO-OCEAN'04, Kobe, pp. 364–369.
Kang, D., Iida, K., Mukai, T., and Kim, J. (**2006**). "Density and sound speed contrasts of the Japanese common squid Todarodes pacificus and their influence on acoustic target strength," Fish. Sci. **72**, 728–736.
Kang, D., Mukai, T., Iida, K., Hwang, D. J., and Myoung, J. G. (**2005**). "The influence of tilt angle on the acoustic target strength of the Japanese common squid (Todarodes pacificus)," ICES J. Mar. Sci. **62**, 779–789.
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**, 2685–2697.
Lavery, A. C., Stanton, T. K., McGehee, D. E., and Chu, D. (**2002**). "Three-dimensional modeling of acoustic backscattering from fluid-like zooplankton," J. Acoust. Soc. Am. **111**, 1197–1210.
Lavery, A. C., Wiebe, P. H., Stanton, T. K., Gareth, L. L., Benfield, M. C., and Copley, N. (**2007**). "Determining dominant scatterers of sound in mixed zooplankton populations," J. Acoust. Soc. Am. **122**, 3304–3326.
Lawson, G. L., Wiebe, P. H., Ashjian, C. J., Chu, D., and Stanton, T. K. (**2006**). "Improved parametrization of Antarctic krill target strength models," J. Acoust. Soc. Am. **119**, 232–242.
Madsen, P. T., Johnson, M., de Soto, N., Zimmer, W. M. X., and Tyack, P. (**2005**). "Biosonar performance of foraging beaked whales (Mesoplodon densirostris)," J. Exp. Biol. **208**, 181–194.
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**, 1273–1294.
Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York).
Mukai, T., Iida, K., Sakaguchi, K., and Abe, K. (**2000**). "Estimations of squid target strength using a small cage and theoretical scattering models," Proc. JSPS-DGHE Int. Symp. **10**, 135–140.
Nash, D. M., Sun, Y., and Clay, C. S. (**1987**). "High resolution acoustic structure of fish," J. Cons., Cons. Int. Explor. Mer **43**, 23–31.
Roper, C., Sweeney, M., and Nauen, C. (**1984**). *Cephalopods of the World. An Annotated and Illustrated Catalogue of Species of Interest to Fisheries*, Vol. **3** (FAO, Rome).
Simmonds, E. J., and MacLennan, D. N. (**2005**). *Fisheries Acoustics*, Fish and Fisheries (Blackwell Science Ltd., Oxford).
Stanton, T. K. (**1988**). "Sound scattering by cylinders of finite length. I. Fluid cylinders," J. Acoust. Soc. Am. **83**, 55–63.
Stanton, T. K., and Chu, D. (**2000**). "Review and recommendations for the modelling of acoustic scattering by fluid-like elongated zooplankton: Euphausiids and copepods," ICES J. Mar. Sci. **57**, 793–807.
Stanton, T. K., Chu, D., and Wiebe, P. H. (**1998**). "Sound scattering by several zooplankton groups. II. Scattering models," J. Acoust. Soc. Am. **103**, 236–253.
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.
Starr, R., and Thorne, R. (**1998**). "Acoustic assessment of squid stocks," in *Squid Recruitment Dynamics. The Genus Illex as a Model: The Commercial Illex Species. Influences on Variability*, edited by E. O. R. Rodhouse and P. G. Dawe (FAO, Rome).
Tatarski, V. I. (**1961**). *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York).
Williams, L. W. (**1909**). *The Anatomy of the Common Squid, Loligo pealii, Lesueur* (Library and Printing Office, Leiden).
Yeh, C. (**1967**). "Scattering of acoustic waves by a penetrable prolate spheroid. I. Liquid prolate spheroid," J. Acoust. Soc. Am. **42**, 518–521.

Benchmark problems for acoustic scattering from elastic objects in the free field and near the seafloor

Mario Zampolli,[a] Finn B. Jensen, and Alessandra Tesei
NATO Undersea Research Centre, Viale San Bartolomeo 400, 19126 La Spezia, Italy

(Received 6 June 2008; revised 11 September 2008; accepted 26 October 2008)

Results from a workshop organized in 2006 to assess the state of the art in target scatter modeling are presented. The problem set includes free-field scenarios as well as scattering from targets that are proud, half-buried, or fully buried in the sediment. The targets are spheres and cylinders, of size $O(1\ \mathrm{m})$, which are insonified by incident plane waves in the low-frequency band 0.1–10 kHz. In all cases, the quantity of interest is the far-field target strength. The numerical techniques employed fall within three classes: (i) finite-element (FE) methods and (ii) boundary-element (BE) techniques, with different approaches to computing the far field via discretizations of the Helmholtz–Kirchhoff integral in each case, and (iii) semianalytical methods. Reference solutions are identified for all but one of the seven test problems considered. Overall, FE- and BE-based models emerge as those being capable of treating a wider class of problems in terms of target geometry, with the FE method having the additional advantage of being able to deal with complex internal structures without much additional effort. These capabilities are of value for the study of experimental scenarios, which can essentially be envisioned as variations of the problem set presented here.
 [DOI: 10.1121/1.3027446]

PACS number(s): 43.30.Jx, 43.40.Fz, 43.20.Fn [JGM] Pages: 89–98

I. INTRODUCTION

High-fidelity computer models of acoustic scattering from three-dimensional (3D) elastic objects included inside layered acoustic media are of high importance in underwater acoustics, as well as in other fields such as medical ultrasound and nondestructive testing. The models allow quick studies of many insonification scenarios for a variety of objects within different background media configurations, with the scope of determining optimal detection and classification strategies for different frequencies. In underwater acoustic applications, low-frequency sonars (typically operating at frequencies <100 kHz for meter-sized objects) penetrate the outer shell of the target and hence provide information about its internal structure and material properties.

A number of modeling techniques for computing the acoustic scattering from fluid-loaded objects have been developed over the years. Examples include boundary-element (BE) methods,[1–4] and other methods based on the Helmholtz–Kirchhoff (HK) boundary integral formulation of the scattering problem, such as the *T*-matrix target scattering theory,[5–7] which has been developed to compute the scattering from simple objects such as spheres, disks, and hemispherically end-capped cylinders, Kirchhoff approximation models,[8] and wavefield superposition methods.[9–11] An alternative semianalytical technique is based on the recovery of the solution for irregular 3D bodies from solutions of the Helmholtz equation on separable geometries via conformal mapping.[12,13] In addition to these, finite-element (FE) methods[2,14–25] have been and continue to be the focus of large research efforts, mostly because of their capability of treating complex geometries and internal structure within the same numerical framework. Further references on FE methods in acoustics can be found in a review paper by Thompson[26] and in Ref. 27, where both FE and BE methods in acoustics are covered.

Particularly in underwater acoustic studies, the presence of boundaries or interfaces in the vicinity of the scatterer makes it necessary to match the near-field target scattering model with a technique capable of describing the interaction between the target-scattered field, computed by analytical methods or, for example, by one of the techniques described above, and the surrounding complex layered medium. This has been achieved by matching the target-scattered field with the normal mode description of the waveguide,[18,28,29] by describing the multiple scattering between the target and the interface via spherical harmonic addition theorems,[30,31] by combining the target *T*-matrix scattering theory with Green's function representation of the surrounding medium,[32–35] including approximations for treating rough interfaces,[36,37] and by the wavefield superposition method,[38] particularly in conjunction with the wavenumber spectral integral representation of the field in the stratified background medium.[39–45]

Among the aforementioned target scattering models, those based on semianalytical approaches (exact modal expansions and *T*-matrix methods) are mostly capable of treating simple target geometries such as spheres, ellipsoids, and different finite cylinders. Wavefield superposition methods, BE formulations, and FE formulations can be used to study more complex target shapes, while the addition of internal structural details such as the discontinuous transitions between a solid and a void (or, e.g., ribs and flooded partitions) makes it necessary to resort to FE-based models. The presence of domain boundaries and interfaces further restricts the number of techniques capable of treating situations such as

[a]Electronic mail: zampolli@nurc.nato.int

proud, buried, or partially buried targets with internal structural detail. In this sense, there is a need for reference solutions to a set of test problems, which gradually expose the increasing complexity of the scattering physics (ranging from simple perfectly rigid shapes to shells and eventually to objects with internal structure), as well as the presence of boundaries in the vicinity of the target. With the intention of making such solutions available to the community, and to establish the ranges of validity, the accuracy, and the computational speed of different solution methods, a benchmark workshop was organized at NURC (Ref. 46) in September 2006.

This paper presents solutions to a set of the workshop test problems in the frequency range 0.1–10 kHz, computed at NURC, NUWC (NAVSEA Newport, RI), and FOI, the Swedish Defense Research Institute in Stockholm. These solutions have been established after a follow-up to the first round of computations presented in Ref. 46. In all cases, the quantity of interest is the far field, which is more closely related to the quantities usually required in practical applications. A brief overview of the numerical methods employed is given in Sec. II. Section III addresses the problem of scattering from a finite cylinder submerged in an ideally infinite volume of water. The structural degree of complexity of the problem is progressively increased from a perfectly rigid body (homogeneous Neumann boundary conditions), to a void elastic cylindrical shell, and finally to a partially solid-filled elastic cylindrical shell. In Sec. IV the cylindrical shell is placed proud on the interface between a water half-space and a half-space of sand, modeled as a fluid with damping. Section V deals with scattering from a solid-filled elastic spherical shell, which is completely buried or half-buried in the layered water-sand background medium. The challenge in this case lies particularly in the computation of the far field inside the layered background medium, with the echo emanating from an object, which is completely or partially buried in the lower medium. Additional solutions for a subset of the workshop test problems not addressed here have already been presented in earlier publications.[25,45]

II. OVERVIEW OF THE NUMERICAL METHODS EMPLOYED IN THE STUDY

A brief overview of the computational methods used at NURC, FOI, and NUWC to obtain the results presented in this paper is given in this section. The numerical tools were used to compute the target strength (TS) in the frequency band 0.1–10 kHz, sampled with 5 Hz increments. Occasionally, some results may cover slightly different bandwidths or frequency samplings, based on constraints imposed by the code efficiency and the computational resources available in each case.

The standard definition of TS in decibels is $\mathrm{TS}=20\log|rp^{\mathrm{scat}}/p^{\mathrm{inc}}|$, where p^{scat} is the amplitude of the scattered field evaluated in the far field of the target at a distance r (in meters) from the "acoustic center," and p^{inc} is the amplitude of the incident plane wave (see Ref. 47). The distance, r, is chosen here so that the receivers lie in the far field of the scatterer. At such distances, the TS converges to values that become independent of r. A good choice, applicable to all the results presented here, is $r=50$ m. The incident plane wave has a nominal amplitude of $|p^{\mathrm{inc}}|=1$.

A. NURC

AXISCAT, a frequency-domain FE model for computing the radiation and scattering from axially symmetric fluid-loaded structures subject to a nonsymmetric forcing field,[25,45] was used to obtain results for each of the test problems presented here. Using an azimuthal Fourier series expansion of the 3D acoustic and elastic fields, the 3D FE problem is separated into a series of independent smaller two-dimensional (2D) problems. The resulting solution time is smaller by two orders of magnitude compared to the time required for the direct solution of the original 3D problem. The model is implemented by modifying the weak equations in the 2D acoustics and the 2D structural application modes of COMSOL MULTIPHYSICS, a commercially available FE modeling tool.[48] The Bérenger perfectly matched layer (PML) (Refs. 25, 49, and 50) is used to emulate the Sommerfeld radiation condition for free-field targets as well as for targets inside layered fluid media. To maximize the computational efficiency of the tool, the PML should be applied as close as possible to the target. The study presented in Ref. 25 shows, in particular, that the PML can be applied directly on the target-fluid interface without affecting the accuracy of the numerical result.

The TS is computed from the FE solution sampled on the target surface via the HK integral, by employing appropriate Green's functions for the free field or for the layered medium. For those cases where the problem includes the presence of a fluid-bottom interface, an efficient approximation of Green's functions based on the superposition of point sources and image point sources is used. The solutions obtained with the NURC FE tool are referred to as "NURC-FE" below.[25] The label "FE-HK" denotes the use of the NURC-FE model in conjunction with the approximations of Green's function spectral transmission integrals (buried sources) obtained by the method of steepest descent.[45]

For those cases where an axisymmetric structure is located inside a layered medium without breaking the overall axial symmetry of the geometry, such as the problems of Sec. V, the FE-HK approach models all multiple interactions between the target-scattered field and the sediment. If the overall symmetry is broken, but the target is still axisymmetric, such as in Sec. IV, the single-scattering approximation is obtained by computing the scattered field components for the target in the free field, generated by the directly incident field and by the bottom-reflected incident field, respectively.

OASES is a tool for computing the acoustic propagation inside fluid- and elastic-layered media via the wavenumber integral representation of Green's functions.[39–42] The scattering module of OASES includes analytical models for computing the scattering from layered elastic spheres. Solutions can be computed by either using a single-scattering approximation or by computing the full solution using a nodal stiffness matrix describing the target surface impedance, in conjunction with a discrete representation of the HK integral for

the layered medium. The full solution of the scattering problem with OASES requires the inversion of a dense matrix at each frequency, the size of which is equal to twice the number of surface points required to characterize the target impedance. Because of the computational cost of this approach, it is preferrable to adopt the single-scattering approximation whenever the solution is not strongly affected by second-order scattering effects between the target and the water-sediment interface. In the present paper, this is the case for the completely buried sphere problem of Sec. V A, while the higher-order scattering effects, which have a strong effect on the solution of the half-buried sphere problem of Sec. V B, made it necessary to use the computationally more demanding stiffness matrix solution. The OASES solutions are represented by the TS curves labeled "OAS" in the plots.

B. FOI

BE solutions for the rigid cylinder in Sec. III A, and the empty cylindrical shell problems in Secs. III B and IV, were obtained with a high-order boundary integral equation (BIE) tool developed at FOI.[3,4,51] The program is capable of computing the scattering from 3D perfectly rigid bodies and solid-shell bodies. The boundary integrals are computed by a global high-order quadrature technique using tilted spherical coordinates to remove the singularities in the BIE kernel. The target geometry is represented by a map of *B*-splines.

The FOI tool can model targets in layered solid and fluid media by inserting appropriate Green's functions into the HK integral.[44] Green's functions for layered background media are computed by separating the range- and depth-dependent equations via Hankel transforms in range. The depth-dependent ordinary differential equations for each homogenous layer are solved with FEs.[43,52] The basis functions used in this particular FE technique are obtained from the analytical solutions of the source-free ordinary differential equation associated with each layer. For convenience, the solutions obtained at FOI are labeled "FOI-BIE" further below.

C. NUWC

The NUWC solutions were computed with SONAX, the axisymmetric structural acoustic FE modeling tool developed at NRL during the 1990s.[15,16,18,29] The version of SONAX used to obtain the results presented in Sec. III employs unconjugated infinite elements (UIEs) (Refs. 2, 15, and 21) to emulate the Sommerfeld radiation condition. SONAX and AXISCAT both use similar implementations of the azimuthal Fourier mode decomposition of the 3D FE equations, with the main difference being that SONAX uses high-order hierarchic FE basis functions, while AXISCAT uses standard Lagrange polynomials. The comparison of the SONAX-UIE solutions with the AXISCAT-PML solutions makes it possible to verify the effectiveness of the PML, particularly in those cases where it is applied in direct contact with the target. For completeness, it should be mentioned that PMLs have been added into SONAX independently by Shirron and Giddings,[18] in conjunction with a normal mode model to obtain the far field inside a water waveguide overlying a fluid sediment half-space.[29] This new capability, however, is not immediately applicable to the problems studied in Secs. III and V, where the water depth is ideally infinite.

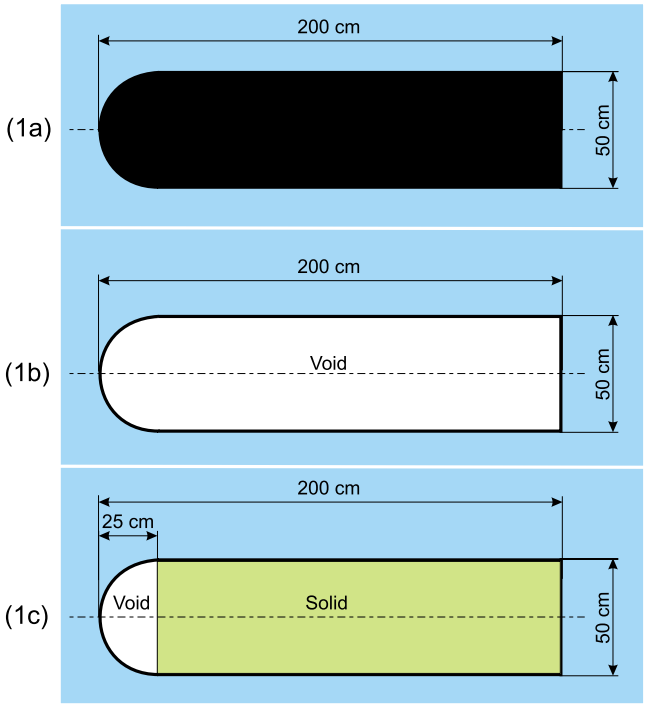


FIG. 1. (Color online) Three cylinders are considered. Perfectly rigid cylinder (a), void cylindrical shell (b), and 90% filled cylindrical shell (c).

III. SUBMERGED CYLINDER

The scatterer is a 2.0 m long cylinder of 0.5 m outer diameter, with one hemispherical and one flat endcap [Fig. 1(a)], submerged in an ideally infinite volume of water. A plane wave insonifies the cylinder at three angles with respect to the cylinder axis: 0°, 70°, and 90°, where 90° is broadside insonification (Fig. 2). The scattered field is computed at predefined spatial positions around the object over the frequency bands of interest. There are three subcases: (a)

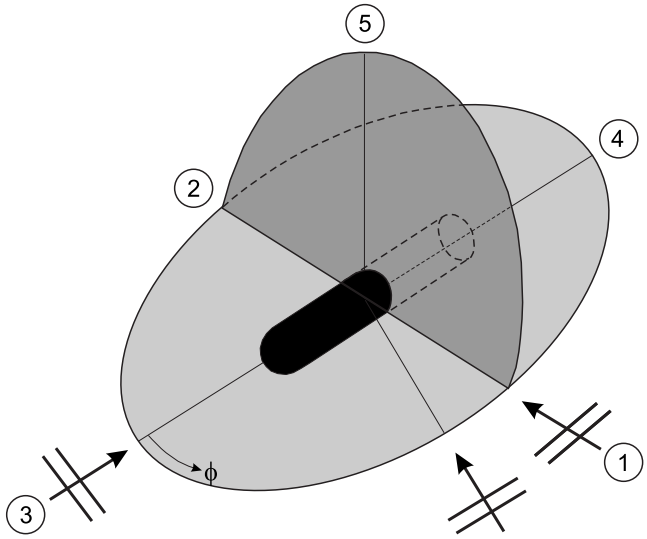


FIG. 2. Geometry of the cylinder scattering problem.

TABLE I. Material properties. A p-wave attenuation of 0.5 dB/λ is assumed for the sediment.

	c_p (m/s)	c_s (m/s)	ρ (kg/m^3)
Water	1500	0	1000
Sediment	1600	0	1800
Fiberglass casing	3500	1400	3000
Solid filling	2500	1200	2000

a perfectly rigid cylinder (zero normal velocity on target surface), (b) a void cylindrical shell of 1 cm wall thickness having the same dimensions as the cylinder of case (a), and (c) the cylindrical shell of case (b) filled with solid material for 90% of the volume. The filling and the casing are assumed to be perfectly bonded. The acoustic properties of the materials used in the models are given in Table I. No damping was considered for the fiberglass shell and the solid filling. The reason for this is that no reliable information on attenuation values for these types of materials were found in literature. Simulations show that even a small damping will have an appreciable effect on the elastic response, causing mostly a smoothing of sharp resonance peaks.

The examples presented here compare the results obtained for the TS as a function of frequency for backscattering at broadside insonification (ϕ=90°), scattering with the receiver in the same direction as for the broadside case, but with the wave incident obliquely from ϕ=70°, and backscattering at end-on insonification (ϕ=0°). These results should give a good basis for benchmarking comparisons to readers interested in using other numerical models than the ones presented here. Additional receiver angles are covered in Ref. 46.

A. Perfectly rigid cylinder

The free-space cylinder is considered to be perfectly rigid (i.e., homogeneous Neumann boundary conditions on the wet surface). This is a calibration case that serves the purpose of checking the general code setup with plane-wave insonification from different angles and far-field TS computations at different look directions. Since the cylinder is considered a rigid body, there is no wave propagation inside the cylinder (no elastic effects) and the scattering problem reduces to a simple reflection problem in the exterior domain for the cylinder. In this sense, the perfectly rigid cylinder example serves also as a first test case for the numerical approximations of the radiation boundary conditions implemented in the various codes.

The solutions for the three different insonification directions, obtained with the NURC-FE, the NUWC-FE, and the FOI-BIE model in the band 0.1–10 kHz, are presented in Fig. 3. The three models give almost identical answers to this scattering problem, and any one of these results could therefore be designated a reference solution. The small differences seen in some of the comparisons could presumably be eliminated by using the same frequency sampling and by pushing the numerical convergence. In particular, these results suggest that the radiation boundary conditions implemented in the three tools, namely, the NURC PML, used in the

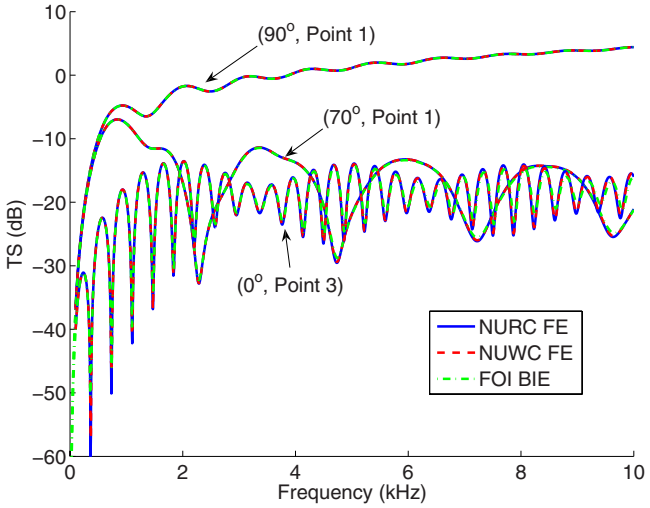


FIG. 3. (Color online) Scattering from a perfectly rigid cylinder.

NURC-FE model,[25] the UIE,[21] used in the NUWC-FE computations,[2,15] and the FOI BE method,[4] yield a satisfactory numerical agreement.

B. Cylindrical shell

A fiberglass shell of 1 cm thickness, of the same shape as the cylinder studied in Sec. III A, is considered in this test case. The shell material is fiberglass with the parameters given in Table I. Note that material damping is ignored. In this case a host of shell-borne Lamb modes will be excited, which, in turn, give rise to elastic resonance effects in the computed TS spectrum. Results for the three different insonification angles, obtained with the NURC-FE, the NUWC-FE, and the FOI-BIE models, are presented in Figs. 4–6. The results agree very well in each of the three test cases, with the agreement between the NURC-FE and the NUWC-FE model being closest. The FOI-BIE results, which have been computed for a set of more sparsely sampled frequencies (20 Hz spacing) compared to the other two models (5 Hz spacing), miss some of the narrow resonance peaks. The dis-

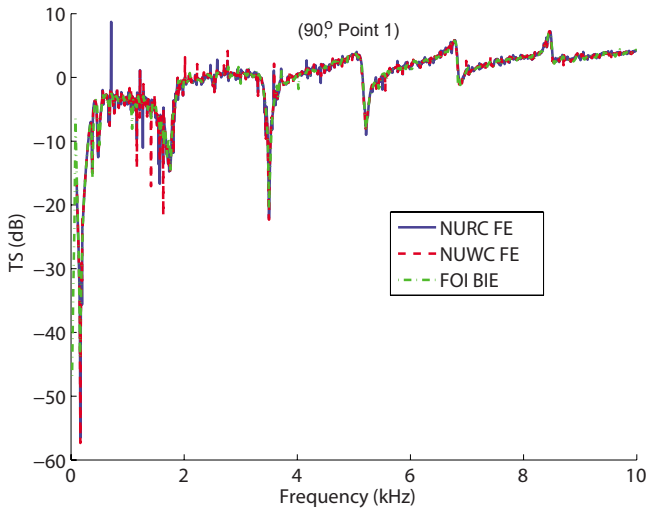


FIG. 4. (Color online) Backscattering from a cylindrical shell, broadside insonification.

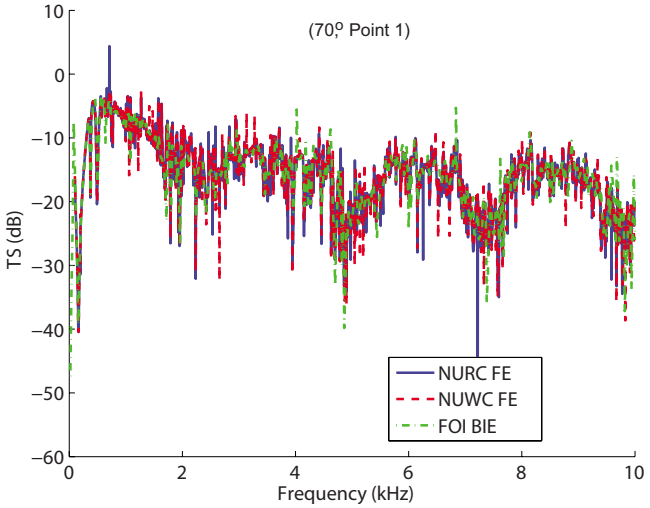


FIG. 5. (Color online) Scattering from a cylindrical shell, oblique insonification at 70°.

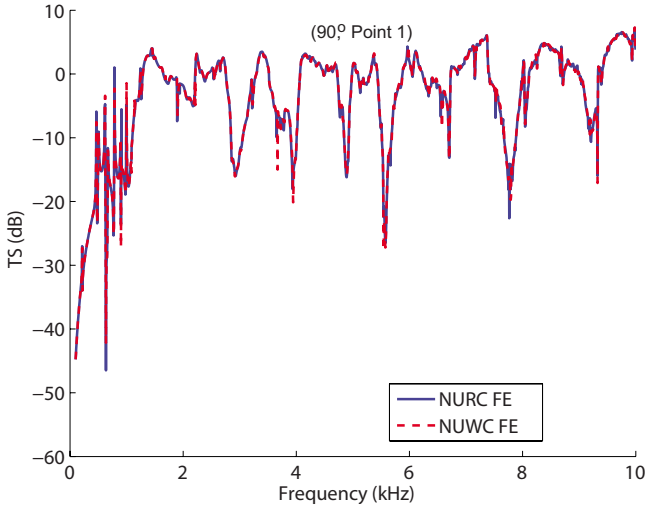


FIG. 7. (Color online) Backscattering from a solid-filled cylindrical shell, broadside insonification.

agreement is more evident in Fig. 5, where the presence of helical shell-borne Lamb waves gives rise to a particularly complex resonance structure. In conclusion, the solutions computed with the NURC-FE and with the NUWC-FE model can be regarded as reference solutions for the cylindrical shell scattering problem.

C. Partially solid-filled cylindrical shell

This is the same cylindrical shell as in Sec. III B, but with a solid filling, as shown in Fig. 1(c). The shell material is fiberglass and the filling is a material with the parameters given in Table I. Note that material damping is ignored in both the shell and the solid filling. In this case, a complex resonance structure is expected due to the coupling of shell-borne waves with body waves in the interior. This test problem is computationally challenging for 3D codes, since the interior of the cylinder is part of the solution domain (more elements, many more degrees of freedom, much longer computation times). For this reason, the problem was not solved with the FOI-BIE code. Figures 7–9 show almost identical results for the NURC-FE and the NUWC-FE model. Due to the computational efficiency of these codes, a check on numerical convergence was possible over the entire frequency band, and the results can be regarded as benchmarks for this structurally complex test problem.

IV. CYLINDRICAL SHELL ON THE INTERFACE BETWEEN TWO FLUIDS

The same cylinder as in Sec. III B is placed on a flat seabed (modeled as a fluid half-space) extending below a half-space of water, as described in Fig. 10. Sediment properties corresponding to sand are listed in Table I, with a realistic p-wave attenuation of 0.5 dB/wavelength assumed for the sandy sediment. By introducing a water-sediment interface near the cylinder, the scattering geometry is no longer axisymmetric, which complicates the issue of generating accurate results for this test problem. In fact, none of the modeling tools applied to the free-space cylinder, except for the 3D BIE code, seem directly applicable to such a problem.

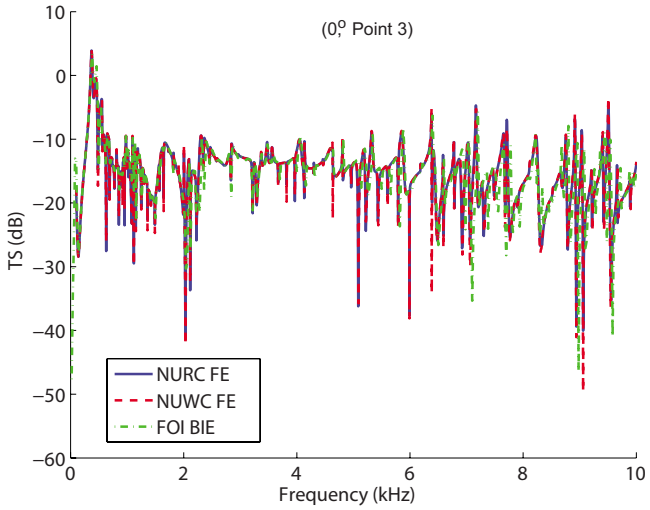


FIG. 6. (Color online) Backscattering from a cylindrical shell, end-on insonification.

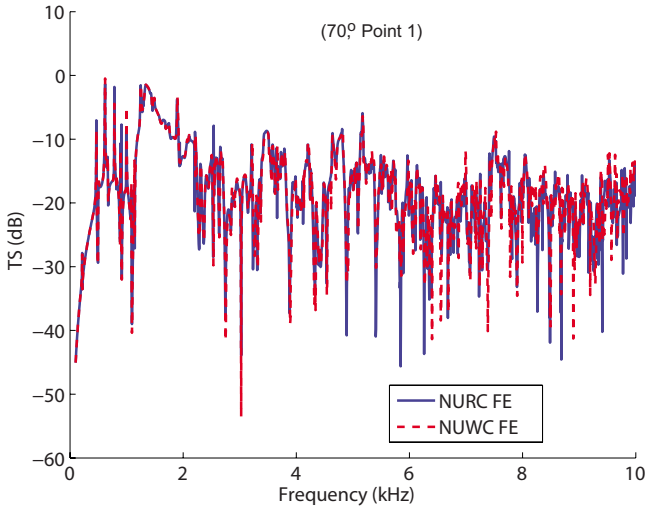


FIG. 8. (Color online) Scattering from a solid-filled cylindrical shell, oblique insonification at 70°.

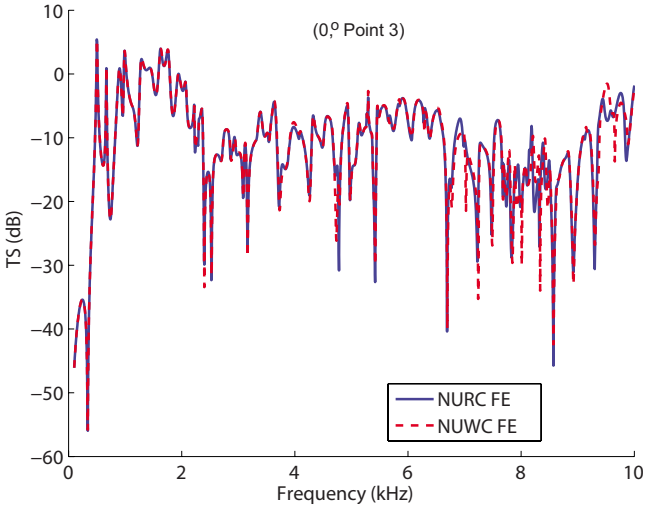


FIG. 9. (Color online) Backscattering from a solid-filled cylindrical shell, end-on insonification.

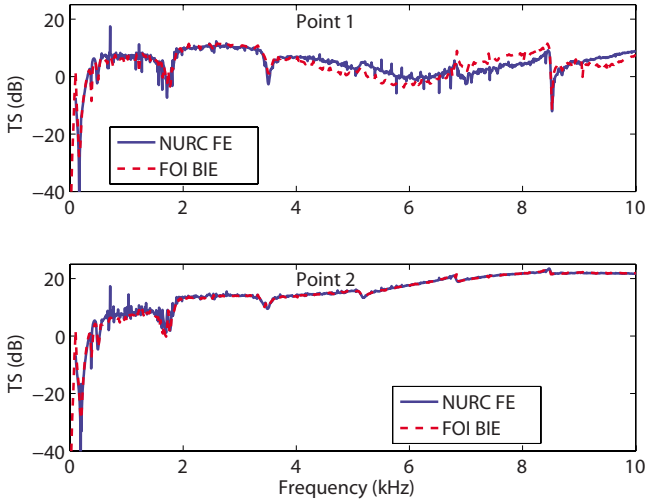


FIG. 11. (Color online) Scattering from a cylindrical shell, proud on the seafloor. Subcritical insonification 15°.

Solutions are presented in Fig. 11 for subcritical (grazing angle $\theta=15°$ insonification) and in Fig. 12 for supercritical ($\theta=25°$ insonification). The agreement between the FOI-BIE and NURC-FE results, particularly for the forward scattered (point 2) solutions, is quite close.

Since this is not an axisymmetric scattering problem, the approximate NURC-FE answer was generated as a superposition of two plane-wave solutions, one at an angle θ with respect to the bottom, and one at an angle $-\theta$, with modified amplitude and phase corresponding to the bottom-reflected incident plane wave. In the successive step, the two scattering problems are solved for the cylinder in free space (axisymmetric, no interface), using the NURC FE code. Finally, the far-field TS is computed via the HK integral using a raylike first-order steepest descent approximation for the two-layer Green's function spectral integral.[25] This approach has been shown to provide accurate results for other scattering problems, despite the two approximations involved: no multiples between cylinder and water-bottom interface, and the use of the approximation in the far-field computation.

The results of the fully 3D FOI-BIE model runs for this case appear to be the most accurate ones. The NURC-FE solution in Fig. 11 for backscattering at a grazing incidence of 15° has the same resonance structure as the FOI solution, but levels are off by approximately 3 dB in the frequency range 4.0–10.0 kHz. The agreement is much better for $\theta=25°$, Fig. 12, where only an interference null around 2.75 kHz is missed by the NURC solution. It is probably the result of ignoring multiple scattering in the NURC solution that causes these disagreements, with stronger effects seen at low grazing angles. In summary, the scattering from a cylinder on the seabed can be computed with good accuracy by both models, and even though one cannot claim benchmark accuracy as for Sec. III, the FOI/NURC solutions in Figs. 11 and 12 represent a reference to within a few decibels over the entire frequency range. The FOI-BIE solutions plotted in Figs. 11 and 12 are smoother compared to the NURC solution, because the computation times associated with the solution of the fully 3D BIE model made it necessary to select a sampling of the frequency spectrum, which is four times coarser compared to the NURC computations.

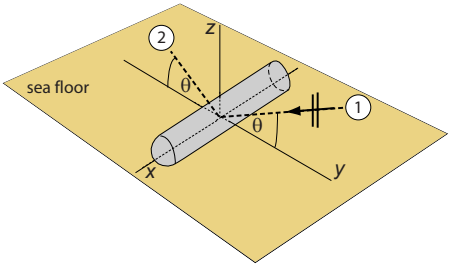


FIG. 10. (Color online) Cylinder proud on a flat seafloor.

A. Partially solid-filled cylindrical shell proud on the seafloor

The partially solid-filled cylindrical shell of Sec. III C is placed on the seafloor according to Fig. 10. The results for backscattering and forward scattering, at subcritical insonification ($\theta=15°$) and supercritical insonification ($\theta=25°$), are

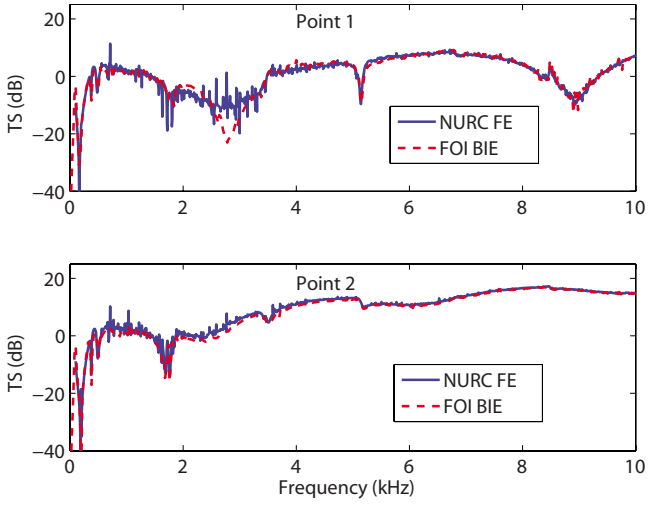


FIG. 12. (Color online) Scattering from a cylindrical shell, proud on the seafloor. Supercritical insonification 25°.

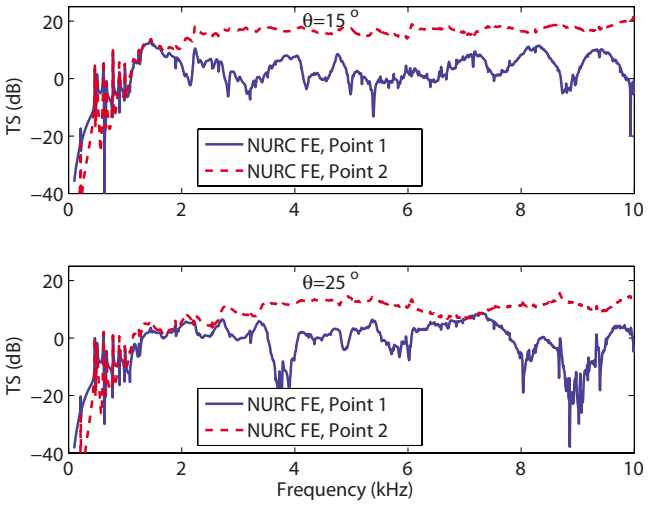


FIG. 13. (Color online) Scattering from a partially solid-filled cylindrical shell, proud on the seafloor. Subcritical insonification 15° and supercritical insonification 25°. No alternative solutions for this problem were available.

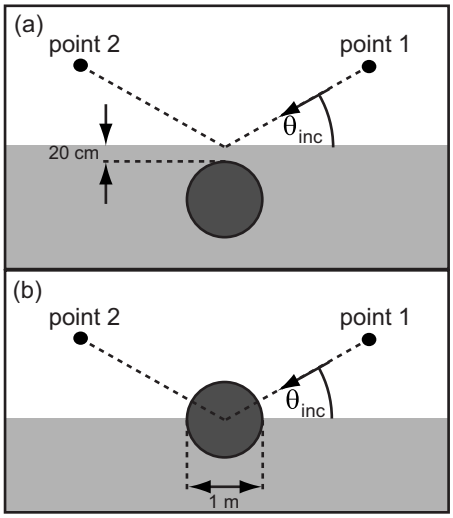


FIG. 14. Scattering from a filled spherical shell. Buried sphere (a) and partially buried sphere (b).

presented in Fig. 13. Given the computational complexity of the problem, the only available solution was obtained with the NURC-FE tool AXISCAT. Although the comparison with alternative solutions was not possible in this case, there is some degree of confidence in the results shown in the figure, which have been obtained using a mesh with the same spatial discretization as the one used for the void cylindrical shell. Hence, by extension of the conclusions drawn for the comparison between the FE result and the BIE result for the proud void cylindrical shell, it is reasonable to expect that the FE solutions of Fig. 13 can be considered as accurate solutions within the limitations of the single-scattering approach.

V. BURIED AND HALF-BURIED SPHERES

For spherically symmetric objects (shells or filled spheres), the plane-wave scattering in a homogeneous medium can be solved analytically. Moreover, specialized codes have been developed to treat this scattering problem in the presence of a water-sediment interface.[18,25,29,32–34,39–42,45] In this section, the attention is turned to a spherical shell of 1.0 m outer diameter and wall thickness 1 cm, filled with the solid material of which the material properties are given in Table I. The sphere is embedded in a two-layered background fluid medium, with the upper medium being water and the lower medium being sand (Table I). The scattered field is computed in the far field away from the object, over the frequency band 0.1–10 kHz. Two subcases are considered (Fig. 14): (a) sphere buried at 20 cm depth, and (b) half-buried sphere. The far-field receivers are located in the backscattering direction and in the forward scattering direction, at a distance of 50 m from the projection of the sphere centroid on the water-sediment interface plane. Solutions are computed for subcritical insonification (grazing angle $\theta_{\text{inc}}=15°$) and for supercritical insonification (grazing angle $\theta_{\text{inc}}=25°$).

Benchmark results for scattering from the buried and half-buried void spherical shells have already been presented in Ref. 45. Solving the problem for a filled spherical shell is a more challenging task, because of the increased structural complexity of the target. This narrows down the number of computational techniques, which can be employed. The results presented here compare the coupled FE with HK postprocessing[25,45] (labeled FE-HK in the plots) to solutions obtained by the wavenumber integration tool OASES (Refs. 39–42) (labeled OAS in the plots).

A. Buried sphere

The far-field TS for the buried sphere is plotted in Fig. 15. The OASES solution was computed with the single-scattering approximation of the scattering module, which computes the field scattered by the sphere included in an ideally infinite sand medium by a spherical harmonic superposition method. The far-field TS is successively computed via the HK integral, which takes into account the two-layered background medium by employing Green's functions obtained through the numerical integration of horizontal wavenumber spectral integrals.

The agreement between the FE-HK and the OAS result is overall very good, for both subcritical insonification [Fig. 15(a)] and supercritical insonification [Fig. 15(b)]. Some disagreement can be seen in the subcritical backscattering results, in the low TS region above 4 kHz. Similar discrepancies have been observed also in previous model comparison studies[45] for a buried spherical shell. Nevertheless, the trend in the low TS region above 4 kHz agrees well between the two models.

The overall good agreement between the FE-HK solution, which is a full scattering solution, and the OAS solution, which takes into account only single scattering, suggests that the interference between the scattered field and the water-sediment interface is negligible in this case.

B. Half-buried sphere

In this more complicated test problem, the multiple-scattering interactions between the spherical target and the

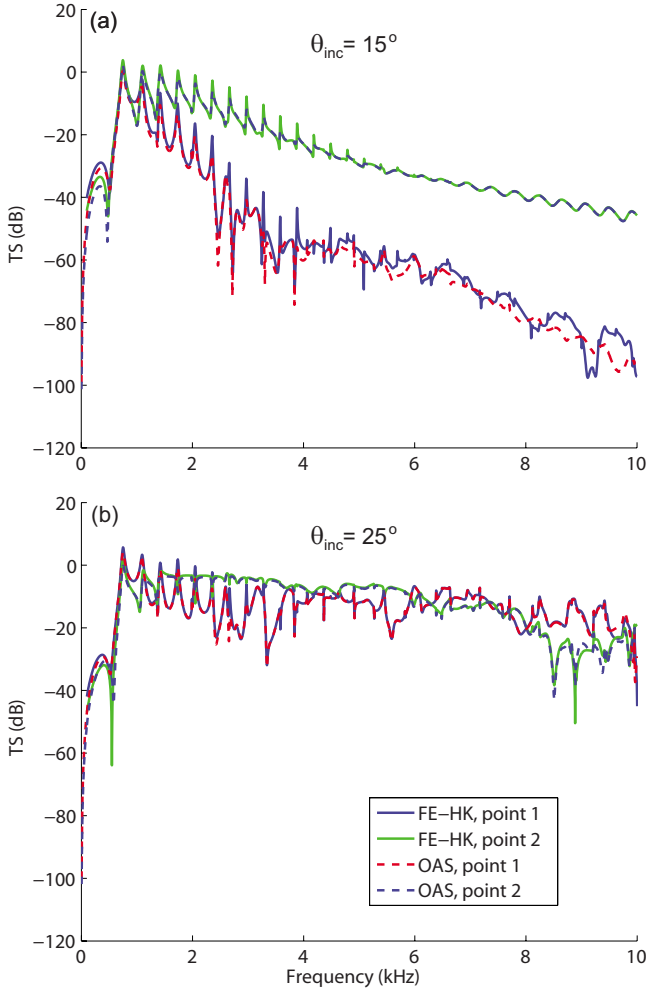


FIG. 15. (Color online) Scattering from the buried filled sphere at subcritical (a) and supercritical (b) incidences.

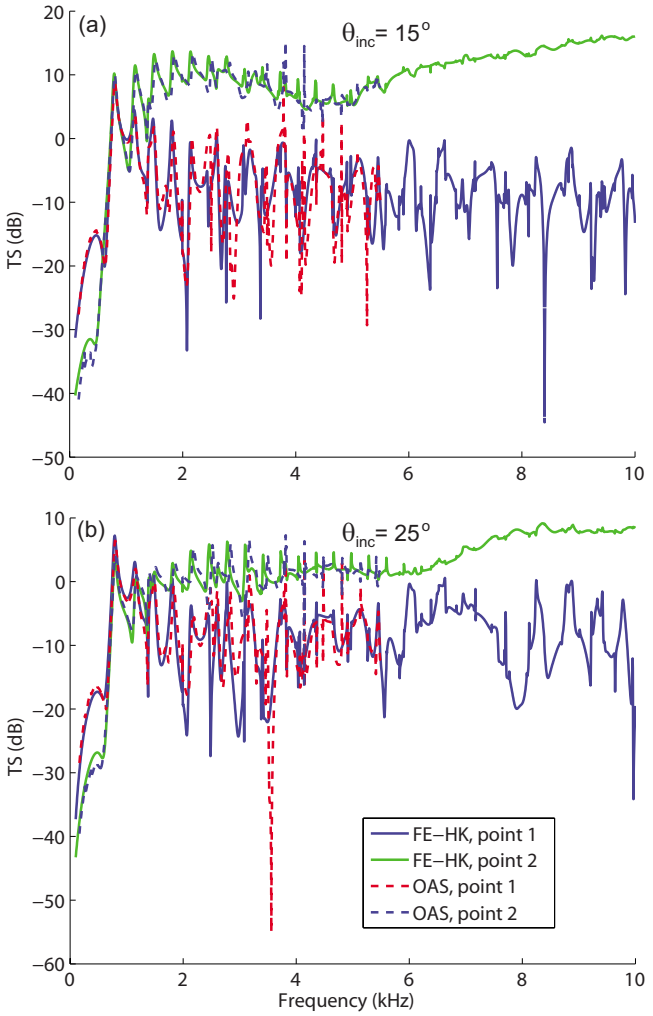


FIG. 16. (Color online) Scattering from the half-buried filled sphere at subcritical (a) and supercritical (b) incidences.

seafloor cannot be neglected. For this reason, it is necessary to compute the OAS solution using a nodal stiffness matrix describing the target surface impedance, in conjunction with the discrete representation of the HK integral for the layered medium. The full solution of the scattering problem with OASES requires the inversion of a dense matrix at each frequency, the size of which is equal to twice the number of surface points (pressure and pressure-gradient sources) required to characterize the target impedance. The computational burden of this approach has made it necessary to limit the maximum frequency of the OASES computations to 5.5 kHz. In Fig. 16 the FE-HK solution and the OASES solution are compared. The agreement in both backscattering and forward scattering is good at the low frequencies, with some differences, caused by the progressive degradation of the OASES stiffness matrix solution convergence, appearing as the frequency increases. For the lower part of the frequency band, the solutions provided can be regarded overall as benchmark solutions. The FE-HK solution was obtained by discretizations analogous to those used for the complete burial computations, which gives some degree of confidence in the accuracy of the FE-HK results throughout the entire frequency band.

VI. CONCLUSIONS

This paper gives an overview of a number of different numerical techniques currently employed in low-frequency [0.1–10 kHz for $O(1\ \mathrm{m})$-sized objects] target scatter modeling. Results are presented for a set of test problems comprising free-field scattering scenarios as well as scattering from targets that are proud, half-buried, or fully buried in the sediment.

Solutions to axisymmetric free-field problems, such as the submerged cylinder and shell of Sec. III, are easily obtained with high accuracy by the FE methods and by the BE method for the cases without the internal structure. The problem becomes computationally intensive when the cylinder is partially solid filled, in which case the addition of an internal structure in the target is modeled in a straightforward way by the FE methods. Interaction with an interface is a nontrivial problem: Only two solutions (NURC FE and the fully 3D FOI BIE) for a void cylindrical shell could be obtained, for which one is a single-scattering approximation resulting from the postprocessing of the free-field results computed by the axially symmetric NURC FE. Because of the presence of the internal structure, only the NURC FE single-scattering solution is available for the filled cylindrical shell proud on

the seabed. For the solid-filled spheres, the FE-HK model and the OASES model provide reference solutions for the buried and the half-buried configuration. The wavenumber integration solution coupled to the stiffness matrix approach for the half-buried sphere is limited to the lower part of the frequency band, for constraints imposed by the available computational resources. Nevertheless, the comparison at the lower frequencies together with the convergence of the FE solution give a basis of confidence for the reliability of the FE-HK solution also at higher frequencies in this case.

From this study, it is also apparent that full 3D modeling is computationally intensive and generally prohibitive without the availability of a multinode computing facility. Overall, the low-frequency solution quality is very good and reference solutions are identified for all but one of the seven test problems.

The accuracy of the solutions presented, and the capability of the FE method and of the BE method to model more complex shapes than the ones presented in this paper, make it possible to envision the application of such techniques to the analysis of targets commonly encountered during field work. The FE method has also the advantage of being able to deal in an immediate manner with discontinuous internal structures (such as the filler-void transition inside the cylindrical shell of Sec. III C), which makes the technique amenable for studying complex targets containing, for example, ribs, solid partitions, and flooded partitions. The study of such problems is of interest in connection with experimental work, for example, the study of a composite inhomogeneously filled cylinder on the seafloor presented in Ref. 53, or the detection of internal flaws in composite objects.[54]

ACKNOWLEDGMENTS

The authors are grateful to John B. Blottman of NUWC and to Ilkka Karasalo of FOI, Stockholm for providing, respectively, the SONAX and the BIE model results presented in this paper.

[1]H. A. Schenck, "Improved integral formulation for acoustic radiation problems," J. Acoust. Soc. Am. **44**, 41–58 (1968).
[2]J. J. Shirron and I. Babuska, "A comparison of approximate boundary conditions and infinite element methods for exterior Helmholtz problems," Comput. Methods Appl. Mech. Eng. **164**, 121–139 (1998).
[3]S. Amini and P. J. Harris, "A comparison between various boundary integral formulations of the exterior acoustic problem," Comput. Methods Appl. Mech. Eng. **84**, 59–75 (1990).
[4]I. Karasalo, "On evaluation of hypersingular integrals over smooth surfaces," Comput. Mech. **40**, 617–625 (2007).
[5]P. Waterman, "New formulation of acoustic scattering," J. Acoust. Soc. Am. **45**, 1417–1429 (1969).
[6]G. Kristensson and P. C. Waterman, "The T matrix for acoustic and electromagnetic scattering by circular disks," J. Acoust. Soc. Am. **72**, 1612–1625 (1982).
[7]V. K. Varadan, V. V. Varadan, J. H. Su, and T. Pillai, "Comparison of sound scattering by rigid and elastic obstacles in water," J. Acoust. Soc. Am. **71**, 1377–1383 (1982).
[8]J. A. Fawcett, "Modeling of high-frequency scattering from objects using a hybrid Kirchhoff/diffraction approach," J. Acoust. Soc. Am. **109**, 1312–1319 (2001).
[9]R. Jeans and I. C. Mathews, "The wave superposition method as a robust technique for computing acoustic fields," J. Acoust. Soc. Am. **92**, 1156–1166 (1992).
[10]D. T. Wilton, I. C. Mathews, and R. A. Jeans, "A clarification of nonexistence problems with the superposition method," J. Acoust. Soc. Am. **94**, 1676–1680 (1993).
[11]R. Jeans and I. Mathews, "A unique coupled boundary element/finite element method for the elastoacoustic analysis of fluid-filled thin shells," J. Acoust. Soc. Am. **94**, 3473–3479 (1993).
[12]D. B. Reeder and T. K. Stanton, "Acoustic scattering by axisymmetric finite-length bodies: An extension of a two-dimensional conformal mapping method," J. Acoust. Soc. Am. **116**, 729–746 (2004).
[13]D. B. Reeder, J. M. Jech, and T. K. Stanton, "Broadband acoustic backscatter and high-resolution morphology of fish: Measurement and modeling," J. Acoust. Soc. Am. **116**, 747–761 (2004).
[14]F. Ihlenburg, *Finite Element Analysis of Acoustic Scattering* (Springer-Verlag, New York, 1998).
[15]J. J. Shirron, "Solution of exterior Helmholtz problems using finite and infinite elements," Ph.D., thesis, University of Maryland, College Park, MD (1995).
[16]S. Dey, J. J. Shirron, and L. S. Couchman, "Mid-frequency structural acoustic and vibration analysis in arbitrary, curved three-dimensional domains," Comput. Struct. **79**, 617–629 (2001).
[17]D. K. Datta, S. Dey, and J. J. Shirron, *Scalable Three-Dimensional Acoustics Using hp-Finite/Infinite Elements and FETI-DP*, Springer Series: Lecture Notes in Computational Science and Engineering (Springer-Verlag, Berlin, 2005).
[18]J. J. Shirron and T. E. Giddings, "A finite element model for acoustic scattering from objects near a fluid-fluid interface," Comput. Methods Appl. Mech. Eng. **196**, 279–288 (2006).
[19]C. Farhat, A. Macedo, and M. Lesoinne, "A two-level domain decomposition method for the iterative solution of high frequency exterior Helmholtz problems," Numer. Math. **85**, 283–308 (2000).
[20]J. A. Fawcett, "Summary of object/waveguide scattering modeling at SACLANTCEN," Technical Report No. SR-298, SACLANTCEN, La Spezia, Italy, 1998.
[21]D. S. Burnett, "A three-dimensional acoustic infinite element based on a prolate spheroidal multipole expansion," J. Acoust. Soc. Am. **96**, 2798–2816 (1994).
[22]D. S. Burnett and R. L. Holford, "Prolate and oblate spheroidal acoustic infinite elements," Comput. Methods Appl. Mech. Eng. **158**, 117–141 (1998).
[23]D. S. Burnett and M. Zampolli, "Development of a finite-element, steady-state 3-D acoustics code for target scattering," Technical Report SR-379, NATO Undersea Research Center, La Spezia, Italy, 2003.
[24]D. S. Burnett and M. Zampolli, "FESTA: A 3-D finite element program for acoustic scattering from undersea targets," Technical Report SR-394, NATO Undersea Research Center, La Spezia, Italy, 2004.
[25]M. Zampolli, A. Tesei, F. B. Jensen, N. Malm, and J. B. Blottman, "A computationally efficient finite element model with perfectly matched layers applied to scattering from axially symmetric objects," J. Acoust. Soc. Am. **122**, 1472–1485 (2007).
[26]L. L. Thompson, "A review of finite-element methods for time-harmonic acoustics," J. Acoust. Soc. Am. **119**, 1315–1330 (2006).
[27]S. Marburg and B. Nolte, *Computational Acoustics of Noise Propagation in Fluids, Finite and Boundary Eelement Methods* (Springer-Verlag, Heidelberg, 2008).
[28]F. Ingenito, "Scattering from an object in a stratified medium," J. Acoust. Soc. Am. **82**, 2051–2059 (1987).
[29]T. E. Giddings and J. J. Shirron, "A model for sonar interrogation of complex bottom and surface targets in shallow-water waveguides," J. Acoust. Soc. Am. **123**, 2024–2034 (2008).
[30]G. C. Gaunaurd and H. Huang, "Acoustic scattering by a spherical body near a plane boundary," J. Acoust. Soc. Am. **96**, 2526–2536 (1994).
[31]H. Huang and G. C. Gaunaurd, "Acoustic point source scattering by a spherical elastic shell submerged beneath a free surface," J. Acoust. Soc. Am. **99**, 2720–2726 (1996).
[32]R. Lim, J. L. Lopes, R. H. Hackman, and D. G. Todoroff, "Scattering by objects buried in underwater sediments: Theory and experiment," J. Acoust. Soc. Am. **93**, 1762–1783 (1993).
[33]R. Lim, "Acoustic scattering by a partially buried three-dimensional elastic obstacle," J. Acoust. Soc. Am. **104**, 769–782 (1998).
[34]H. J. Simpson, B. H. Houston, and R. Lim, "Laboratory measurements of sound scattering from a buried sphere above and below the critical angle (L)," J. Acoust. Soc. Am. **113**, 39–42 (2003).
[35]G. C. Bishop and J. Smith, "Scattering from rigid and soft targets near a planar boundary: Numerical results," J. Acoust. Soc. Am. **105**, 130–143 (1999).
[36]G. C. Bishop and J. Smith, "Scattering from an elastic shell and a rough

fluid-elastic interface: Theory," J. Acoust. Soc. Am. **101**, 767–788 (1997).
[37]R. Lim, K. L. Williams, and E. I. Thorsos, "Acoustic scattering by a three-dimensional elastic object near a rough surface," J. Acoust. Soc. Am. **107**, 1246–1262 (2000).
[38]A. Sarkissian, "Method of superposition applied to scattering from a target in shallow water," J. Acoust. Soc. Am. **95**, 2340–2345 (1994).
[39]H. Schmidt and J. Glattetre, "A fast field model for the three-dimensional wave propagation in stratified environments based on the global matrix method," J. Acoust. Soc. Am. **78**, 2105–2114 (1985).
[40]A. Tesei, A. Maguer, W. L. J. Fox, R. Lim, and H. Schmidt, "Measurements and modeling of acoustic scattering from partially and completely buried spherical shells," J. Acoust. Soc. Am. **112**, 1817–1830 (2002).
[41]H. Schmidt, "Virtual source approach to scattering from partially buried elastic targets," AIP Conf. Proc. **728**, 456–463 (2004).
[42]I. Lucifredi and H. Schmidt, "Subcritical scattering from buried elastic shells," J. Acoust. Soc. Am. **120**, 3566–3583 (2006).
[43]S. Ivansson and I. Karasalo, "Computation of modal wavenumbers using an adaptive winding-number integral method with error control," J. Sound Vib. **161**, 173–180 (1993).
[44]I. Karasalo and J. Mattsson, "Accurate numerical modelling of scattering by 3D bodies and shells in a fluid-solid medium," in *Proceedings of Fourth European Conference on Underwater Acoustics*, edited by A. Alippi and G. B. Cannelli (CNR-IDAC, Rome, Italy, 1998), pp. 691–696.
[45]M. Zampolli, A. Tesei, G. Canepa, and O. A. Godin, "Computing the far field scattered or radiated by objects inside layered fluid media using approximate Greens functions," J. Acoust. Soc. Am. **123**, 4051–4058 (2008).
[46]F. B. Jensen, M. Zampolli, and A. Tesei, "Benchmarking scattering from spheres and cylinders near the seafloor: A numerical comparative study," Technical Report No. NURC-FR-2007-005, NATO Undersea Research Center, La Spezia, Italy, 2007.
[47]R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (Peninsular, Los Altos, CA, 1983).
[48]COMSOL AB, COMSOL MULTIPHYSICS 3.2—User's guide, 2005; http://www.comsol.com (Last viewed 6/6/2008).
[49]J. P. Bérenger, "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200 (1994).
[50]J. P. Bérenger, "Perfectly matched layer for the FDTD solution of wave-structure interaction problems," IEEE Trans. Antennas Propag. **44**, 110–117 (1996).
[51]I. Karasalo and K. Otto, "Acoustic scattering from submerged elastic objects," Technical Report No. FOI-R-0307-SE, FOI, Swedish Defence Research Agency, Stockholm, 2001.
[52]I. Karasalo, "Exact finite elements for wave propagation in range-independent fluid-solid media," J. Sound Vib. **172**, 671–688 (1994).
[53]A. Tesei, M. Zampolli, and G. Canepa, "At sea measurements of acoustic-elastic scattering by a 1.5 m long cylinder made of composite materials," in Proceedings of Second International Conference on Underwater Acoustic Measurements (2007), pp. 505–510.
[54]A. Tesei, P. Guerrini, and M. Zampolli, "Tank measurements of scattering from a resin-filled fiberglass spherical shell with internal flaws," J. Acoust. Soc. Am. **124**, 827–840 (2008).

Coherence function of a sound field in an oceanic waveguide with horizontally isotropic statistics

A. G. Voronovich[a] and V. E. Ostashev
NOAA/Earth System Research Laboratory, 325 Broadway, Boulder, Colorado 80303

(Received 5 November 2007; revised 31 October 2008; accepted 31 October 2008)

The mean value and the coherence function of a sound field propagating in an oceanic waveguide with random inhomogeneities are important statistical characteristics of this field, which are needed for many practical applications. Closed equations for the coherence function were obtained in many works for both two dimensional and three dimensional geometries. For the 3D case, these equations are too involved even for a numerical treatment. In this paper, explicit expressions for the mean field and the coherence function due to a point omnidirectional monochromatic source in a 3D waveguide are derived for the case of random inhomogeneities, which are statistically isotropic in a horizontal plane. The solutions are much simpler than those obtained previously due to the cylindrical symmetry of the problem. The theory developed is used to study numerically the mean field and the coherence function in an oceanic waveguide perturbed by a random field of internal waves with the Garrett–Munk spectrum. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3035832]

PACS number(s): 43.30.Re, 43.30.Bp, 43.30.Qd [JAC] Pages: 99–110

I. INTRODUCTION

Coherence of low-frequency sound waves propagating over long ranges in the ocean diminishes due to sound scattering by random internal waves (IWs) and spice. Important statistical characteristics of a sound field are the mean value and the coherence function. These first two statistical moments of a sound field have been studied in many works using different approaches, e.g., Refs. 1–8. For low-frequency, long-range sound propagation in a fluctuating ocean, a modal representation of a sound field seems to be the most adequate. This representation was used in studies of sound propagation through a fluctuating ocean both for a two dimensional (2D) geometry (e.g., see Refs. 6 and 9) and for a three dimensional (3D) geometry.[5,7,8,10,11] In the latter papers, closed equations for the mean sound field and the coherence function in a 3D fluctuating ocean were derived using different techniques. The mean sound field can be calculated relatively easy with these equations. However, most of the equations for the coherence function are too involved to be solved numerically.

For example, in Ref. 5, the Chernov method[12] was used to calculate the first two statistical moments of a monochromatic sound field. The equations derived for the coherence function allowed us to obtain some numerical results, e.g., the depth-averaged coherence function. However, a dimension of the set of these equations is still too large for a detailed numerical analysis of the dependence of the coherence function on the parameters of the problem.

The goal of the present paper is to simplify the theory developed in Ref. 5 and to make it readily amenable for numerical studies of the coherence function of a low-frequency sound wave propagating in a fluctuating ocean. To this end, we will assume that random inhomogeneities in the ocean are statistically homogeneous and isotropic in a horizontal plane and that a sound source is omnidirectional. Both these assumptions are valid with good accuracy for many problems of low-frequency, long-range sound propagation. With these assumptions, it is worthwhile to use a cylindrical coordinate system and to decompose the sound field into a sum of acoustic modes and azimuthal harmonics similar to the approach of Ref. 7. Then, the Chernov method is used to obtain closed equations for the first and second moments of the mode amplitudes. As a result, closed equations for the mean field and the coherence function of a sound wave propagating in a random ocean are derived. The solution of the equation for the mean field agrees with those known in the literature. The dimension of a scattering matrix appearing in the solution of the equation for the coherence function is much smaller than those in the previous works due to the symmetry of the problem used in the derivation. This makes the theory developed readily amenable for numerical calculations of the coherence function. Examples of such calculations are presented.

The paper is organized as follows. In Sec. II, starting equations for the considered problem are presented. The mean field and the coherence function are calculated in Secs. III and IV, respectively. In Sec. V, the theory developed is used for numerical studies of the coherence function of a sound field propagating in an oceanic waveguide with a random field of IW. Finally, Sec. VI summarizes the results obtained in the paper.

II. PROBLEM FORMULATION

A. Starting equations

Let a point omnidirectional monochromatic source be located at $z=z_0$ and $\boldsymbol{r}=0$ in an oceanic waveguide with random inhomogeneities (see Fig. 1). Here, z and $\boldsymbol{r}=(x,y)$ are vertical and horizontal coordinates. The ocean surface and the upper surface of the ocean bottom are located at $z=0$ and

[a]Electronic mail: alexander.voronovich@noaa.gov

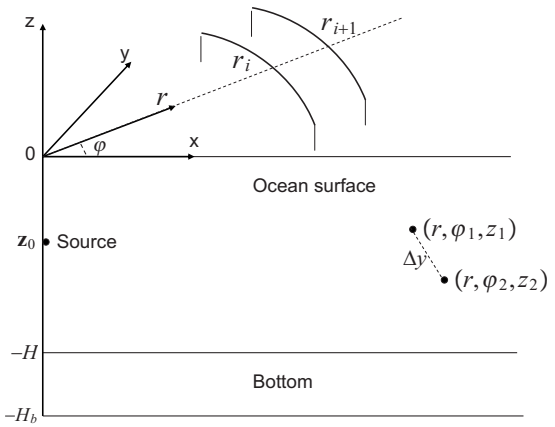


FIG. 1. The geometry of sound propagation in the ocean. Two receivers are located at $(r,\varphi_1;z_1)$ and $(r,\varphi_2;z_2)$. The horizontal distance between them is $\Delta y=r(\varphi_1-\varphi_2)$. The $(i+1)$ cylindrical slab $r_i<r<r_{i+1}$ is also shown.

$z=-H$, respectively. The ocean bottom is modeled as a stratified fluid located within the layer $-H_b\leqslant z\leqslant -H$, which lies on an acoustically hard (basalt) plane surface located at $z=-H_b$. Within the water column (i.e., at $-H\leqslant z\leqslant 0$), the sound pressure $p(\boldsymbol{r},z)$ due to the omnidirectional point source satisfies the Helmholtz equation

$$\left(\frac{\partial^2}{\partial z^2}+\nabla_{\boldsymbol{r}}^2+\frac{\omega^2}{c^2(\boldsymbol{r},z)}\right)p=\delta(z-z_0)\,\delta(\boldsymbol{r}), \tag{1}$$

where ω is the sound frequency, δ is the delta function, and $c(\boldsymbol{r},z)$ is the sound speed, which includes both a regular sound-speed profile in a stratified ocean and random fluctuations due to, e.g., IW.

In Eq. (1), we set $\omega^2/c^2(\boldsymbol{r},z)=\omega^2/c_0^2(z)-\mu(\boldsymbol{r},z)$. Here, $1/c_0^2(z)=\langle 1/c^2(\boldsymbol{r},z)\rangle$, where the brackets $\langle\ \rangle$ denote averaging over an ensemble of realizations of sound-speed fluctuations, and μ is a random function given by

$$\mu(\boldsymbol{r},z)=\frac{\omega^2}{c_0^2(\omega)}-\frac{\omega^2}{c^2(\boldsymbol{r},z)}. \tag{2}$$

The random function μ is assumed to be statistically homogeneous and isotropic in a horizontal plane; note that $\langle\mu\rangle=0$.

For low-frequency, long-range sound propagation, it is worthwhile to decompose the sound field p into a sum of acoustic modes $u_n(z)$ corresponding to a stratified ocean without fluctuations ($\mu=0$) and a stratified fluid bottom. Within the layer $-H_b\leqslant z\leqslant 0$, these modes satisfy the equation[13]

$$\left[\frac{d}{dz}\left(\frac{1}{\varrho_0(z)}\frac{d}{dz}\right)+\frac{1}{\varrho_0(z)}\left(\frac{\omega^2}{c_0^2(z)}-\xi_n^2\right)\right]u_n=0 \tag{3}$$

and boundary conditions at the ocean surface and hard bottom: $u_n(0)=0$ and $(\partial u_n/\partial z)_{z=-H_b}=0$. Here, ϱ_0 is the dimensionless density, which is equal to 1 within the water column but can depend on z within the fluid bottom, ξ_n are propagation constants, and $n=1,2,\ldots$. In the absence of sound absorption in the bottom, the acoustic modes u_n and squares of the horizontal wave numbers ξ_n^2 are real. Note that a generalization of the theory developed below to account for sound absorption (which results in complex u_n and ξ_n) is straightforward. However, for simplicity and to be able to check energy conservation, it is assumed that sound absorption in the bottom is absent. The set of the acoustic modes is complete and orthonormal,

$$\sum_n u_n(z)u_n(z_0)=\varrho_0(z_0)\,\delta(z-z_0),\qquad \int_{-H_b}^{0}\frac{u_n(z)u_m(z)}{\varrho_0(z)}dz=\delta_{nm}. \tag{4}$$

We will consider sound propagation in the forward scattering approximation, according to which the sound field p satisfies the following integral equation:

$$p(\boldsymbol{r},z)=G_0(r;z,z_0)+\int_{-H}^{0}dz'\int_{0\leqslant r'\leqslant r}G_0(|\boldsymbol{r}-\boldsymbol{r}'|;z,z')\times\mu(\boldsymbol{r}',z')p(\boldsymbol{r}',z')d\boldsymbol{r}'. \tag{5}$$

Here, $r=|\boldsymbol{r}|$ and G_0 is the Green's function of Eq. (1) in a stratified ocean without random inhomogeneities [i.e., when $c=c_0(z)$],

$$G_0(|\boldsymbol{r}-\boldsymbol{r}'|;z,z')=-\frac{i}{4}\sum_n u_n(z)u_n(z')H_0^{(1)}(\xi_n|\boldsymbol{r}-\boldsymbol{r}'|), \tag{6}$$

where $H_0^{(1)}$ is the Hankel function. Note that since in the fluid bottom $\mu=0$, in Eq. (5) the integration over z' is, in fact, limited to the water column.

The Helmholtz equation [Eq. (1)] is equivalent to Eq. (5) if the upper limit of integration over r' in Eq. (5) is set to infinity. By setting this limit to r, the effect of the random inhomogeneities $\mu(\boldsymbol{r},z')$ with $r'>r$ on $p(\boldsymbol{r},z)$ is ignored; i.e., backscattering of sound waves is neglected. Equation (5) is similar to Eq. (7) in Ref. 14 where the case of a sound wave incident on an inhomogeneous medium located at $x>0$ is considered. Note that since the backscattering is ignored, the sound wavelength λ must be significantly smaller than the characteristic horizontal scale of sound-speed fluctuations.

The goal of this paper is to calculate the first two statistical moments of the sound field. In all previous works known to us, except for Ref. 7, these statistical moments were studied in the Cartesian coordinate system. It will be shown below that calculation of the coherence function is significantly simpler in a cylindrical coordinate system r,φ,z (see Fig. 1) when one can take advantage of the rotational symmetry of the problem. Note that in the cylindrical coordinate system, $\boldsymbol{r}=(r\cos\varphi,r\sin\varphi)$. In this system, it is worthwhile to decompose the sound field into the sum of both acoustic modes and azimuthal harmonics,[7]

$$p(r,\varphi,z)=\sum_n\sum_{q=-\infty}^{\infty}a_{nq}(r)u_n(z)H_q^{(1)}(\xi_n r)e^{iq\varphi}, \tag{7}$$

where $a_{nq}(r)$ are the mode amplitudes. Equations for $a_{nq}(r)$ are presented in the next subsection.

B. Equations for mode amplitudes

Let the propagation range be divided into a number of cylindrical slabs $(0,r_1),(r_1,r_2),\ldots,(r_{L-1},r_L=r)$, where L is

the number of slabs. The $(i+1)$ slab is shown in Fig. 1. In the first slab, the sound field is calculated using Eq. (5), where $r=r_1$. This allows us to determine the sound field $p(r_1,\varphi,z)$ at the internal surface (beginning) of the second cylindrical slab. In the second slab, the sound field is also calculated using the forward scattering approximation, which allows us to determine the sound field at the internal surface of the third cylindrical slab, $p(r_2,\varphi,z)$, and so on. Inside the $(i+1)$ slab, where $i=0,1,\ldots(L-1)$, the sound field $p(r,\varphi,z)$ satisfies the following integral equation:

$$p(\boldsymbol{r},z)=p_i(\boldsymbol{r},z)+\int_{-H}^{0}dz'\int_{r_i<r'<r}G_0(|\boldsymbol{r}-\boldsymbol{r}'|;z,z')\times\mu(\boldsymbol{r},z')p(\boldsymbol{r},z')d\boldsymbol{r}'. \tag{8}$$

Here, $r_i<r<r_{i+1}$ and $p_i(\boldsymbol{r},z)$ is the sound field inside the slab without random inhomogeneities, which at $r=r_i$ coincides with the sound field at the internal surface of the ith cylindrical slab: $p_i(r_i,\varphi,z)=p(r_i,\varphi,z)$. Equation (8) differs from Eq. (5) only by limits of integration over r' (which correspond to a particular slab) and by the first term on the right-hand side, which corresponds to the field at the beginning of a slab. In the first slab, where $i=0$, Eq. (8) coincides with Eq. (5), provided that $p_0(\boldsymbol{r},z)=G_0(\boldsymbol{r};z,z_0)$ and $r_0=0$.

Similar to Eq. (7), the sound field $p_i(\boldsymbol{r},z)$ is decomposed into a sum of acoustic modes and azimuthal harmonics,

$$p_i(r,\varphi,z)=\sum_n\sum_{q=-\infty}^{\infty}a_{nq}(r_i)u_n(z)H_q^{(1)}(\xi_n r)e^{iq\varphi}. \tag{9}$$

Note that in this equation, the mode amplitudes $a_{nq}(r_i)$ do not depend on r. In Eq. (6), the Hankel function is also decomposed into the sum of azimuthal harmonics,

$$H_0^{(1)}(\xi|\boldsymbol{r}-\boldsymbol{r}'|)=\sum_{q=-\infty}^{\infty}J_q(\xi r')H_q^{(1)}(\xi r)e^{iq(\varphi-\varphi')}, \tag{10}$$

where J_q is the Bessel function. Substituting Eqs. (7), (9), and (10) into Eq. (8), we obtain an integral equation for the mode amplitudes within the $(i+1)$ cylindrical slab,

$$a_{nq}(r)=a_{nq}(r_i)+i\sum_{n',q'}\int_{r_i}^{r}V_{nq}^{n'q'}(r')a_{n'q'}(r')dr'. \tag{11}$$

Here, $V_{nq}^{n'q'}$ are given by

$$V_{nq}^{n'q'}(r')=-\frac{1}{4}\int_0^{2\pi}d\varphi' r' J_q(\xi_n r')H_{q'}^{(1)}(\xi_{n'}r')e^{-i(q-q')\varphi'}\times\int_{-H}^{0}u_n(z')\mu(\boldsymbol{r}',z')u_{n'}(z')dz'. \tag{12}$$

In this formula, the Bessel function can be written as $J_q(\xi_n r')=\frac{1}{2}H_q^{(1)}(\xi_n r')+\frac{1}{2}(H_q^{(1)}(\xi_n r'))^*$. Substituting this formula into Eq. (12), we express $V_{nq}^{n'q'}$ as a sum of two terms, the first of which contains the product $H_q^{(1)}(\xi_n r')H_{q'}^{(1)}(\xi_{n'}r')$. If the arguments of both Hankel functions are large, this product is a rapidly oscillating function of r', which after integration with respect to r' in Eq. (11) gives a negligible contribution to a_{nq}. (Here, we take into account that $\mu(\boldsymbol{r},z')$ varies slowly on the scale of a sound wavelength.) The second term contains the product $H_q^{(1)}(\xi_n r')(H_{q'}^{(1)}(\xi_{n'}r'))^*$. In this product, we replace the Hankel functions with their asymptotics,

$$H_q^{(1)}(x)\approx\sqrt{\frac{2}{\pi x}}\exp\left(ix-\frac{i\pi q}{2}-\frac{i\pi}{4}\right), \tag{13}$$

which are valid for $q\ll x$.

The lowest order correction to the asymptotics [Eq. (13)] would be given by the term $iq^2/2x$ in the exponential in this equation. (This correction term easily follows from Eq. (8.453.3) in Ref. 15.) For small φ, one has $\exp(iq\varphi)=\exp(iq\Delta y/r)$, where Δy is a cross-range distance. Therefore, for typical values of q one finds $q\sim r/\rho_{\mathrm{corr}}$, where ρ_{corr} is a correlation radius of the sound field. Thus, in Eq. (13) the correction term can be neglected if the following condition is fulfilled:

$$\rho_{\mathrm{corr}}\gg\sqrt{r/k_0}\sim\rho_F, \tag{14}$$

where k_0 is a reference sound wave number and ρ_F is the radius of the first Fresnel zone. Since ρ_{corr} diminishes with distance (usually as $r^{-1/2}$), the condition given by Eq. (14) breaks down for sufficiently large r.

Note that a rather restrictive inequality given by Eq. (14) is a *sufficient* condition of applicability of the results obtained in the paper. *Necessary* conditions are probably much less restrictive. For example, in Sec. III C the decrements of attenuation of acoustic modes calculated without the correction term $iq^2/2x$ are compared with those calculated using the diagram technique, which is a more general approach than that in the present paper. It is shown that the results coincide regardless of the condition given by Eq. (14). Furthermore, when considering the coherence function, the correction term $iq^2/2x$ corresponding to one point of observation should be subtracted from that corresponding to another point. As a result, these two correction terms are partially canceled. A rigorous analysis of such cancellation is rather involved and not presented in this paper.

Substituting Eq. (13) into Eq. (12), we have

$$V_{nq}^{n'q'}(r')=-\frac{1}{4\pi}\frac{e^{i(\xi_{n'}-\xi_n)r'}}{\sqrt{\xi_n\xi_{n'}}}\int_0^{2\pi}d\varphi' e^{i(q'-q)(\varphi'-\pi/2)}\times\int_{-H}^{0}u_{n'}(z')\mu(\boldsymbol{r}',z')u_n(z')dz'. \tag{15}$$

It follows from this equation that the matrix V is Hermitian

$$V_{nq}^{n'q'}=(V_{n'q'}^{nq})^*. \tag{16}$$

[Note that the Hermitian property, Eq. (16), holds only in the absence of sound absorption in the bottom when u_n and ξ_n are real.]

We now assume that the sound-speed fluctuations are small and, following the Chernov method,[12] iterate Eq. (11) twice,

$$a_{nq}(r) = a_{nq}(r_i) + i\sum_{n',q'}\int_{r_i}^{r} dr' V_{nq}^{n'q'}(r') a_{n'q'}(r_i) - \sum_{n',q',n'',q''}\int_{r_i}^{r} dr' V_{nq}^{n'q'}(r')\int_{r_i}^{r'} dr'' V_{n'q'}^{n''q''}(r'') \times a_{n''q''}(r_i) + O(\mu^3). \quad (17)$$

Note that to obtain nontrivial results after averaging, one has to include terms that contain products of two μ-terms. To include all such terms, Eq. (11) was iterated twice. The third iteration of this equation is not needed if scattering in a cylindrical slab with the width of order of the correlation radius l_0 of the sound-speed fluctuations in the horizontal plane is relatively weak. [Equation (11) could be iterated an infinite number of times, resulting in a diagram technique,[16] which is not considered in this paper.] Equation (17) for the mode amplitudes $a_{nq}(r)$ will be a starting point in derivation of closed equations for the mean field and the coherence function.

III. MEAN FIELD

In this section, the mean sound field $\bar{p}$ is calculated. To this end, both sides of Eq. (7) are averaged over an ensemble of realizations of μ. The right-hand side of the resulting formula for $\bar{p}$ contains the mean mode amplitudes $\bar{a}_{nq}$, which are calculated in Sec. III A using the Chernov method.

A. The Chernov method

Averaging both sides of Eq. (17), we obtain

$$\bar{a}_{nq}(r) = \bar{a}_{nq}(r_i) + i\sum_{n',q'}\int_{r_i}^{r} dr' \langle V_{nq}^{n'q'}(r') a_{n'q'}(r_i)\rangle - \sum_{n',q',n'',q''}\int_{r_i}^{r} dr' \int_{r_i}^{r'} dr'' \langle V_{nq}^{n'q'}(r') V_{n'q'}^{n''q''}(r'') \times a_{n''q''}(r_i)\rangle + O(\mu^3). \quad (18)$$

The matrices $a(r_i)$ and $V(r')$ correlate only within a narrow cylindrical slab located at $r=r_i$ with the width of the order of l_0. Assuming that the width of the $(i+1)$th slab significantly exceeds l_0, in Eq. (18) the correlation between $a(r_i)$ and $V(r')$ can be neglected. In this approximation, correlators containing both $V(\mathbf{r}')$, $V(\mathbf{r}'')$, etc., and $a(r_i)$ are split into a product of two correlators, one of which contains $V(\mathbf{r}')$, $V(\mathbf{r}'')$, etc., and the other contains the mode amplitudes $a(r_i)$. This approximation is the cornerstone of the Chernov method (e.g., see Refs. 12 and 5), and it is similar to approximations used in other approaches associated with the Markov approximation.[16] In this approximation, the second term on the right-hand side of Eq. (18) vanishes since $\langle V\rangle=0$. As a result, we obtain the following formula for the mean mode amplitudes $\bar{a}_{nq}(r)$:

$$\bar{a}_{nq}(r) = \bar{a}_{nq}(r_i) - \sum_{n',q',n'',q''}\int_{r_i}^{r} dr' \int_{r_i}^{r'} dr'' \langle V_{nq}^{n'q'}(r') \times V_{n'q'}^{n''q''}(r'')\rangle \bar{a}_{n''q''}(r_i). \quad (19)$$

Here, the terms of order $O(\mu^3)$ are neglected.

Equation (19) contains the sum

$$\mathcal{V} = \sum_{q'} \langle V_{nq}^{n'q'}(r') V_{n'q'}^{n''q''}(r'')\rangle, \quad (20)$$

which is calculated in the rest of this subsection. Having in mind the calculation of the coherence function in Sec. IV, let us consider a more general V-correlator than that in Eq. (20),

$$\langle V_{nq}^{n'q'}(r') V_{m'q'}^{mq''}(r'')\rangle = \frac{1}{16\pi^2}\frac{\exp[i(\xi_{n'}-\xi_n)r' + i(\xi_m-\xi_{m'})r'']}{\sqrt{\xi_n\xi_{n'}\xi_{m'}\xi_m}} \times \int_0^{2\pi} d\varphi' d\varphi'' e^{i(q'-q)(\varphi'-\pi/2)+i(q''-q')(\varphi''-\pi/2)} \times \int_{-H}^{0} u_n(z')u_{n'}(z')u_{m'}(z'')u_m(z'') \times B_\mu(|\mathbf{r}'-\mathbf{r}''|;z',z'')dz', \quad (21)$$

where, to simplify notations, the index n'' is replaced with m. This correlator coincides with that in Eq. (20) if $m'=n'$. In Eq. (21),

$$B_\mu(|\mathbf{r}'-\mathbf{r}''|;z',z'') = \langle \mu(\mathbf{r}',z')\mu(\mathbf{r}'',z'')\rangle \quad (22)$$

is a horizontally isotropic correlation function of random inhomogeneities. Using the Fourier–Bessel transform, the correlation function is expressed as follows:

$$B_\mu(|\mathbf{r}'-\mathbf{r}''|;z',z'') = \int_0^\infty J_0(k|\mathbf{r}'-\mathbf{r}''|)\hat{B}_\mu(k;z',z'')kdk. \quad (23)$$

Here, k is the wave number, and the horizontal spectrum of random inhomogeneities is given by

$$\hat{B}_\mu(k;z',z'') = \int_0^\infty J_0(kr)B_\mu(r;z',z'')rdr. \quad (24)$$

Since $|\mathbf{r}'-\mathbf{r}''| = \sqrt{r'^2+r''^2-2r'r''\cos(\varphi'-\varphi'')}$, in Eq. (21) it is worthwhile to replace the integrals over φ' and φ'' with the integrals over φ'' and $\theta=\varphi'-\varphi''$. Then, the integration over φ'' results in $2\pi\delta_{qq''}$ (where $\delta_{qq''}$ is the Kronecker symbol), and Eq. (21) simplifies

$$\langle V_{nq}^{n'q'}(r') V_{m'q'}^{mq''}(r'')\rangle = \frac{\delta_{qq''}}{8\pi}\frac{\exp[i(\xi_{n'}-\xi_n)r' + i(\xi_m-\xi_{m'})r'']}{\sqrt{\xi_n\xi_{n'}\xi_{m'}\xi_m}} \times \int_0^{2\pi} d\theta e^{i(q'-q)\theta}\int_0^\infty J_0(k|\mathbf{r}'-\mathbf{r}''|) \times \Phi_{nn',mm'}(k)kdk, \quad (25)$$

where

$$\Phi_{nn',mm'}(k) = \int_{-H}^{0} u_n(z')u_{n'}(z')u_m(z'')u_{m'}(z'') \times \hat{B}_\mu(k;z',z'')dz'dz''. \quad (26)$$

Obviously, the matrix Φ is symmetric with respect to the permutation of indices within the first and the second groups as well as the permutations of the group themselves,

$$\Phi_{nn',mm'} = \Phi_{n'n,mm'} = \Phi_{nn',m'm} = \Phi_{mm',nn'}. \quad (27)$$

The V-correlator appearing in Eq. (20) is obtained by setting $q''=q$ and $m'=n'$ in Eq. (25). The resulting formula is substituted into Eq. (20). In that equation, the sum over the index q' is calculated using the Fourier series for the delta function,

$$\sum_{q'=-\infty}^{\infty} e^{iq'\theta} = 2\pi\delta(\theta), \quad -\pi < \theta < \pi. \quad (28)$$

Then, the integral over θ can also be calculated. As a result, we obtain the following formula for the sum $\mathcal{V}$:

$$\mathcal{V} = \sum_{q'} \langle V_{nq}^{n'q'}(r')V_{n'q'}^{mq}(r'')\rangle = \frac{\exp[i(\xi_{n'}-\xi_n)r' + i(\xi_m-\xi_{n'})r'']}{4\sqrt{\xi_n\xi_m}\xi_{n'}} \times \int_0^\infty J_0(k(r'-r''))\Phi_{nn',m'm}(k)kdk. \quad (29)$$

Note, that $\mathcal{V}$ does not, in fact, depend on the index q.

B. Formula for the mean field

Since the sound source is assumed to be isotropic and $\mathcal{V}$ does not depend on q, the solution of Eq. (19) has the following form:

$$\bar{a}_{nq}(r) = \bar{a}_n(r)\delta_{q0}. \quad (30)$$

Here, $\bar{a}_n(r)$ are the mean mode amplitudes, which do not depend on the azimuthal index q. The mean sound field $\bar{p}$ is expressed in terms of $\bar{a}_n$ via the following formula:

$$\bar{p}(r,z) = \sum_n \bar{a}_n(r)u_n(z)H_0^{(1)}(\xi_n r). \quad (31)$$

This formula can be obtained by averaging both sides of Eq. (7) and using Eq. (30) in the resulting equation.

Substituting Eq. (30) into Eq. (19), the following equation for the $\bar{a}_n(r)$ is obtained:

$$\bar{a}_n(r) = \bar{a}_n(r_i) - \sum_m S_{nm}(r_i,r)\bar{a}_m(r_i). \quad (32)$$

Here, the matrix S is given by

$$S_{nm}(r_i,r) = \int_{r_i}^{r} dr' \int_{r_i}^{r'} dr'' \sum_{n',q'} \langle V_{nq}^{n'q'}(r')V_{n'q'}^{mq}(r'')\rangle. \quad (33)$$

Substituting the value of the sum $\mathcal{V}$ from Eq. (29) into Eq. (33), we have

$$S_{nm}(r_i,r) = \frac{1}{4\sqrt{\xi_n\xi_m}}\sum_{n'} \int_{r_i}^{r} dr' \int_{r_i}^{r'} dr'' e^{i(\xi_{n'}-\xi_n)r' + i(\xi_m-\xi_{n'})r''} \times \int_0^\infty J_0(k(r'-r''))\Phi_{nn',n'm}(k)kdk. \quad (34)$$

The expression for S_{nm} can be simplified by introducing new integration variables $\rho=(r'+r'')/2$ and $\eta=r'-r''$,

$$S_{nm}(r,r_i) = s_{nm}\int_{r_i}^{r} e^{-i\rho b_{nm}}d\rho. \quad (35)$$

Here, the matrix s_{nm} does not depend on r and is given by

$$s_{nm} = \frac{1}{4\sqrt{\xi_n\xi_m}}\sum_{n'}\frac{1}{\xi_{n'}}\int_0^\infty dkk\Phi_{mn',n'n}(k) \times \int_0^\infty d\eta e^{i\eta b_{nmn'}}J_0(k\eta). \quad (36)$$

Furthermore, the coefficients $b_{nmn'}$ and b_{nm} are given by

$$b_{nm} = \xi_n - \xi_m, \quad b_{nmn'} = \frac{\xi_n+\xi_m}{2} - \xi_{n'}. \quad (37)$$

Strictly speaking, in Eq. (36) the upper limit of integration over η is $r-r_i-|r+r_i-2\rho|$. However, since $(r-r_i)\gg l_0$, for characteristic values of $k\sim l_0^{-1}$ the argument of the Bessel function in Eq. (36) is generally large. Since the integral over η converges, the upper limit of integration can be extended to infinity. Note that such an extension of the limits of integration is always used in theories of waves in random media (e.g., see Ref. 16).

Using Eq. (6.671.1-2) from Ref. 15, the integral over η in Eq. (36) is calculated explicitly,

$$\int_0^\infty d\eta e^{i\eta b_{nmn'}}J_0(k\eta) = \begin{cases} (k^2-b_{nmn'}^2)^{-1/2}, & k > |b_{nmn'}| \\ -i\,\mathrm{sgn}(b_{nmn'})(b_{nmn'}^2-k^2)^{-1/2}, & k < |b_{nmn'}|. \end{cases} \quad (38)$$

As a result, we obtain a final formula for the matrix s,

$$s_{nm} = \frac{1}{4\sqrt{\xi_n\xi_m}}\sum_{n'}\frac{1}{\xi_{n'}}\left(\int_{|b_{nmn'}|}^\infty \frac{\Phi_{mn',n'n}(k)kdk}{\sqrt{k^2-b_{nmn'}^2}} - i\,\mathrm{sgn}(b_{nmn'})\int_0^{|b_{nmn'}|}\frac{\Phi_{mn',n'n}(k)kdk}{\sqrt{b_{nmn'}^2-k^2}}\right). \quad (39)$$

In the limit $r\to r_i$, Eq. (32) reduces to the following differential equation:

$$\frac{d\bar{a}_n}{dr} = -\sum_m s_{nm}e^{-irb_{nm}}\bar{a}_m. \quad (40)$$

Let us compare this equation with Eq. (32). Integrating both sides of Eq. (40) and iterating the resulting equation, we obtain

$$\bar{a}_n(r) = \bar{a}_n(r_i) - \sum_m \int_{r_i}^{r} dr' s_{nm} e^{-ir' b_{nm}} \bar{a}_m(r_i) + O(s^2). \quad (41)$$

This equation coincides with Eq. (32) except for terms of order $O(s^2)$. Note that $O(s^2) \sim O(\mu^4)$ and terms of order $O(\mu^4)$ were neglected when deriving Eq. (32) from Eq. (17). Therefore, the differential equation [Eq. (40)] is equivalent to the finite-difference equation [Eq. (32)] to order $O(\mu^4)$. Also, we conclude that the Chernov method is valid if a variation in the mean field $\bar{p}$ over the distance l_0 is small. The differential equation [Eq. (40)] is more convenient for the analysis than Eq. (32) since it does include a width of cylindrical slabs $(r_{i+1} - r_i)$, which can be chosen arbitrarily within certain limits.

C. The matrix Rytov method

Equation (40) can be represented in the following general form:

$$\frac{du(r)}{dr} = M(r)u(r), \quad (42)$$

where u and M are matrices. In this subsection, an approximate solution of Eq. (42) is first considered and then applied to the solution of Eq. (40).

Assuming that the norm of the matrix $\int_0^r M(r')dr'$ is small, Eq. (42) can be solved in the Born approximation,

$$u(r) = \left(1 + \int_0^r M(r')dr'\right) u(0) + O(M^2). \quad (43)$$

To the same accuracy, this formula can be written as

$$u(r) \approx \exp\left(\int_0^r M(r')dr'\right) u(0). \quad (44)$$

In wave propagation in inhomogeneous media, a transition from Eq. (43) to Eq. (44) is well known and results in a formula for $u(r)$, which coincides with a solution of a corresponding propagation problem in the Rytov approximation (e.g., see Ref. 17). Therefore, in the present paper, $u(r)$ given by Eq. (44) is called a solution of Eq. (42) in the *matrix* Rytov method. Note that a solution of Eq. (42) given by Eq. (44) is more general than that given by Eq. (43). Indeed, the former solution is valid provided that the commutators of the matrix M for different r' are small; however it does not require that the norm of the matrix $\int_0^r M(r')dr'$ is small. In particular, when the matrix M is diagonal, Eq. (44) is an exact solution of Eq. (42).

Applying the matrix Rytov method to the solution of Eq. (40), we have

$$\bar{a}(r) = \exp(-D)\bar{a}(0), \quad (45)$$

where the matrix D is given by

$$D_{nm} = s_{nm} \int_0^r e^{-ir' b_{nm}} dr'. \quad (46)$$

In Eq. (45), the values of $\bar{a}_n(0)$ follow from Eqs. (6), (7), and (30),

$$\bar{a}_n(0) = -iu_n(z_0)/4. \quad (47)$$

Calculating the integral on the right-hand side of Eq. (46), one gets $D_{nm} = s_{nm} f_{nm}(r)$, where $f_{nm}(r) = [1 - \exp(-irb_{nm})]/ib_{nm}$. If $n = m$, then $f_{nn}(r) = r$. If $n \neq m$ and $r \gg |b_{nm}|^{-1}$, then $|f_{nm}(r)| \ll r$. For example, for the canonical sound-speed profile and the sound frequency $f = 75$ Hz, one has $\max(|b_{nm}|^{-1}) = 7.8$ km. Thus, for $r \gg \max(|b_{nm}|^{-1})$, the matrix D becomes diagonal: $D_{nm} \approx -r\delta_{nm} s_{nm}$. In this case, Eq. (45) simplifies, and the mean mode amplitudes are given by an explicit formula,

$$\bar{a}_n(r) = -\frac{i}{4} e^{-s_{nn} r} u_n(z_0). \quad (48)$$

It follows from this formula that the mean mode amplitudes do not interact and exponentially decay with the range r. The decrement of attenuation of the mode n is equal to $\mathrm{Re}\, s_{nn}$. Substitution of Eq. (48) into Eq. (31) results in an explicit formula for the mean sound field $\bar{p}(r)$.

Let us compare the decrement of attenuation $\mathrm{Re}\, s_{nn}$ with that obtained in Ref. 4 [see Eq. (27) from that reference], where a more general approach (a diagram technique) was used in derivation. In Ref. 4, the following form of the horizontal spectrum of IW was employed:

$$\hat{B}_\mu(k; z', z'') = \sum_j P_j(k) w_j(k, z') w_j(k, z''). \quad (49)$$

Here, $P_j(k)$ is the modal spectrum of the refraction index fluctuations and w_j are the IW modes (see Sec. V for more details). Substituting this expression into Eqs. (26) and (39), one finds

$$\mathrm{Re}\, s_{nn} = \frac{1}{4\xi_n} \sum_{n',j} \int_{|\xi_n - \xi_{n'}|}^{\infty} \frac{P_j(k) \left(\int_{-H}^{\infty} u_n(z) w_j(k, z) u_{n'}(z) dz\right)^2}{\xi_{n'} \sqrt{k^2 - (\xi_n - \xi_{n'})^2}} k dk. \quad (50)$$

In this formula, the upper limit of integration over k can be replaced with $|\xi_n + \xi_{n'}|$ since due to the forward scattering approximation, the random inhomogeneities with the size of the order of or smaller than λ should be absent. Furthermore, in Eq. (50) $\xi_{n'}$ in front of the square root can be replaced with $\sqrt{(\xi_n + \xi_{n'})^2 - k^2/2}$ since $k \ll \xi_n$ and the propagation constants ξ_n vary within relatively narrow limits: $\max(|\xi_n - \xi_{n'}|) \ll \xi_n$. With these replacements, Eq. (50) coincides with Eq. (27) from Ref. 4.

IV. COHERENCE FUNCTION

A. Formula for the coherence function

In the cylindrical coordinate system, the coherence function is defined as follows:

$$\Gamma(r; \varphi_1, z_1; \varphi_2, z_2) = \langle p(r, \varphi_1; z_1) p^*(r, \varphi_2; z_2) \rangle, \quad (51)$$

where two points of observation (r, φ_1, z_1) and (r, φ_2, z_2) are located at the same horizontal distance r from the source. For the horizontally isotropic problem considered in this paper, the coherence function Γ should depend only on the difference $\varphi_1 - \varphi_2$ of the azimuthal angles of the two points of observation and, in what follows, will be denoted as

$\Gamma(r,\varphi_1-\varphi_2;z_1;z_2)$. Substituting Eq. (7) into the right-hand side of Eq. (51), we have

$$\Gamma(r;\varphi_1-\varphi_2,z_1;z_2)=\sum_{n,q,m,q'}\langle a_{nq}(r)a^*_{mq'}(r)\rangle u_n(z_1)u_m(z_2)\times H_q^{(1)}(\xi_n r)(H_{q'}^{(1)}(\xi_m r))^* e^{i(q\varphi_1-q'\varphi_2)}. \tag{52}$$

It follows from this equation that for Γ to be dependent on the difference $\varphi_1-\varphi_2$, the complex amplitudes a_{nq} and $a_{mq'}$ with different azimuthal indices q and q' should be uncorrelated,[7]

$$\langle a_{nq}(r)a^*_{mq'}(r)\rangle=\langle a_{nq}(r)a^*_{mq}(r)\rangle\delta_{qq'}. \tag{53}$$

This exact result is a direct consequence of the statistical isotropy of the sound field in the horizontal plane. Substituting Eq. (53) into the right-hand side of Eq. (52) and denoting $\varphi=\varphi_1-\varphi_2$, we obtain the following expression for the coherence function of a sound field propagating in a fluctuating ocean:

$$\Gamma(r,\varphi;z_1,z_2)=\frac{2}{\pi r}\sum_{n,m}\frac{1}{\sqrt{\xi_n\xi_m}}e^{i(\xi_n-\xi_m)r}I_{nm}(r,\varphi)u_n(z_1)u_m(z_2), \tag{54}$$

where the modal correlation matrix I_{nm} is given by

$$I_{nm}(r,\varphi)=\sum_{q=-\infty}^{\infty}\langle a_{nq}(r)a^*_{mq}(r)\rangle e^{iq\varphi}. \tag{55}$$

When deriving Eq. (54) from Eq. (52), we also replaced the Hankel functions $H_q^{(1)}(\xi_n r)$ and $H_q^{(1)}(\xi_m r)$ with their asymptotics, assuming that $\xi_n r\gg q$ and $\xi_m r\gg q$. Thus, in order to calculate the coherence function $\Gamma(r,\varphi;z_1,z_2)$ given by Eq. (54), we need to know the modal correlation matrix $I_{nm}(r,\varphi)$. In Sec. IV B, this matrix will be calculated using the Chernov method.

Note that in the Cartesian coordinates x,y,z, the (transverse) coherence function is defined similarly to Eq. (51): $\Gamma(x;y_1,z_1;y_2,z_2)=\langle p(x,y_1,z_1)p^*(x,y_2,z_2)\rangle$. This coherence function depends not only on the difference y_1-y_2 between transverse coordinates y_1 and y_2 but also on their sum (e.g., see Ref. 5). Therefore, in the Cartesian coordinate system, the coherence function $\Gamma(x;y_1,z_1;y_2,z_2)$ has an extra argument in comparison with that in the cylindrical coordinates [see Eq. (54)], so that a dimension of the problem considered increases by 1. For this reason, calculation and analysis of the coherence is much simpler in cylindrical coordinates than in the Cartesian.

B. Modal correlation matrix I_{nm}

To derive a closed equation for the matrix I_{nm}, we start with Eq. (17). In that equation, replacing n with m and taking complex conjugation, one gets

$$a^*_{mq}(r)=a^*_{mq}(r_i)-i\sum_{m',q'}\int_{r_i}^{r}dr'[V^{m'q'}_{mq}(r')]^*a^*_{m'q'}(r_i)-\sum_{m',q',m'',q''}\int_{r_i}^{r}dr'[V^{m'q'}_{mq}(r')]^*\times\int_{r_i}^{r'}dr''[V^{m''q''}_{m'q'}(r'')]^*a^*_{m''q''}(r_i)+O(\mu^3). \tag{56}$$

Multiplying this equation by Eq. (17) and then averaging both sides of the resulting equation, a formula for $\langle a_{nq}(r)a^*_{mq}(r)\rangle$ within the $(i+1)$th slab is obtained. The correlators appearing on the right-hand side of this formula are split similarly to those in Eq. (18). As a result, an explicit expression for $\langle a_{nq}(r)a^*_{mq}(r)\rangle$ takes the form

$$\begin{aligned}\langle a_{nq}(r)a^*_{mq}(r)\rangle=&\langle a_{nq}(r_i)a^*_{mq}(r_i)\rangle\\&-\sum_{n',n'',q'}\int_{r_i}^{r}dr'\int_{r_i}^{r'}dr''\langle V^{n'q'}_{nq}(r')V^{n''q}_{n'q'}(r'')\rangle\times\langle a_{n''q}(r_i)a^*_{mq}(r_i)\rangle\\&-\sum_{m',m'',q'}\int_{r_i}^{r}dr'\int_{r_i}^{r'}dr''\langle V^{m'q'}_{mq}(r')V^{m''q}_{m'q'}(r'')\rangle^*\times\langle a_{nq}(r_i)a^*_{m''q}(r_i)\rangle\\&+\sum_{n',q',m'}\int_{r_i}^{r}dr'\int_{r_i}^{r}dr''\langle V^{n'q'}_{nq}(r')V^{mq}_{m'q'}(r'')\rangle\times\langle a_{n'q'}(r_i)a^*_{m'q'}(r_i)\rangle.\end{aligned} \tag{57}$$

When deriving this expression, Eq. (16) was used.

It follows from Eq. (33) that in the second and third terms on the right-hand side of Eq. (57), the sum over n',q' and the sum over m',q' are equal to $S_{nn''}(r_i,r)$ and $S^*_{mm''}(r_i,r)$, respectively. Let us multiply both sides of Eq. (57) by $e^{iq\varphi}$ and then sum over q. As a result, one finds

$$\begin{aligned}I_{nm}(r,\varphi)=&I_{nm}(r_i,\varphi)-\sum_{n''}S_{nn''}(r_i,r)I_{n''m}(r_i,\varphi)\\&-\sum_{m''}S^*_{mm''}(r_i,r)I_{mm''}(r_i,\varphi)\\&+\sum_{n',q',m'}\int_{r_i}^{r}dr'\int_{r_i}^{r}dr''\sum_q e^{iq\varphi}\langle V^{n'q'}_{nq}(r')\\&\times V^{mq}_{m'q'}(r'')\rangle\langle a_{n'q'}(r_i)a^*_{m'q'}(r_i)\rangle.\end{aligned}$$

In the last term on the right-hand side of this equation, the V-correlator is replaced with its value from Eq. (25). Then, the sum over q is calculated using Eq. (28), which gives the factor $2\pi\delta(\theta-\varphi)$. After integration over θ, we obtain the factor $e^{iq'\varphi}$ and, using Eq. (55), calculate the sum over q'. As a result, we obtain the following equation:

$$I_{nm}(r,\varphi) = I_{nm}(r_i,\varphi) - \sum_{n''} S_{nn''}(r_i,r) I_{n''m}(r_i,\varphi) - \sum_{m''} S^*_{mm''}(r_i,r) I_{nm''}(r_i,\varphi) + \sum_{n',m'} \Pi_{nm,n'm'}(r_i,r;\varphi) I_{n'm'}(r_i,\varphi), \quad (58)$$

where the matrix Π is given by

$$\Pi_{nm,n'm'}(r_i,r;\varphi) = \frac{1}{4}\int_{r_i}^{r} dr' \int_{r_i}^{r} dr'' \frac{\exp[i(\xi_{n'}-\xi_n)r' + i(\xi_m - \xi_{m'})r'']}{\sqrt{\xi_n \xi_{n'} \xi_{m'} \xi_m}} \times \int_0^\infty J_0(k\sqrt{r'^2 + r''^2 - 2r'r''\cos\varphi}) \Phi_{nn',mm'}(k) k dk. \quad (59)$$

Similar to the case of the mean sound field, Eq. (59) can be simplified by introducing new integration variables $\rho=(r'+r'')/2$ and $\eta=r'-r''$,

$$\Pi_{nm,n'm'}(r_i,r;\varphi) = \frac{1}{4}\int_{r_i}^{r} d\rho \frac{e^{i\rho b_2}}{\sqrt{\xi_n \xi_m \xi_{n'} \xi_{m'}}} \int_0^\infty dk k \Phi_{nn',mm'}(k) \times \int_{-\infty}^{\infty} d\eta e^{i\eta b_1} J_0\left(k\sqrt{4\rho^2 \sin^2\frac{\varphi}{2} + \eta^2 \cos^2\frac{\varphi}{2}}\right). \quad (60)$$

Here, b_1 and b_2 are the following combinations of the horizontal wave numbers:

$$b_1 = \frac{\xi_{n'} - \xi_n - \xi_m + \xi_{m'}}{2}, \quad b_2 = \xi_{n'} - \xi_n + \xi_m - \xi_{m'}. \quad (61)$$

Strictly speaking, in Eq. (60) the limits of integration over η are given by the inequality $|\eta| \leqslant r - r_i - |r + r_i - 2\rho|$. However, similar to the case of the mean field, these limits of integration can be extended to $(-\infty,\infty)$.

Equation (60) can be simplified further. First, for small values of φ, the argument of the Bessel function J_0 in this equation can be replaced with $k\sqrt{\rho^2\varphi^2 + \eta^2}$. Then, the integral over η can be calculated with the use of Eq. (6.677.3) in Ref. 15. As a result, we obtain

$$\Pi_{nm,n'm'}(r_i,r;\varphi) = \frac{1}{2}\int_{r_i}^{r} d\rho \frac{e^{i\rho b_2}}{\sqrt{\xi_n \xi_m \xi_{n'} \xi_{m'}}} \times \int_{|b_1|}^\infty dk k \Phi_{nn',mm'}(k) \frac{\cos(\rho\varphi\sqrt{k^2 - b_1^2})}{\sqrt{k^2 - b_1^2}}. \quad (62)$$

Now we can consider the limit $r \to r_i$ and, similar to Sec. III B, replace Eq. (58) with the following differential equation for the modal correlation matrix:

$$\frac{dI_{nm}(r,\varphi)}{dr} = -\sum_{n'} e^{i(\xi_{n'}-\xi_n)r} s_{nn'} I_{n'm}(r,\varphi) - \sum_{m'} e^{i(\xi_m - \xi_{m'})r} s^*_{mm'} I_{nm'}(r,\varphi) + \sum_{n',m'} e^{i(\xi_{n'} - \xi_n + \xi_m - \xi_{m'})r} \sigma_{nm,n'm'}(\varphi) I_{n'm'}(r,\varphi). \quad (63)$$

Here, the matrix σ does not depend on r and is given by

$$\sigma_{nm,n'm'}(\varphi) = \frac{1}{2\sqrt{\xi_n \xi_m \xi_{n'} \xi_{m'}}} \int_0^\infty dk k \Phi_{nn',mm'}(k) \frac{\cos(\rho\varphi\sqrt{k^2 - b_1^2})}{\sqrt{k^2 - b_1^2}}. \quad (64)$$

The matrix σ has the same symmetries, Eq. (27), as the matrix Φ does. Note that in Eq. (63), the azimuthal angle φ appears as a parameter. The initial condition to Eq. (63) is formulated at $r=0$. Similar to the derivation of Eq. (47), we have

$$I_{nm}(0,\varphi) = \tfrac{1}{16} u_n(z_0) u_m(z_0). \quad (65)$$

This matrix does not depend on φ and, in what follows, is denoted as $I_{nm}(0)$.

To check the energy conservation, in Eq. (63) we set $\varphi=0$ and $m=n$ and sum the resulting equation over n,

$$\frac{d}{dr}\sum_n I_{nn}(r,0) = -\sum_{n,m}\left[s_{mn} + s^*_{nm} - \sum_{n'} \sigma_{n'n',nm}(0)\right] e^{i(\xi_n - \xi_m)r} I_{nm}(r,0). \quad (66)$$

Using Eqs. (39) and (64) and taking into account the symmetries of the matrix $\sigma_{mn,n'm'}$, Eq. (27), it can be easily seen that the expression in the square brackets in Eq. (66) is equal to zero,

$$s_{mn} + s^*_{nm} - \sum_{n'} \sigma_{n'n',nm}(0) = 0. \quad (67)$$

[This relationship is called the optical theorem (e.g., see Refs. 16 and 18).] Thus, Eq. (66) becomes

$$\sum_n I_{nn}(r,0) = \text{const}$$

This equality manifests conservation of the acoustic energy.[5]

Equation (63) can be represented in a general form of Eq. (42),

$$\frac{dI_{nm}(r,\varphi)}{dr} = -\sum_{n',m'} T_{nm,n'm'}(r,\varphi) I_{n'm'}(r,\varphi), \quad (68)$$

where the matrix T is given by

$$T_{nm,n'm'}(r,\varphi) = e^{-i(\xi_n - \xi_{n'})r}\delta_{mm'} s_{nn'} + e^{i(\xi_m - \xi_{m'})r}\delta_{nn'} s^*_{mm'} - e^{-i(\xi_n - \xi_{n'} - \xi_m + \xi_{m'})r}\sigma_{nm,n'm'}(\varphi). \quad (69)$$

Equation (68) can be solved approximately using the matrix

Rytov method. As a result, we obtain the following explicit formula for the modal correlation matrix I:

$$I^R(r,\varphi) = \exp\left(-\int_0^r T(r',\varphi)dr'\right) I(0). \quad (70)$$

Here, the calculation of a matrix exponent is considered as an explicit operation, and the superscript R stands for "Rytov." Note that in the last term on the right-hand side of Eq. (69), a linear combination of propagation constants can be very small for the canonical sound-speed profile and other realistic profiles. Therefore, the integration over r' of the last term cannot be simplified as in the case for the mean field.

V. NUMERICAL IMPLEMENTATION

In this section, the theory developed above is used for numerical studies of sound propagation in an oceanic waveguide with a random field of IW.

A. Garrett–Munk spectrum and ocean stratification

To calculate matrices s_{nm} and T, which appear in the equations for the mean field and coherence function, we need to know the correlation function $B_\mu(|\boldsymbol{r}'-\boldsymbol{r}''|;z',z'') = \langle \mu(\boldsymbol{r}',z')\mu(\boldsymbol{r}'',z'')\rangle$ of the random field μ due to IW. In Eq. (2) for μ, we write $c(\boldsymbol{r},z)=c_0(z)+\delta c(\boldsymbol{r},z)$, where $\delta c(\boldsymbol{r},z)$ describes sound-speed fluctuations due to IW. Taking into account that $\delta c/c_0 \ll 1$, the correlation function B_μ is written as

$$B_\mu(|\boldsymbol{r}'-\boldsymbol{r}''|;z',z'') = 4\frac{\omega^4}{c_0^3(z')c_0^3(z'')}\langle \delta c(\boldsymbol{r}',z')\delta c(\boldsymbol{r}'',z'')\rangle. \quad (71)$$

Following Ref. 19, in this formula δc is represented as $\delta c(\boldsymbol{r},z)=[c_0(z)\mu N^2(z)/g]\xi(\boldsymbol{r},z)$. Here, $\mu \approx 24.5$, $g=9.8\ \mathrm{m/s^2}$ is the gravitational acceleration, $N(z)$ is the Brunt–Väisälä frequency, and ξ is the vertical displacement of a fluid parcel due to IW. As a result, we obtain the following expression for B_μ:

$$B_\mu(r;z',z'') = 4\frac{\omega^4}{c_0^2(z')c_0^2(z'')}\frac{\mu^2 N^2(z')N^2(z'')}{g^2} \times \sum_j \int d\boldsymbol{k} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} G_\xi(j,k) w_j(k,z') w_j(k,z''). \quad (72)$$

Here, $w_j(k,z)$ are internal wave modes, where the index j indicates the IW mode number, and G_ξ is the modal spectrum of IW. Note that due to statistical isotropy of the sound-speed fluctuations, the 3D spectrum $B_\mu(r;z',z'')$ is expressed in terms of the 2D spectrum $G_\xi(j,k)$. (Individual realizations of the sound-speed fluctuations are, of course, fully 3D.) The IW modes satisfy the equation

$$\frac{d^2 w_j(k,z)}{dz^2} + k^2\frac{N^2(z)-\omega^2}{\omega^2-f_i^2} w_j(k,z) = 0,$$

the boundary conditions $w_j(k,0)=w_j(k,-H)=0$, and the normalization conditions

$$\int_{-H}^0 [N^2(z)-f_i^2] w_j(k,z) w_{j'}(k,z)dz = \delta_{jj'}.$$

Here, f_i is the inertial frequency.

Comparing Eqs. (24) and (72), we find that the horizontal spectrum of the fluctuations is given by

$$\hat{B}_\mu(k;z',z'') = \omega^4\frac{8\pi\mu^2}{g^2}\frac{N^2(z')N^2(z'')}{c_0^2(z')c_0^2(z'')}\sum_j G_\xi(j,k) w_j(k,z') w_j(k,z''). \quad (73)$$

Substituting Eq. (73) into Eq. (26), we obtain the following expression for the matrix Φ:

$$\Phi_{nn',mm'}(k) = \sum_j G_\xi(j,k) Q^j_{nn'}(k) Q^j_{mm'}(k),$$

where the matrix Q^j_{nm} is given by

$$Q^j_{nm}(k) = Q^j_{mn}(k) = \omega^2\frac{\sqrt{8\pi}\mu}{g}\int_{-H}^0 \frac{N^2(z)}{c_0^2(z)} u_n(z) w_j(k,z) u_m(z) dz.$$

In numerical calculations below, we will assume that $f_i = 7.3\times 10^{-5}\ \mathrm{s^{-1}}$ and that the Brunt–Väisälä frequency and the sound speed are given by the Munk canonical profiles,

$$N(z) = N_0 \exp(-|z|/B),$$

$$c_0(z) = c_{00}[1 + 0.0057(e^{-2(|z|-z_1)/B} + 2(|z|-z_1)/B - 1)],$$

where $N_0 = 5.2\times 10^3\ \mathrm{s^{-1}}$, $B \approx 1000$ m, $z_1 = 890$ m, and $c_{00} = 1500$ m/s. Furthermore, we will assume that G_ξ is given by the Garrett-Munk spectrum (e.g., see Ref. 19),

$$G_\xi(j,k) = \frac{2B^3 N_0^2 E}{\pi^2 M}\frac{jj_*}{j^2+j_*^2}\frac{q_* k}{(k^2+j^2 q_*^2)^2}. \quad (74)$$

Here, $j_*=3$ and $E=6.3\times 10^{-5}$ are empirical dimensionless constants, $q_* = \pi f/(BN_0)$, and $M = \Sigma_{j=1}^\infty (j^2+j_*^2)^{-1} \approx (\pi j_* - 1)/(2j_*^2)$.

B. Numerical results

First, the dependence of the solution of the differential equation for the modal correlation matrix $I(r,\varphi)$, Eq. (68), on the range step Δr is studied numerically. (As we saw in Sec. III B, this solution should not depend on the range step provided that Δr is greater than l_0 but less than the horizontal scale of variation of the coherence function.) To this end, we consider the normalized difference $d(\Delta r_1,\Delta r_2) = (2\|I_{nm}(\Delta r_1) - I_{nm}(\Delta r_2)\|)/(\|I_{nm}(\Delta r_1)\| + \|I_{nm}(\Delta r_2)\|)$, which characterizes the difference in matrices I_{nm} calculated for different Δr. The values of $d(\Delta r_1,\Delta r_2)$ are presented in Table I for different values of Δr_1 and Δr_2. The results in the table correspond to the sound propagation range $r=1000$ km and the azimuthal angle $\varphi=0$. Furthermore, for this table and all numerical results below, the sound frequency $f=75$ Hz, the source depth $z_0=-807$ m, and only the first 25 propagating acoustic modes are taken into account. The values of Δr were obtained by splitting the propagation range into L

TABLE I. The normalized difference $d(\Delta r_1, \Delta r_2)$ between the modal correlation matrices I calculated for different Δr. For this table, the sound propagation range $r=1000$ km, the sound frequency $f=75$ Hz, the source depth $z_0=-807$ m, and the azimuthal angle $\varphi=0$.

Δr_1, Δr_2 (km)	50, 33.3	33.3, 25	25, 20	20, 16.7	16.7, 14.3	14.3, 12.5	12.5, 11.1
$d(\Delta r_1, \Delta r_2)\times 10^{-3}$	6.33	4.05	1.88	1.28	1.27	1.20	0.87

$=20,30,\ldots,90$ slabs. The results presented in the table (and similar results obtained but not presented here) indicate that the matrix I_{nm} depends on Δr only slightly. Therefore, the subsequent numerical calculations were done for the range step $\Delta r=25$ km. This value of Δr was chosen as a trade-off between calculation accuracy and time.

Next the difference between the matrix I_{nm} obtained as a solution of Eq. (68) and the matrix I^R_{nm} calculated with the use of the matrix Rytov method, Eq. (70), is studied. In Fig. 2, the normalized difference $(2\|I_{nm}-I^R_{nm}\|)/(\|I_{nm}\|+\|I^R_{nm}\|)$ between I_{nm} and I^R_{nm} is plotted as a function of the sound propagation range r for $\varphi=0$. It follows from the figure that the normalized difference between I_{nm} and I^R_{nm} is small. Therefore, either of two approaches can be used here.

In Fig. 3, the magnitudes of the entries of the modal correlation matrix $I_{nm}(r,\varphi=0)$ are plotted versus the indices n and m for four propagation ranges $r=25$, 250, 500, and 1000 km. It follows from the figure that the cross-modal correlations decrease with range. Furthermore, equipartition of the mode intensities I_{nn} with range can be seen in Fig. 3. More clearly, this effect is depicted in Fig. 4, where the mode intensities $I_{nn}(r,\varphi=0)$ are plotted versus the mode number n for different r. Note that equipartition of the mode intensities was theoretically predicted by Dozier and Tappert[6] for the case of sound propagation in a 2D ocean. Figures 3 and 4 (and other numerical calculations done but not shown here) confirm that this result is also valid for the 3D case.

If in Eq. (51) $\varphi_1=\varphi_2$ and $z_1=z_2=z$, the coherence function becomes the mean sound intensity $\mathcal{I}(r,z)=\langle p(r,\varphi_1;z)p^*(r,\varphi_1;z)\rangle$ at the range r and depth z. Figure 5 shows the dependence of the mean sound intensity (normalized by its maximal value) on depth z for four values of the range r. For $r=25$ km, the interference structure of the mean intensity $\mathcal{I}$ can be seen in the figure. As r increases, scattering tends to smooth this structure of $\mathcal{I}$ and slightly broaden it.

To study the dependence of the coherence function $\Gamma(r,\varphi;z_1,z_2)$ on the horizontal separation between hydrophones, it is worthwhile to set $z_1=z_2=z$ and average the coherence function over depth z (see Ref. 5). The resulting depth-averaged coherence function is denoted as $\bar{\Gamma}(r,\varphi)$ and can be obtained by integrating the right-hand side of Eq. (54) over z and using Eq. (4),

$$\bar{\Gamma}(r,\varphi)=\frac{2}{\pi r}\sum_n \frac{I_{nn}(r,\varphi)}{\xi_n}. \tag{75}$$

Figure 6 depicts the dependence of the depth-averaged coherence function $\bar{\Gamma}(r,\varphi)$ (normalized by its maximum value) on the hydrophone separation $\Delta y=r\varphi$ in the horizontal plane for the four ranges considered above. It follows from Fig. 6 that at $r=25$ km the coherence function depends only slightly on Δy. An increase in the range r results in a decrease of the coherence, as it should be. Figure 6 allows us to calculate the correlation radius ρ_{corr} of the sound field. For $r=500$ and 1000 km, the correlation radii are $\rho_{\text{corr}}=3$ and 1.8 km, respectively. Note that for very small Δy, an analytical expression for the depth-averaged coherence function is derived in the Appendix and compared with that in Ref. 5.

Note that the sufficient condition of the applicability of the theory developed, Eq. (14), is approximately valid up to the ranges $r\lesssim 250\sim 500$ km for the numerical examples above. Beyond these ranges, strictly speaking, the results obtained can be considered only as an extrapolation. However,

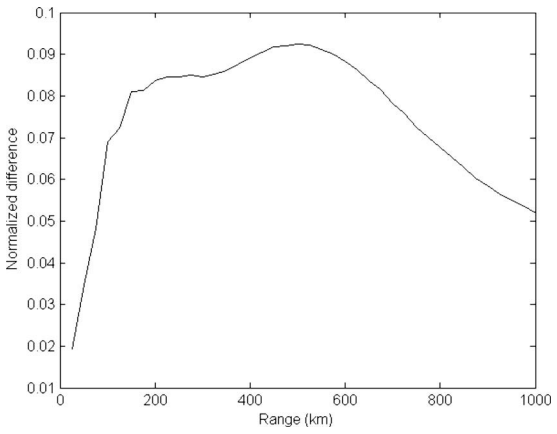


FIG. 2. The normalized difference between the modal correlation matrices I_{nm} and I^R_{nm} vs the sound propagation range r for $f=75$ Hz, $z_0=-807$ m, and $\varphi=0$.

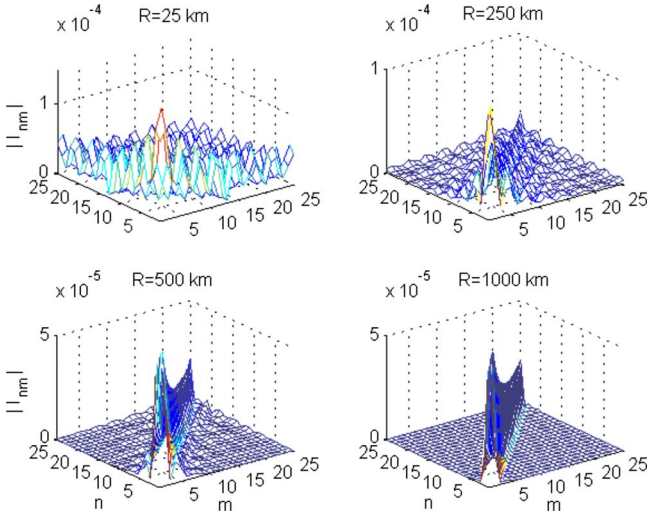


FIG. 3. (Color online) The magnitudes of the entries of the modal correlation matrix I_{nm} vs the indices n and m for four ranges r and $f=75$ Hz, $z_0=-807$ m, and $\varphi=0$.

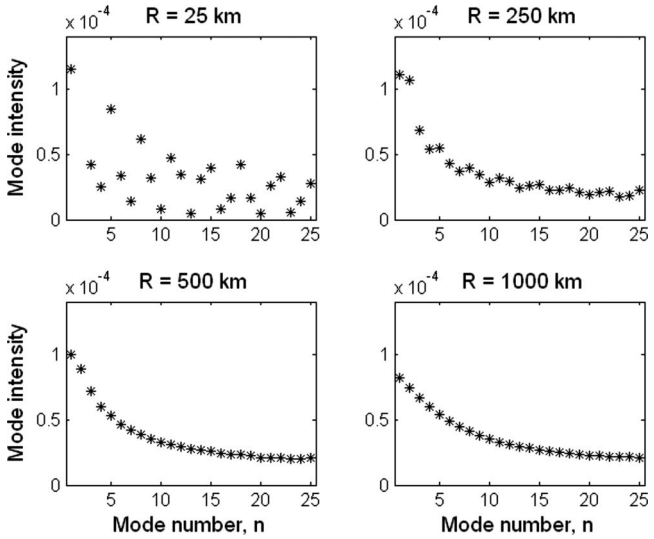


FIG. 4. The mode intensities I_{nn} vs the mode number n for four ranges r and $f=75$ Hz, $z_0=-807$ m, and $\varphi=0$.

as indicated in Sec. II B, the sufficient condition Eq. (14) is probably too restrictive so that the numerical results obtained for $r=1000$ km could still be valid.

VI. CONCLUSIONS

In this paper, a theory was developed for the calculation of the first two statistical moments of a sound field due to an omnidirectional monochromatic source in a 3D fluctuating ocean with random inhomogeneities, which are statistically isotropic in a horizontal plane. To account for this symmetry, a cylindrical coordinate system was used in which the sound field was decomposed into a sum of acoustic modes and azimuthal harmonics. The equations governing the evolution of the mean field and coherence function were obtained using the Chernov method. The equation for the coherence function is much simpler than those known in the literature due to the axial symmetry of the problem, which was used in the derivation. It was shown that this equation conserves

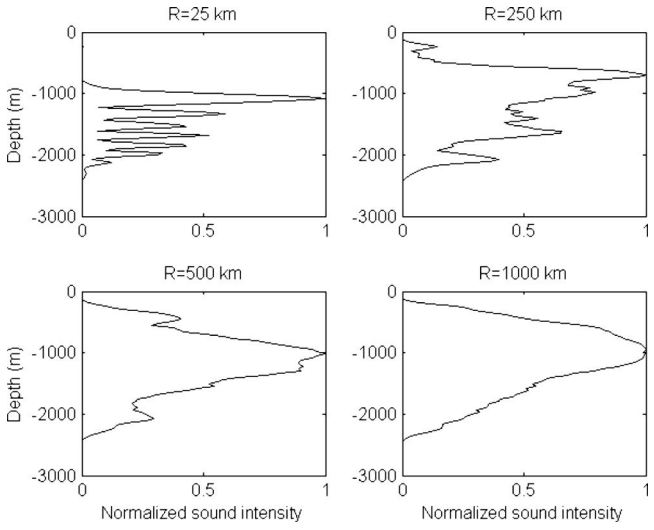


FIG. 5. The mean sound intensity $\mathcal{I}$ vs the ocean depth z for four ranges r and $f=75$ Hz and $z_0=-807$ m.

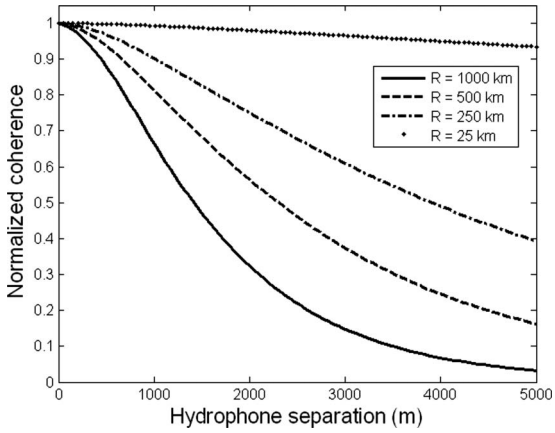


FIG. 6. The normalized depth-averaged coherence function $\bar{\Gamma}$ vs the horizontal hydrophone separation $\Delta y=r\varphi$ for four ranges r and $f=75$ Hz and $z_0=-807$ m.

acoustic energy. Using the matrix Rytov method, explicit expressions for the mean field and coherence function were also derived.

Using the equations derived, the dependence of the coherence function on parameters of the problem was studied numerically. In particular, equipartition of the mode intensities with range was revealed in 3D fluctuating ocean. Furthermore, the dependence of the depth-averaged coherence function on the hydrophone separation and range was studied.

ACKNOWLEDGMENTS

This work was supported by ONR Grant No. N0001407IP20046. We would like to thank three anonymous reviewers for their critical comments. In particular, the importance of the correction term in Eq. (13) was indicated in one of the reviews.

APPENDIX: THE DEPTH-AVERAGED COHERENCE FUNCTION

In this appendix, an analytical formula for the depth-averaged coherence function is derived for the case of small φ. Due to the optical theorem, Eq. (67), the operator $T(r,\varphi)$ given by Eq. (69) has an eigenvalue equal to zero, which corresponds to the following normalized eigenvector:

$$\tilde{I}_{nm}=\frac{1}{N_{\mathrm{mod}}^{1/2}}\delta_{nm}, \tag{A1}$$

where N_{mod} is the number of acoustic modes. This state corresponds to equipartition of energy between modes and absence of cross-mode correlations. The operator $\tilde{T}=\int_0^r T(r',\varphi)dr'$ is generally not self-adjoint. However, for $\varphi=0$ the vector $\tilde{I}$ is an eigenvector of the adjoint operator $\tilde{T}'$ corresponding also to the zero eigenvalue. Then, for small φ, according to the standard perturbation analysis, this eigenvalue becomes

$$\left\langle \tilde{I} \left| \frac{\partial \tilde{T}}{\partial(\varphi^2)} \right| \tilde{I} \right\rangle \varphi^2$$

$$= \varphi^2 \frac{1}{N_{\mathrm{mod}}} \sum_{n,m} \frac{\partial \tilde{T}_{nn,mm}}{\partial(\varphi^2)}$$

$$= \varphi^2 \frac{\pi k_0^4}{2N_{\mathrm{mod}}} \sum_{n,m} \frac{1}{\xi_n \xi_m} \int_0^\infty dk k \Phi_{nm,nm}(k) \int_0^R d\rho \rho^2$$

$$\times \int_{-\infty}^{\infty} d\eta e^{i(\xi_m - \xi_n)\eta} \frac{J_1(k\eta)}{2\eta}$$

$$= \varphi^2 \frac{\pi k_0^4}{2N_{\mathrm{mod}}} \frac{R^3}{3} \sum_{n,m} \frac{1}{\xi_n \xi_m} \int_{k>|\xi_n - \xi_m|}^{\infty} dk k \Phi_{nm,nm}(k)$$

$$\times \sqrt{k^2 - (\xi_n - \xi_m)^2}. \tag{A2}$$

Here, we have used the formula [see Eq. (6.693) in Ref. 15]

$$\int_0^\infty \cos \beta x \frac{J_1(x)}{x} dx = \begin{cases} \sqrt{1-\beta^2}, & \beta < 1 \\ 0, & \beta > 1. \end{cases} \tag{A3}$$

Let us assume that in Eq. (70) the zero eigenvalue contribution dominates. Then from Eq. (54), for the depth-averaged coherence function, we find that

$$\bar{\Gamma}(r, \Delta y) = \bar{\Gamma}(r, 0) \exp[-(\Delta y / y_*)^2].$$

[Here, the argument φ of $\bar{\Gamma}(r, \varphi)$ was replaced with $\Delta y = r\varphi$.] In this equation,

$$\frac{1}{y_*^2} = R \frac{\pi k_0^4}{6N_{\mathrm{mod}}} \sum_{n,m} \frac{1}{\xi_n \xi_m} \int_{k>|\xi_n - \xi_m|}^{\infty} dk k \Phi_{nm,nm}(k) \sqrt{k^2 - (\xi_n - \xi_m)^2}. \tag{A4}$$

Here, one can replace integration variable $k^2 = k'^2 + (\xi_n - \xi_m)^2$ and compare the result with Eqs. (107) and (111) from Ref. 5. The results will coincide completely if in Eq. (A4) one replaces factor $\pi/6$ by $\pi/4$. The difference is due to the accurate treatment of cylindrical spreading here; in Ref. 5 the case of plane wave propagation was actually considered [see Eq. (100) in that reference].

[1] W. E. Kohler and G. C. Papanicolau, "Wave propagation in a randomly inhomogeneous ocean," in *Wave Propagation and Underwater Acoustics*, Lecture Notes in Physics, Vol. **70**, edited by J. B. Keller and J. S. Papadakis (Springer-Verlag, Berlin, 1977), pp. 153–223.

[2] R. Dashen, "Path-integral treatment of acoustical mutual coherence functions for rays in a sound channel," J. Acoust. Soc. Am. **77**, 1716–1722 (1985).

[3] S. M. Flatte, R. Dashen, W. H. Munk, K. M. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, Cambridge, England, 1979).

[4] A. G. Voronovich and V. E. Ostashev, "Mean field of a low-frequency sound wave propagating in a fluctuating ocean," J. Acoust. Soc. Am. **119**, 2101–2105 (2006).

[5] A. G. Voronovich and V. E. Ostashev, "Low-frequency sound scattering by internal waves in the ocean," J. Acoust. Soc. Am. **119**, 1406–1419 (2006).

[6] L. Dozier and F. Tappert, "Statistics of normal mode amplitudes in a random ocean," J. Acoust. Soc. Am. **63**, 353–365 (1978).

[7] C. Penland, "Acoustic normal mode propagation through a three-dimensional internal wave field," J. Acoust. Soc. Am. **78**, 1356–1365 (1985).

[8] S. Frankenthal and M. J. Beran, "Propagation in random stratified waveguides: A modal-spectral treatment," J. Acoust. Soc. Am. **104**, 3282–3295 (1998).

[9] A. K. Morozov and J. A. Colosi, "Stochastic differential equation analysis for sound scattering by random internal waves in the ocean," Acoust. Phys. **53**, 335–347 (2007).

[10] A. G. Sazontov, "A quasi-classical solution to the radiation transport equation in a scattering medium with regular refraction," Acoust. Phys. **42**, 487–494 (1996).

[11] E. Yu. Gorodetskaya, A. I. Malekhanov, A. G. Sazontov, and N. K. Vdovicheva, "Deep-water acoustic coherence at long-ranges: Theoretical prediction and effects on large-array signal processing," IEEE J. Ocean. Eng. **24**, 156–170 (1999).

[12] L. A. Chernov, *Waves in Randomly-Inhomogeneous Media* (Nauka, Moscow, 1975) (in Russian).

[13] L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media* (Springer-Verlag, Berlin, 1992).

[14] V. E. Ostashev and V. I. Tatarskii, "Representation of the Helmholtz equation solution in the form of a series based on backscattering multiplicity," Waves Random Media **5**, 125–135 (1995).

[15] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1994).

[16] S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radio Physics. Part 4, Wave Propagation Through Random Media* (Springer, Berlin, 1989).

[17] M. L. Oristaglio, "Accuracy of the Born and Rytov approximations for reflection and refraction at a plane interface," J. Opt. Soc. Am. A **2**, 1987–1993 (1985).

[18] L. M. Landau and E. M. Liftshitz, *Quantum Mechanics (Non-Relativistic Theory)*, Course of Theoretical Physics, Vol. **3** (Elsevier, Amsterdam, 2005).

[19] J. Colosi and M. Brown, "Efficient numerical simulation of stochastic internal-wave induced sound-speed perturbation fields," J. Acoust. Soc. Am. **103**, 2232–2235 (1998).

Empirical dependence of acoustic transmission scintillation statistics on bandwidth, frequency, and range in New Jersey continental shelf

Mark Andrews
Department of Electrical and Computer Engineering, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Tianrun Chen
Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Purnima Ratilal
Department of Electrical and Computer Engineering, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

(Received 11 March 2008; revised 1 November 2008; accepted 6 November 2008)

The scintillation statistics of broadband acoustic transmissions are determined as a function of signal bandwidth B, center frequency f_c, and range with experimental data in the New Jersey continental shelf. The received signal intensity is shown to follow the Gamma distribution implying that the central limit theorem has led to a fully saturated field from independent multimodal propagation contributions. The Gamma distribution depends on the mean intensity and the number of independent statistical fluctuations or coherent cells μ of the received signal. The latter is calculated for the matched filter, the Parseval sum, and the bandpassed center frequency, all of which are standard ocean acoustic receivers. The number of fluctuations μ of the received signal is found to be an order of magnitude smaller than the time-bandwidth product TB of the transmitted signal, and to increase monotonically with relative bandwidth B/fc. A computationally efficient numerical approach is developed to predict the mean intensity and the corresponding broadband transmission loss of a fluctuating, range-dependent ocean waveguide by range and depth averaging the output of a time-harmonic stochastic propagation model. This model enables efficient and accurate estimation of transmission loss over wide areas, which has become essential in wide-area sonar imaging applications. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3037228]

PACS number(s): 43.30.Re, 43.30.Pc, 43.60.Cg [RCG] Pages: 111–124

I. INTRODUCTION

Acoustic signals transmitted through an ocean waveguide typically scintillate in both time and space upon reception due to multimodal propagation and random variations of the medium and its boundaries. Quantitative knowledge of these received signal fluctuations is often essential in the design of ocean acoustic experiments and in the analysis of subsequent measurements.[1–3] Extensive research have been conducted to investigate the effects of random oceanographic variations on the acoustic field propagated through continental shelf[4–9] as well as in the deep ocean waveguides.[10,11] The corresponding statistical analyses, however, have typically been limited to either narrowband or broadband transmissions with a fixed time-bandwidth product TB and not as a function of transmitted signal's bandwidth, frequency, and time duration.

Here we determine the scintillation statistics of acoustic transmissions as a function of transmitted bandwidth B, center frequency f_c, and range in a continental shelf waveguide for signals where the transmitted time-bandwidth product TB may be large. Signals with large TB are important for obtaining high range resolution and high signal-to-noise ratio in many ocean-acoustic applications by use of matched filter (MF)[12,13] receivers. We use data obtained from the Main Acoustic Experiment (MAE) 2003 (Refs. 2 and 14) of the Office of Naval Research Geoclutter Program conducted at the Strataform site on the New Jersey continental shelf. Broadband pulses of varying bandwidths and time-bandwidth products in the low to midfrequency range transmitted by a vertical source array and measured on a towed horizontal receiving array are analyzed.

We show that received signal intensity follows the Gamma distribution.[15,16] This implies that the central limit theorem has led to fully saturated fluctuations at the receiver which arise from independent multimodal propagation contributions. The Gamma distribution depends on the mean intensity and the number of independent statistical fluctuations or coherence cells μ of the received signal. We determine the number of coherence cells empirically by Rice's method[17] for several standard receivers commonly used in ocean acoustics, such as the MF, the Parseval sum (PS), and the bandpassed center frequency. We show that the number of independent statistical fluctuations of the received signal μ is not equal to the time-bandwidth product TB of the original transmitted signal for large TB but is typically an order of magnitude smaller. This implies that much of a large TB

signals' original bandwidth remains statistically correlated after propagation through the random waveguide. We then show that the number of received fluctuations μ increases monotonically with the relative bandwidth B/f_c of the transmitted signal.

Our analysis provides a useful method for determining the scintillation statistics of signals transmitted with arbitrary bandwidth in the low to midfrequency range through a fluctuating continental shelf waveguide. This knowledge can then be applied to aid in the design of ocean acoustic experiments and in subsequent measurement analysis. It can also be used to determine error bounds on parameter estimates and determine sample sizes necessary to reduce estimation errors so that they fall within tolerable ranges.[13,18]

In active sonar applications, the MF is extensively applied to maximize signal-to-noise ratio in target detection and enhance range resolution in target localization.[2,12,19–21] The *peak* MF output is often used to determine both the scattered field level and the location of a discrete target.[22–25] Matched filtering is a coherent process where the received signal is correlated to the transmitted waveform. Since acoustic signals propagated long ranges through an ocean waveguide are distorted from multipath propagation and modal dispersion, the MF output becomes highly susceptible to mismatch in an ocean waveguide.

Contrary to the situation in free space where the MF output does not vary with source-receiver range, we show that in an ocean waveguide it degrades significantly and with increasing range due to the effects of cumulative modal dispersion. We show that this degradation must be taken into account to reliably infer source level, scattering strength, or target strength from active sonar data in a continental shelf waveguide, and provide an approach to do so. In free space environments, the fact that the peak output of the MF equals[26,27] the PS energy is often used to normalize the MF output. We show that another approach must be used in an ocean waveguide because the peak MF output tends to have a lower mean and higher standard deviation than an incoherent Parseval energy summation of the received signal.

When a wide-area sonar system is used to study scattering from distributed scatterers, such as a large fish shoal or the sea bottom, over areas spanning tens of thousands of km^2 with varying bathymetry, it is necessary to correct for transmission loss (TL) over these vast areas.[2,14,28–31] Since the waveguides are often random and the transmitted signals are often broadband, the number of computations required per radial with a range-dependent waveguide propagation model such as the parabolic equation make brute-force Monte Carlo simulations impractical for many applications that require rapid analysis of hundreds of wide-area sonar images typically collected in a single day at sea.[2,14]

Here, we provide an approach for rapidly and accurately estimating broadband TL expected over wide areas. The approach uses the range-dependent acoustic model (RAM)[32] to calculate the acoustic intensity at the center frequency of the signal, after propagation through a waveguide randomized by sound speed fluctuations. A combination of range and depth averaging is then used to approximate the *expected* broadband intensity in a fluctuating environment. For the New Jersey continental shelf, we show that only three to five Monte Carlo simulations are required to converge on the mean TL for each radial. This approach has proven to be crucial in enabling rapid wide area scattering strength analysis in recent ocean imaging experiments.[2,14]

We describe relevant aspects of the MAE 2003 field experiment in Sec. II and present statistical analysis of the acquired data. In Sec. III, we describe our method for rapidly estimating the TL expected over wide areas and show how the method can be calibrated using the broadband data collected from the MAE 2003 field experiment.

II. BROADBAND TRANSMISSION DATA

A. Data collection

The experiment was conducted from April 27 to May 15, 2003 (Ref. 2 and 14) on the New Jersey Strataform.[33] A moored vertical source array centered at a depth of 47 m transmitted Tukey-shaded linear frequency modulated (LFM) pulses in three distinct frequency bands, 390–440 Hz, 875–975 Hz, and 1250–1400 Hz, each 1 s in duration. The source array consists of seven XF-4 transducer elements spaced 1.63 m apart for the lower band, and ten mod-30 elements spaced 0.8 m apart for the two higher bands. The signals are radiated azimuthally symmetric about the source array. The signals were received on a horizontal hydrophone array of which one hydrophone was desensitized with a lower gain setting, giving it a larger dynamic range for recording the one way propagated waveforms. The receiving array was towed at depths between 30 and 50 m with a speed of 2 m/s along track lines 10 km long for 30 different tracks during the days of May 5–15. On four of these tracks, the two higher bandwidths were simultaneously transmitted, while in the remaining 26 tracks, the lower frequency band was transmitted. The source transmitted pulses at every 50 s interval in which time the receiver array moved approximately 100 m.

The three distinct source locations and tracks of the receiver array for the data analyzed are overlain on the bathymetry at the experiment site shown in Fig. 1. The bathymetric variations are fairly benign, with depths ranging from 65 to 80 m. The seafloor geoacoustic parameters on the New Jersey Strataform are well characterized.[34] The seafloor is mostly sand with a mean density, sound speed, and attenuation of approximately 1.9 g/cm^3, 1700 m/s, and 0.8 dB/λ, respectively, in the region where the data were collected.[34] Internal waves randomize the sound speed profiles over time and space in the continental shelf region. During the experiment, over 100 sound speed profiles were measured at the experiment site, as shown in Fig. 2.

The effect of propagation through the random and dispersive multimodal ocean waveguide is illustrated in Fig. 3 for two typical transmissions received at ranges of 2.5 and 13 km. The signal is reduced in amplitude due to cylindrical spreading loss and attenuation in the sea bottom. The signal envelope, MF signal, and spectral shape are also substantially modified. The MF signal illustrates the dispersion of

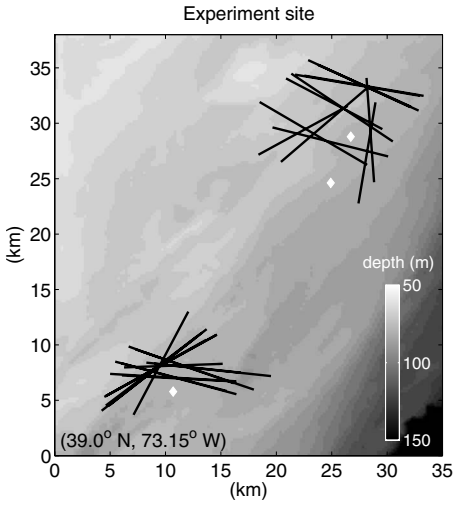


FIG. 1. The locations of the source (white diamond) and receiver tracks (black lines) are shown overlain onto the bathymetry on the New Jersey Strataform. The bathymetry data have 30 m resolution and are obtained from high-resolution bathymetric surveys described in Ref. 33. Coordinates of the southwest corner of the image are 39.0° N, 73.15° W.

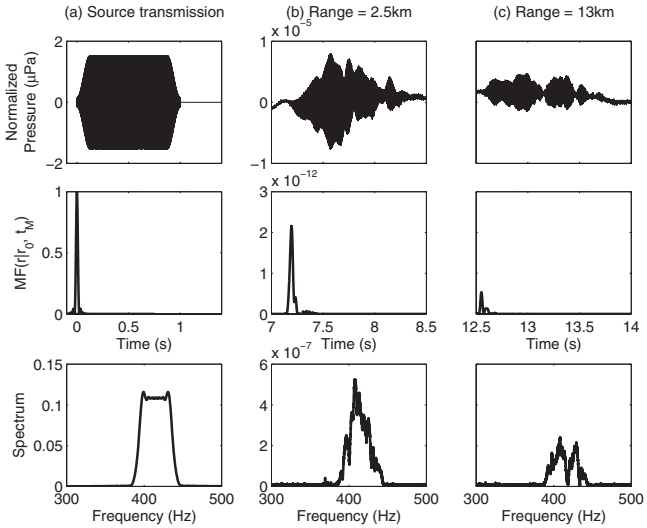


FIG. 3. (a) The transmitted signals Tukey-shaded LFM waveform is compared to the received signal at (b) ranges of 2.5 km and (c) 13 km. The signal envelope (top), the MF signal (middle), and signal spectrum (bottom) are plotted for comparison.

the multipath arrivals; there is one large initial peak formed by low-order modes as well as smaller peaks later in time from higher-order modes.

B. Data processing

Let $\Psi(\boldsymbol{r}|\boldsymbol{r}_o,t)$ be the received pressure in time t at receiver location $\boldsymbol{r}$ from a source at $\boldsymbol{r}_o$ with complex spectral amplitude $\Phi(\boldsymbol{r}|\boldsymbol{r}_o,f)$ for frequency f obtained by Fourier transform analysis,

$$\Phi(\boldsymbol{r}|\boldsymbol{r}_o,f) = \int_T \Psi(\boldsymbol{r}|\boldsymbol{r}_o,t)\exp(-j2\pi ft)dt, \tag{1}$$

where T is the time window used to isolate the direct arrival[35] from reverberation and other noise sources. T is chosen to be 2 s, including 0.5 s before the initial arrival of the signal and 0.5 s after the 1 s duration signal to sufficiently capture the signal and dispersion of the late arrivals.

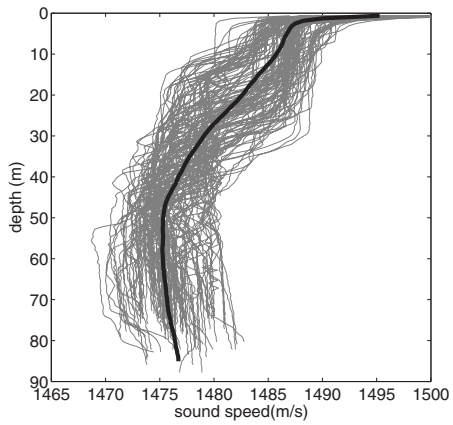


FIG. 2. Over 100 sound speed profiles were experimentally collected during the Main Acoustic Clutter Experiment (2003) on the New Jersey continental shelf by deploying both conductivity-temperature-depth sensors and expendable bathythermographs.

Integrating the magnitude squared spectrum over the signal bandwidth $B=f_f-f_i$, we obtain

$$E_{\mathrm{PS}}(\boldsymbol{r}|\boldsymbol{r}_o) = \int_{f_i}^{f_f} |\Phi(\boldsymbol{r}|\boldsymbol{r}_o,f)|^2 df. \tag{2}$$

By Parseval's theorem $E_{\mathrm{PS}}(\boldsymbol{r}|\boldsymbol{r}_o)$ is proportional to the signal energy. Scattered reverberation and noise integrated into the PS occuring within the time window and frequency band is considered to be negligible in comparison to the direct signal from the source. Figure 4 shows the PS energy for all transmissions in a single track as a function of integration time starting 1 s prior to the pulse arrival. Here we see that increasing the time window beyond the signal duration to in-

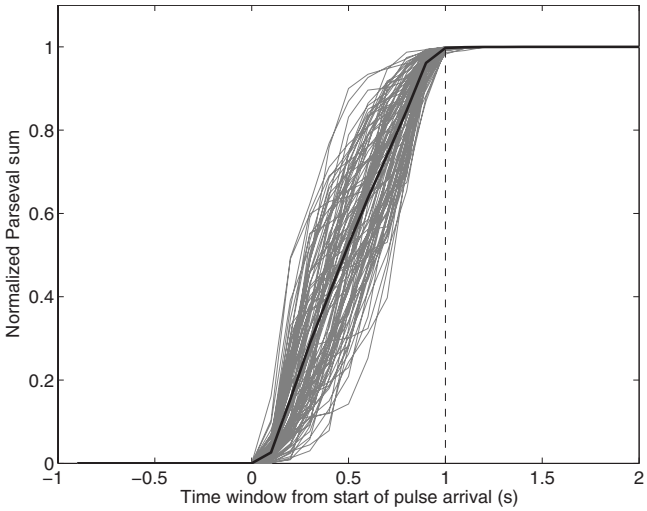


FIG. 4. The normalized PS for all transmissions in a track are illustrated as a function of the integration time window. The time window is measured from 1 s prior to the pulse arrival and increased it to include the 1 s duration signal as well as 1 s afterward to measure the effects of late multipath arrivals, noise, and reverberation on the PS value. Here we see that the entire PS energy is well contained within the 1 s signal duration and therefore has negligible contributions from reverberation and noise.

clude reverberation and noise has negligible effect.

For sonar applications, our primary interest is the MF energy, where the filter is a replica of the original transmitted signal $q(t)$. The normalized MF is given by[19,26] $h(t|t_M)=Kq(t_M-t)$, and its Fourier transform is $H(f|t_M)=KQ^*(f)\exp(-j2\pi f t_M)$, where t_M is the delay time of the MF, $Q(f)$ is the Fourier transform of the source signal, and $K=(\int_{f_i}^{f_f}|Q(f)|^2df)^{-1/2}$ is the normalizing factor. The MF output is then

$$\begin{aligned}\mathrm{MF}(\boldsymbol{r}|\boldsymbol{r}_o,t_M) &= |h(t|t_M) * \Psi(\boldsymbol{r}|\boldsymbol{r}_o,t)|^2 \\ &= \left|\int_{f_i}^{f_f}\Phi(\boldsymbol{r}|\boldsymbol{r}_o,f)H(f|t_M)\exp(j2\pi ft)df\right|^2 \\ &= \left|\int_{f_i}^{f_f}\Phi(\boldsymbol{r}|\boldsymbol{r}_o,f)KQ^*(f)\exp(j2\pi f(t-t_M))df\right|^2. \end{aligned} \tag{3}$$

As was shown in Fig. 3, the MF output has one strong peak due to the direct arrival of the signal. Later multipath arrivals may be present if they can be resolved from the primary peak. To obtain the MF energy contained in the primary arrival, we take the maximum value of Eq. (3),

$$E_{\mathrm{MF}}(\boldsymbol{r}|\boldsymbol{r}_o) = \mathrm{MF}(\boldsymbol{r}|\boldsymbol{r}_o,t_M)|_{\max}. \tag{4}$$

In unbounded free-space media, this maximum occurs when $t_M=|\boldsymbol{r}-\boldsymbol{r}_o|/c$, where c is the sound speed, but in a dispersive waveguide there may be some time delay. For discrete targets, the scattered signal's peak MF output is dominant and often the only peak that stands above background reverberation and ambient noise.

For a LFM signal, the intensity at a single frequency, for instance, the center frequency f_c, can be approximated as an *instantaneous intensity* when the LFM signal sweeps through the signal bandwidth over the signal duration. The pressure magnitude squared spectrum bandlimited to 1 Hz at the center frequency, $|\Phi(\boldsymbol{r}|\boldsymbol{r}_o,f_c)|^2$, multiplied by the signal bandwidth,

$$E_{\mathrm{CF}}(\boldsymbol{r}|\boldsymbol{r}_o,f_c) = |\Phi(\boldsymbol{r}|\boldsymbol{r}_o,f_c)|^2B, \tag{5}$$

then has the same units as the PS and MF energies and can be used for direct statistical comparison.

Taking $10\log_{10}$ of Eqs. (2), (4), and (5) gives the PS, MF, and CF energy levels in decibel units and we denote these by L_{PS}, L_{MF}, and L_{CF}, respectively. The MF degradation is defined here as

$$D_{\mathrm{MF}}(\boldsymbol{r}|\boldsymbol{r}_o) = L_{\mathrm{PS}}(\boldsymbol{r}|\boldsymbol{r}_o) - L_{\mathrm{MF}}(\boldsymbol{r}|\boldsymbol{r}_o). \tag{6}$$

C. Normal mode modeling of matched filtered signal

Here, we show that a significant portion of the signal's MF energy is contained within the initial direct arrival peak. A normal mode propagation model[36] is used to predict modal arrivals of the broadband transmissions. The model determines the acoustic field $\Phi(\boldsymbol{r}|\boldsymbol{r}_o,f)$ for a range-independent shallow water waveguide 80 m deep for all frequencies over the signal bandwidth by coherently combining the contributions from the individual source elements centered at 42 m depth. Applying Eq. (3) provides the modeled MF intensity. Seafloor geoacoustic parameters described in Sec. II A and the mean sound speed profile illustrated in Fig. 2 are used to simulate the New Jersey shelf environment.

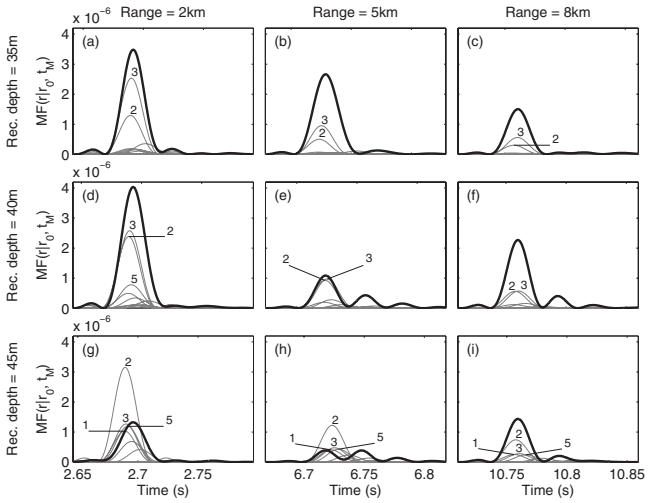


FIG. 5. The Kraken normal mode model is used to simulate the MF signal and compare modal contributions for a receiver at ranges of 2, 5, and 8 km, and depths of 35, 40, and 45 m. The magnitude squared MF output for each mode is shown in gray with the most prominent modes labeled. The modes can combine either constructively or destructively to produce the total matched filtered intensity (black).

Figure 5 illustrates the modeled MF signal for receiver ranges of 2, 5, and 8 km and receiver depths of 35, 40, and 45 m, which were typical during the experiment. The modal contributions for the first ten modes in the ocean waveguide are shown in gray with the dominant mode numbers labeled. In each case, the first four modes are unresolved by the primary signal peak, interfering either constructively or destructively to result in the MF intensity, shown in black. Later modal arrivals generally produce small secondary peaks in the MF signal.

While the normal mode model does not incorporate the effects of range-dependent bathymetry and fluctuating sound speed profiles caused by internal waves, it is still useful in determining which modes are dominant, as well as their relative arrival times. Fluctuations from internal waves may increase the modal dispersion effects, and mode coupling may spread the temporal arrivals from particular modes slightly.[4,7] However, we see that the modal dispersion modeled here is consistent with that seen in the MF data of Fig. 3. For data collected on a single hydrophone, it is impossible to resolve individual modal contributions as is modeled here. At the ranges of interest, we can see that the signal energy is mostly contained within the initial peak.

D. Mean, standard deviation, and number of independent statistical fluctuations

The levels L_{PS}, L_{MF}, and L_{CF} are calculated for each received signal from all 30 tracks. The signal path and corresponding bathymetry from source to receiver for each measurement vary depending on source location, track, and receiver position on the track. The figures will illustrate the

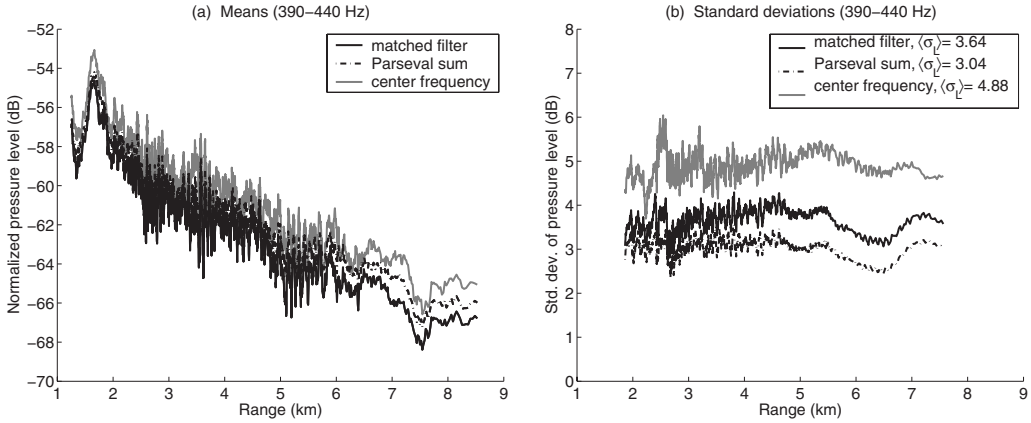


FIG. 6. The one-way propagated broadband transmissions at 390–440 Hz are processed and normalized for a 0 dB re 1 μPa at 1 m source level. The (a) means and (b) standard deviations are calculated using a 2 km running window in range for the MF, PS, and the bandpassed center frequency energies.

statistics of the lowest band as it is the most plentiful dataset. Results from the higher bands are tabulated. The means and standard deviations for the various measurements taken over a running window of 20 pings along each track are illustrated in Fig. 6 for the 390–440 Hz band. The observed fluctuations in the data are due to both temporal and spatial variations in the medium. The source level fluctuations are small compared to those caused by the environment. The time interval for each 20-ping sequence is 16 min and change in receiver location is 2 km. This distance is greater than the spatial coherence length of approximately 500 m for internal waves in shallow water.[4,37] The 20-ping averaging is therefore large enough to sample our data over independent fluctuations of the environment caused by internal waves, while still providing a small enough window to observe trends in the means and standard deviations over range. The means and standard deviations from each track are then averaged together. The means are calculated from antilog data and then converted to decibel levels for plotting. This is to avoid potential bias in the mean of log-transformed quantities.[15] The standard deviations are calculated directly from the log-transformed data since they are expected to have uniform standard deviations.[15]

The means plotted in Fig. 6(a) illustrate the general intensity decay with range due to spreading, absorption, and scattering losses in the waveguide. Modal interference effects are significant for source-receiver separations less than 2 km. The standard deviations in Fig. 6(b) are largest for L_{CF} and smallest for L_{PS}. The PS measurement integrates over statistically independent fluctuations within the signal duration and bandwidth which reduces its standard deviation. The MF also utilizes the entire signal duration and bandwidth, but has a slightly higher standard deviation than the PS because the MF degradation is not constant and introduces its own fluctuation.

Using coherence theory, the energy level standard deviation σ_L is related to the number of independent statistical fluctuations or coherence cells μ (Refs. 15 and 16) of the received signal,

$$\sigma_L = (10 \log_{10} e) \sqrt{\sum_{k=0}^{\infty} \frac{1}{(\mu + k)^2}}. \tag{7}$$

Using the experimentally determined standard deviations as an input to Eq. (7), the corresponding μ for the levels L_{CF}, L_{PS}, and L_{MF} are calculated and summarized in Table I for all three bands. The level L_{CF} has a standard deviation close to 5 dB for each band, which corresponds to $\mu \approx 1.2$. The PS for the 50 Hz bandwidth signal from 390 to 440 Hz, for instance, integrates over approximately $\mu = 2.5$ independent statistical fluctuations over the signal duration and spectrum. This is an order of magnitude smaller than the deterministic transmitted signal's own time-bandwidth product of $TB = 50$ for $T = 1$ s and $B = 50$ Hz. This implies that the signal is remaining largely correlated across the bandwidth after propagation through the random waveguide.

The number of independent statistical fluctuations μ of the received signal depends on the relative bandwidth B/f_c of the transmitted signal for signals with a fixed time duration, smaller than its statistical decorrelation time scale. The standard deviation and μ are plotted as a function of relative bandwidth in Figs. 7(a) and 7(b) for the PS energy. We ob-

TABLE I. Standard deviations σ_L of the measurements for each bandwidth are obtained from data analysis described in Sec. II D. The number of independent statistical fluctuations μ are calculated from Eq. (7).

	390–440 Hz		875–975 Hz		1250–1400 Hz	
	σ_L (dB)	μ	σ_L (dB)	μ	σ_L (dB)	μ
Center frequency	4.98	1.16	4.92	1.18	4.98	1.16
Parseval sum	2.99	2.54	3.62	1.86	3.57	1.91
Matched filter	3.64	1.85	4.35	1.41	4.19	1.49
Relative bandwidth B/f_c	0.121		0.108		0.113	

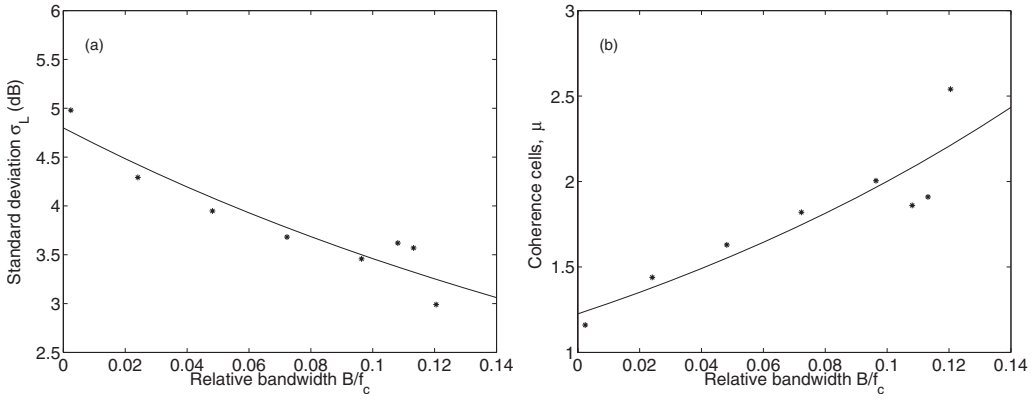


FIG. 7. The PS energy level (a) standard deviations σ_L, and (b) number of coherence cells μ are plotted as a function of relative bandwidth from each of the three bands. The results at other B/f_c ratios are obtained by bandpassing the 390–440 Hz data and performing similar statistical analysis. The equation describing the monotonic fit for (b) is $\mu=A\exp(kB/f_c)$, where parameters $A=1.23$ and $k=4.9$. The curve shown in (a) can be obtained from (b) with Eq. (7).

serve the energy standard deviation decreases monotonically with the relative bandwidth, while μ increases monotonically.

The rate of increase in μ with relative bandwidth obtained here from experimental data is much smaller than that obtained in Ref. 8 based on theoretical modeling. There, the number of coherence cells was found to increase from 1 to 20 as the signal bandwidth increased from 0 to 200 Hz for a 1 kHz signal propagating in a waveguide with 100 m water depth at 15 km range. Figure 7(b) shows that this is physically not achievable unless the statistical decorrelation time of the signal is much shorter than the signal duration in that environment. Extrapolating the results in Fig. 7(b), we would obtain only about three coherence cells for a signal with $B/f_c=0.2$ when its duration is roughly 1 s on the New Jersey shelf, where the acoustic field decorrelation time scale in this waveguide is about 5 min.[38,39]

When a waveguide is randomized, the multimodal interference pattern is randomized in both space and time. Figure 7(a) suggests that even though the modes are randomized, the intensity fluctuation for the broadband signal are roughly bounded by the peak and trough of intensity variation with range and depth in a nonrandom waveguide. This is expected to be the case when the fluctuations are caused by linear internal waves. When nonlinear internal waves or solitons are present, then much larger standard deviations may be obtained.

The relationship derived in Fig. 7(a) between the standard deviation and relative bandwidth can now be used in experimental design to aid in selecting the appropriate frequency bandwidth for the acoustic transmissions in order to achieve the desired standard deviation in the received signal energy.

E. Statistical distributions

Histograms for the 390–440 Hz bandpassed center frequency data E_{CF}, the broadband PS E_{PS}, and MF energies E_{MF} are shown in Figs. 8–10. The histograms illustrate the distribution of measurements separated in range increments of 500 m. The center frequency data follow the exponential distribution, which has nonzero probability density for $E=0$, as shown in Fig. 8. This implies that the complex spectral amplitude $\Phi(\boldsymbol{r}|\boldsymbol{r}_o,f_c)$ can be treated as a circular complex Gaussian random variable.[16] This allows us to approximate the E_{CF} measurements as being fully randomized.

The broadband data follow the Gamma distribution[15] parametrized by the number of coherence cells,

$$P(E(r))=\frac{(\mu/\overline{E(r)})^{\mu}E(r)^{\mu-1}\exp(-\mu E(r)/\overline{E(r)})}{\Gamma(\mu)}, \tag{8}$$

where $\Gamma(\mu)$ is the Gamma function, and $\overline{E(r)}$ is the mean broadband energy at a given range r from the source. The theoretical Gamma distribution of Eq. (8) is plotted over each histogram in Figs. 8–10 using the calibrated results for mean energy $\overline{E(r)}$ from Sec. III B and μ from Sec. II D. For the bandpassed center frequency data, the theoretical exponential distribution is a special case of the Gamma distribution where $\mu=1$. For each case, the theoretical statistical distributions, $P(E(r))$, at six different ranges are each calculated using the range value centered within the range window used for each histogram. The distributions change as the ex-

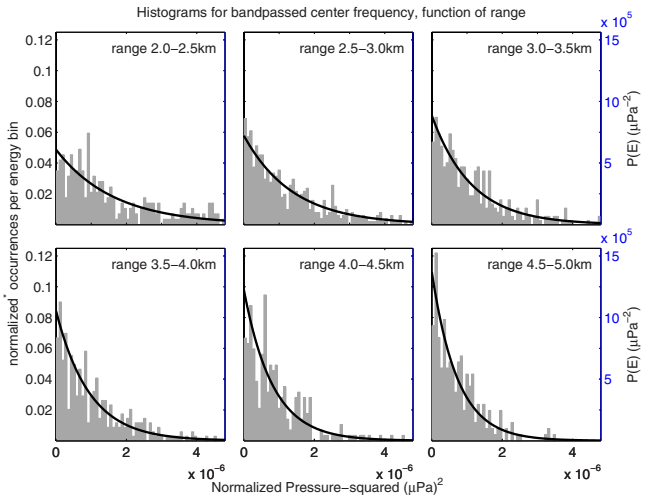


FIG. 8. (Color online) Histograms of the received center frequency energies are separated by range in 500 m increments and plotted against the calibrated model energy distribution. The histograms are normalized by dividing the number of occurrences in each energy bin with the total number of energy measurements at the specified range. The right axis for the theoretical distribution is scaled to provide direct visual comparison of the theoretical distribution with the histogram.

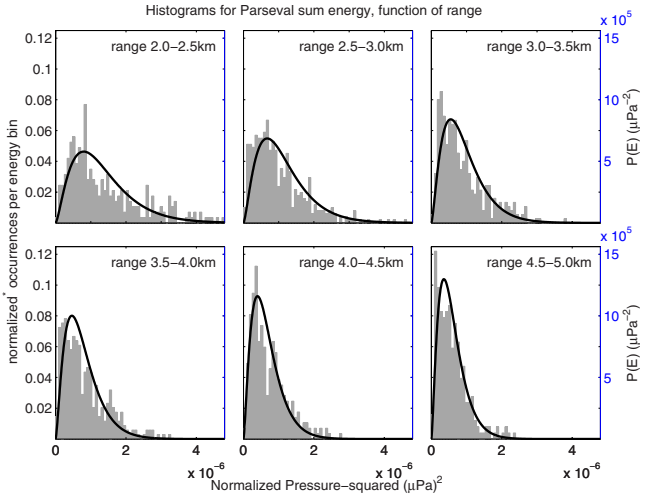


FIG. 9. (Color online) Histograms of the received broadband PS energies, plotted similar to Fig. 5.

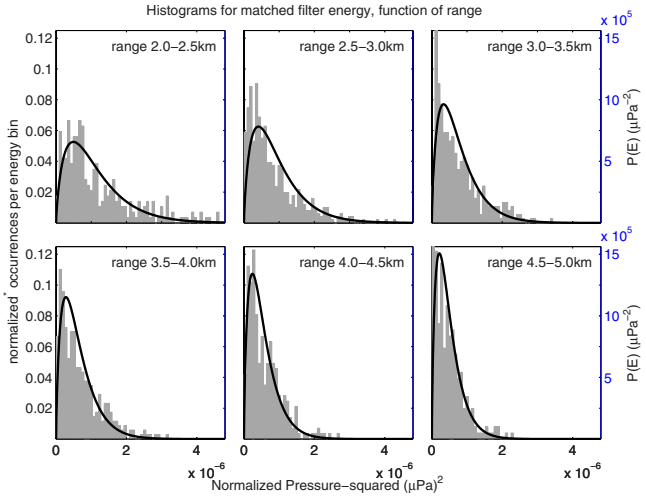


FIG. 10. (Color online) Histograms of the received broadband MF energies, plotted similar to Fig. 5.

pected broadband energy $\overline{E(r)}$ decreases with range. For all three measurements, the histograms closely match the theoretical distributions.

The statistical distributions for the data here are similar to those found in Ref. 9 for the same area. In that experiment, broadband 100 Hz bandwidth signals centered at 400 Hz were transmitted along two different propagation paths of 42 and 60 km. The intensities were found to scintillate according to similar types of distributions. At those ranges, the peak MF energy level in decibels approximated a log-normal distribution, which is a special case of the distribution expressed in Eq. (8) for a large time-bandwidth product. The distribution of the point intensity closely followed an exponential distribution, approximating an instantaneous measurement of a statistically saturated acoustic field. Here, the distributions are plotted on a linear scale and the histograms are grouped over range intervals to show how the distributions vary as a function of range. Also, the distributions here use data collected at different locations in space and are therefore functions of scintillation over both time and space.

F. Matched filter degradation

The 20-ping averaged MF degradation in decibels, given by the difference between the mean PS and MF levels, can be estimated from Fig. 6(a) for the 390–440 Hz data. In addition to modal dispersion, this difference may also be caused by ambient noise or reverberation integrated into the PS measurement. These effects are negligible for our signals within the analysis time T. The MF degradation for the individual transmitted signals plotted in Fig. 11(a) shows significant fluctuation. The 20-ping averaged mean degradation in Fig. 11(b) illustrates an approximate linear increase with range. The range-dependent MF degradation D_{MF} can be approximated using the calibrated results from Sec. III B through the linear relationship

$$D_{\mathrm{MF}}(\rho) = c + m\rho, \tag{9}$$

where c is an initial offset at the onset of modes and $m\rho$ accounts for cumulative dispersion in the waveguide. For these 390–440 Hz data, $c=0.7$ dB and $m=0.05$ dB/km, as shown in Table II. The latter linear decibel increase in degradation with range corresponds to an exponential decay of the MF energy with range when compared to the PS energy. The mean MF degradation in Fig. 11 is biased slightly above the calibrated MF degradation because the mean is taken over log values resulting in a bias that depends on the variance of the MF degradation.

G. Monte Carlo simulations for broadband acoustic transmissions through a random waveguide for field statistics

In this section, Monte Carlo simulations are used to simulate sound speed fluctuations in a random waveguide and propagate the broadband acoustic signals through the waveguide. We show that the statistics of simulated acoustic transmissions through the dynamic environment are consistent with the observed statistics of the data in Sec. II D.

Internal waves randomize the ocean by causing fluctuations in water-column temperature and sound speed profiles over both space and time by adiabatic vertical displacements. This phenomenon has been shown to be the primary source for acoustic field randomization in continental shelf environments.[4,6,35,37,40–42] More recent work has furthered the understanding of intensity fluctuations caused by both linear diffuse internal waves,[43] which create a gradual continuous fluctuation in the sound speed structure, and nonlinear internal waves[4–7] or solitons, which create more discontinuous sound speed fluctuations in range. Both the 1995 SWARM and the 1996 PRIMER experiments, conducted on the New Jersey shelf, found that linear internal waves cause broadband intensity scintillation while nonlinear solitons further randomize and attenuate the intensity through mode coupling and mode stripping.[4–7]

Creating an accurate model for both the linear diffuse internal waves and the nonlinear solitons is challenging with-

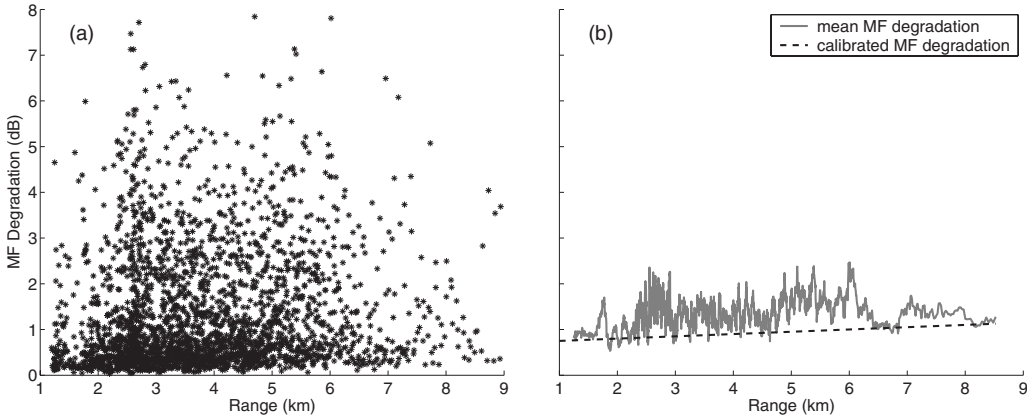


FIG. 11. The MF degradation in the 390–440 Hz data is illustrated here for (a) each received signal and (b) using a running average over 20 transmissions. The average is compared to the calibrated degradation.

out highly detailed environmental data such as real-time sound speed and temperature profiles measured continuously in time and space. The SWARM and PRIMER experiments observed both linear and nonlinear internal waves in the same region as this experiment. Both were conducted in July and August when the internal wave activity is the strongest and most prevalent[5] while the MAE was conducted in April–May when the intensity and prevalence of internal waves are weaker. The experimentally measured sound speed profiles from the New Jersey continental shelf during April–May 2003 shown in Fig. 2 provide a good sample of the profiles measured at different locations and times on the continental shelf; however, they are too aliased to observe specific internal waves.

Given these limitations, a simpler approach is taken for modeling the environment. We assume the statistics of the sound speed profiles are stationary over the region of the experiment. Linear internal waves are simulated by updating the sound speed profiles in range every 500 m, chosen as an approximate correlation length observed for continental shelf environments.[4,37] The profiles are selected randomly from the experimentally collected sound speed profiles. This approach is taken because it is impossible to predict the direction, amplitude, and frequency of specific internal waves. Instead, a sufficient sample of the observed sound speed profiles are used to simulate the *effects* of the internal waves after many independent simulations. Discontinuous solitons are not included in the model because their presence cannot be confirmed between the source and the receiver for any given transmission. The linear diffuse internal waves are largely responsible for the scintillation in the broadband intensity interference pattern and are therefore sufficient for the first order statistics such as the mean and standard deviations of the intensities. This model does not describe the temporal coherence of the scintillating acoustic field or intensity because the model relies only on independent realizations of the environment. The temporal coherence of the acoustic intensity could be calculated if the temporal fluctuations of the environment caused by internal waves were included into the model.

The acoustic field through this fluctuating environment is modeled using the RAM,[32] which takes into account the environmental parameters including sea-bottom geoacoustic properties, source and receiver locations, varying bathymetry between source and receiver, and the dynamic sound speed profile. For each realization, it produces the time-harmonic Green's function $G(\boldsymbol{r}|\boldsymbol{r}_0,f)$ in space at frequency f from which we can determine the complex spectral amplitude,

$$\Phi_m(\boldsymbol{r}|\boldsymbol{r}_o,f) = Q(f)G(\boldsymbol{r}|\boldsymbol{r}_0,f), \tag{10}$$

where the subscript m indicates that it is a modeled output. Here, $Q(f)$ is the Fourier transform for a synthesized source signal. Applying the RAM over the source bandwidth fol-

TABLE II. The results of calibration for source level correction c_{SL} and TL decay rate correction, α, are shown here for both PS and MF energy over three frequency bands. The MF degradation is found as the difference of the PS value and MF value for both source level and range-dependent degradation. For the two higher frequency bands, there were fewer transmissions occurring over a smaller window in range for which to calibrate the expected levels. This results in higher error bounds for the estimates of c_{SL} and α. For the highest frequency band, the transmission data across the array for each transmission are highly correlated, so the error bounds calculated for the 875–975 Hz band are used as an upper limit for the error bounds.

	390–440 Hz		875–975 Hz		1200–1400 Hz	
	c_{SL} (dB)	α (dB/km)	c_{SL} (dB)	α (dB/km)	c_{SL} (dB)	α (dB/km)
Parseval sum	-0.8 ± 0.2	-0.03 ± 0.04	-1.1 ± 0.6	-0.09 ± 0.20	0.8 ± 0.6	-0.06 ± 0.20
Matched filter	-1.5 ± 0.2	-0.08 ± 0.04	-1.7 ± 0.8	-0.26 ± 0.20	-0.8 ± 0.8	-0.08 ± 0.20
MF degradation	0.7	0.05	0.6	0.17	1.6	0.02

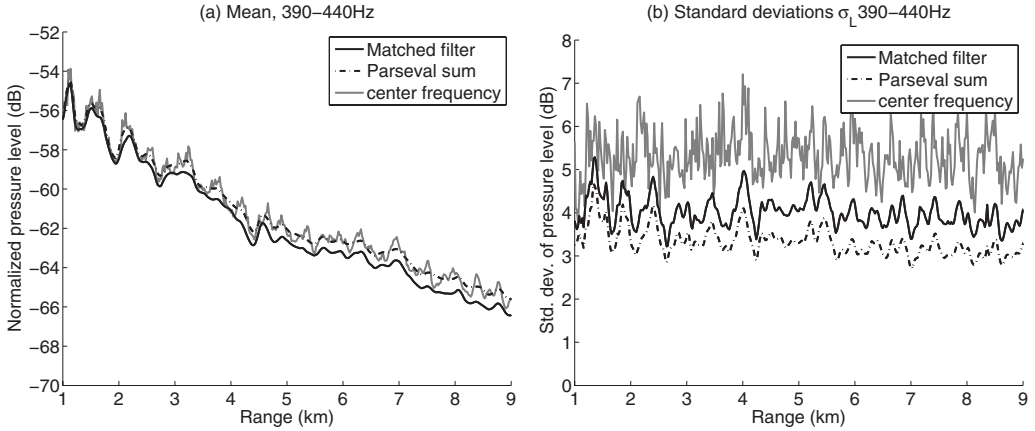


FIG. 12. The (a) mean and (b) standard deviations for the 390–440 Hz MF, PS, and center frequency energies modeled using Monte Carlo simulations in a waveguide randomized by internal waves. The statistics of the model match those of the data in Fig. 6.

lowing Eqs. (2) and (4) provides the modeled PS, $E_{\mathrm{PS},m}$ and MF energies, $E_{\mathrm{MF},m}$. For example the modeled PS energy is expressed as

$$E_{\mathrm{PS},m}(\boldsymbol{r}|\boldsymbol{r}_0) = \int_{f_i}^{f_f} |\Phi_m(\boldsymbol{r}|\boldsymbol{r}_o,f)|^2 df, \tag{11}$$

and the modeled mean broadband PS transmission loss is $10\log_{10}$ of the received mean PS energy, averaged over all realizations, normalized by the source energy E_0,

$$\begin{aligned}\mathrm{TL}_m(\boldsymbol{r}|\boldsymbol{r}_0) &= 10\log_{10}\left(\frac{1}{E_0}\langle E_{\mathrm{PS},m}(\boldsymbol{r}|\boldsymbol{r}_0)\rangle\right)\\ &= 10\log_{10}\left(\frac{1}{E_0}\left\langle \int_{f_i}^{f_f} |Q(f)|^2 |G(\boldsymbol{r}|\boldsymbol{r}_0,f)|^2 df\right\rangle\right),\end{aligned} \tag{12}$$

where $E_0 = \int_{f_i}^{f_f} |Q(f)|^2 df$.

To illustrate the modeled field statistics, we implement this Monte Carlo model for a dynamic waveguide with a constant water depth of 85 m and a sand bottom halfspace. Figure 12 illustrates the means and standard deviations for 100 realizations of the MF, PS, and center frequency energies using a normalized 0 dB re 1 μPa source level for the 390–440 Hz band. Each of the 100 realizations of the broadband energy were computed with a spectral spacing of 1 Hz over the 50 Hz bandwidth, leading to a total of 5000 time-harmonic runs to generate the expected PS and MF energies.

The statistics derived from the model closely match those of the data in Fig. 6. For instance, the model PS energy standard deviation is the smallest, while that for the center frequency is the largest, consistent with the data. This implies that this approach can be applied to reliably model the acoustic field statistics on the New Jersey continental shelf.

III. COMPUTATIONALLY EFFICIENT NUMERICAL APPROACH FOR ESTIMATING MEAN BROADBAND TRANSMISSION LOSS IN A FLUCTUATING RANGE-DEPENDENT OCEAN WAVEGUIDE

In wide area sonar applications, the mean TL must be efficiently estimated in order to detrend sonar imagery and invert for scattering strength or target strength of scatterers and source level of radiating targets over vast areas.[2,14] The brute-force method described in Sec. II G for estimating the expected broadband TL is far too computationally intensive and impractical for realistic sonar operations. Wide area sonar can image circular areas with a diameter exceeding 100 km. Imaged objects include fish shoals that often occupy areas spanning more than 20×20 km^2.[2] In order to invert for areal scattering strengths in these images, the bistatic TL both from the source and to the receiver must be estimated over the entire area and updated for each image with a moving receiver. For example, a large fish shoal extending 20 km in cross range may span 60° at 20 km range from the source, requiring 60 TL transects to be estimated at 1° spacing (roughly equal to the broadside angular resolution of a wide area sonar system).[2,14] For each transect requiring 5000 time harmonic simulations, a total of 300 000 runs are required to estimate the mean broadband TL and subsequently the areal scattering strengths throughout the shoal.

Here we propose a numerically efficient approach for estimating the expected broadband energy and mean TL through a fluctuating range-dependent ocean waveguide. We show that the mean broadband energy can be estimated to within a 0.5 dB standard deviation using three Monte Carlo simulations at only the center frequency and by introducing range and depth averaging for each radial. This significantly reduces the number of time-harmonic Monte Carlo runs from 300 000 to just 180 for the 60° span of the sonar imagery described above. These 180 runs can be computed within minutes to allow near real-time inversion of wide-area sonar imagery for fish areal scattering strengths and population density estimates.[2]

A. Approximating mean broadband transmission loss

We illustrate the computationally efficient approach for estimating the mean broadband TL in a fluctuating ocean waveguide for two different waveguides with sandy bottoms in Figs. 13(a) and 13(b). The first has a constant water depth of 85 m and the second has an upslope bathymetry, as shown in Fig. 13(d). For comparison, the 100 Monte Carlo realiza-

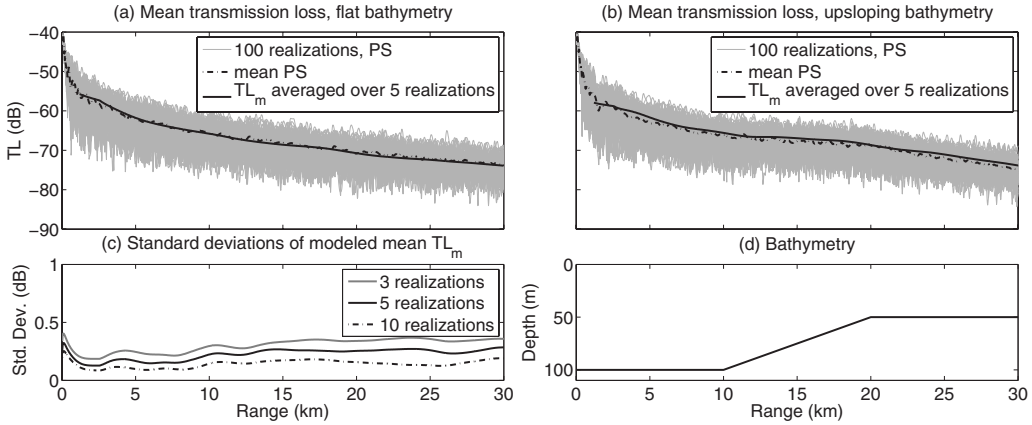


FIG. 13. The mean broadband PS TL over 100 realizations of the Monte Carlo model are compared to the approximate model that uses the center frequency TL found by averaging over the first five realizations and using a spatial average over both range and depth. The results are illustrated for environments with (a) constant and (b) upsloping bathymetry where the bathymetry is plotted in (d). In (c) the standard deviations for this modeled mean TL are shown as a function of range averaging over 3, 5, and 10 Monte Carlo realizations.

tions of the full broadband PS TL are plotted along with the ensemble average over these 100 realizations for both waveguides.

We next approximate the mean broadband TL as

$$\mathrm{TL}_{m,N}(\boldsymbol{r}|\boldsymbol{r}_0) \approx 10\log_{10}(I_{m,N}(\boldsymbol{r}|\boldsymbol{r}_0,f_c)) = 10\log_{10}\left\{\frac{1}{N}\sum_{p=1}^{N}\overline{|G_p(\boldsymbol{r}|\boldsymbol{r}_0,f_c)|^2}\right\}, \quad (13)$$

which is obtained by averaging over N, only a few independent realizations of the acoustic field at the center frequency with the overline indicating spatial averaging of the Green's function magnitude square over both depth and range. Here, the magnitude squared Green's function output of the RAM is averaged about the receiver over ± 20 m in depth and in range using a range dependent Hanning window with an extent given by $1.5\log(r/(1\ \mathrm{km}))$ km. We first average the model over $N=5$ independent realizations to reduce the standard deviation of the approximate mean TL to less than 0.5 dB. At ranges greater than 2 km, this approximate $\mathrm{TL}_{m,5}$ follows a monotonic decay and matches the averaged broadband PS TL from 100 Monte Carlo realizations (with 5000 time-harmonic runs) to within 0.5 dB, as shown in Figs. 13(a) and 13(b).

Figure 13(c) illustrates how averaging over different numbers of independent Monte Carlo realizations for the environment with constant bathymetry reduces the standard deviation of the approximate TL model. In this case, the standard deviation is still less than 0.5 dB out to 30 km range even when using only $N=3$ realizations.

The spatial window used for depth and range averaging is chosen to be large enough to subtend at least one modal cycle, but not too large as to average out the attenuation in range or the effects caused by channeling of the acoustic energy in the water column. This spatial averaging bears some resemblance to the range averaging used by Harrison and Harrison,[44] where range averaging is used to replace a frequency averaging over the bandwidth in a *deterministic* environment. Here, both range and depth averaging are applied in order to estimate the *mean* transmission loss in a *random* environment. Averaging the intensity over depth has a similar effect as averaging over independent realizations of the environment; both reduce the variance caused by modal interference in a waveguide randomized by internal waves.

B. Calibration of broadband expected intensity model

Here, we calibrate the computationally efficient model for mean broadband TL with data from the New Jersey Strataform. The PS energy level differs from the TL by the source level, $L_{\mathrm{PS}}=\mathrm{SL}+\mathrm{TL}$. A maximum likelihood estimator (MLE) is employed to correct the modeled expected broadband intensity for any inaccuracies in source level and the waveguide attenuation decay rate.

In long range sonar applications, the source level is determined by the coherent interaction of multiple source elements in the far field of the array. While the individual transducer elements have been calibrated by the manufacturer in tanks, the resultant output source power still needs to be calibrated in the field since it is dependent on the number of elements operating at any given time and the input power applied to each element.

The mean TL decay trend, discussed in Sec. III A, depends not only on geometrical spreading loss and absorption into the seafloor, but also on absorption and scattering losses of the ocean environment from bubble clouds, nonlinear solitons, fish, and other scatterers in the water column. These parameters may not be completely characterized in the propagation model. We introduce an attenuation correction coefficient a to the model to account for incomplete knowledge of all absorption and scattering loss mechanisms in the environment,

$$\overline{E_m(\boldsymbol{r}|\boldsymbol{r}_0)} \approx E_0 I_{m,N}(\boldsymbol{r}|\boldsymbol{r}_0,f_c)\exp(2a|\boldsymbol{\rho}-\boldsymbol{\rho}_0|), \quad (14)$$

where $|\boldsymbol{\rho}-\boldsymbol{\rho}_0|$ is the horizontal separation between the source and the receiver. Equation (14) is a good approximation for waveguides at ranges where the field is dominated by low order modes that propagate at elevation angles close to the horizontal where the attenuation is approximately constant over the modes.

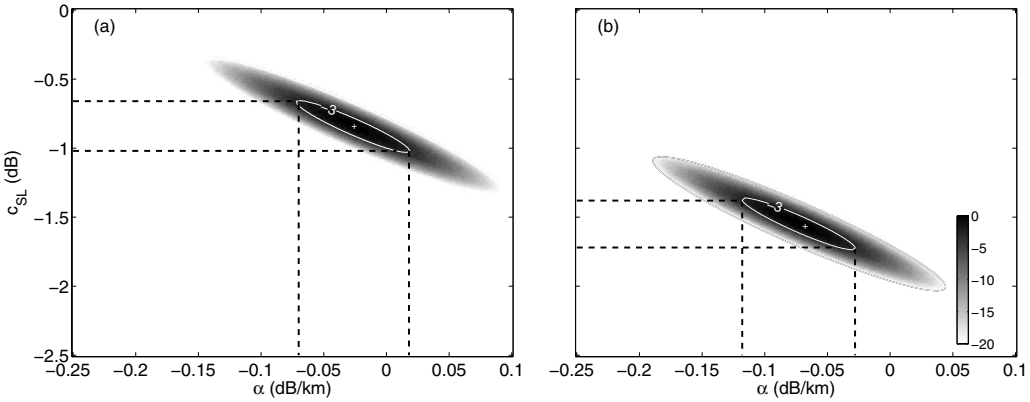


FIG. 14. The normalized log-likelihood function shown here for both (a) PS and (b) MF energy levels to calibrate for both source level and the range-dependent attenuation correction.

This model is applied to estimate the expected intensity level over range for the MAE. A MLE estimates parameters a and E_0 by comparing the broadband data from Sec. II D with the modeled mean intensity of Eq. (14). These parameters can then be applied, respectively, as corrections to the mean TL decay with range $\alpha|\boldsymbol{\rho}-\boldsymbol{\rho}_0|$, where $\alpha=20\log_{10}(e)a$, and correction to the source level, $c_{SL}=10\log_{10}(E_0/E_{0(\text{uncorrected})})$, where $E_{0(\text{uncorrected})}$ is an uncalibrated nominal value for source energy which may be obtained, for instance, from the transducer manufacturer's specification. Note that a calibration relying entirely on data that does not employ an acoustic propagation model would fail to take into account first order changes in the TL such as focusing effects caused by bathymetric variations along different transects in range-dependent environments, as well as changes in mean water-column sound speed and bottom geoacoustic parameters.

We assume that the received broadband signal from each transmission is statistically independent. The joint probability density function for the N received energy measurements from Eq. (8) is

$$p(E(r_1),E(r_2),\ldots,E(r_N)) = \prod_{i=1}^{N} \frac{(\mu/\overline{E(r_i)})^{\mu}E(r_i)^{\mu-1}\exp(-\mu E(r_i)/\overline{E(r_i)})}{\Gamma(\mu)}, \quad (15)$$

where $E(r_i)$ is the broadband intensity measured for the ith received signal with source-receiver separation r_i. The corresponding expected intensity $\overline{E(r_i)}$ is obtained from Eq. (14) incorporating the modeled expected TL for each transmission with source-receiver separation of r_i. Taking $10\log_{10}$ of Eq. (15) we obtain the log-likelihood function for the data. Equation (15) allows for a method to correct for unknown parameters in the model, in this case E_0 and a. The log-likelihood function depends nonlinearly on these two parameters, making it challenging to obtain a MLE for these parameters analytically. Here, the log-likelihood function is plotted onto a two dimensional ambiguity surface for E_0 and a, where the maximum value indicates the most probable value for the correction to source level and range-dependent correction to transmission loss.

The log-likelihood function is plotted for the model calibration with the 390–440 Hz PS energy data in Fig. 14(a), normalized by its maximum value. The contour line marks the −3 dB value, which indicates where the statistical likelihood is half that at the maximum. The correction to source level is found to be −0.8 dB for PS, estimated within an error bound of ±0.2 dB, and the attenuation correction α is found to be −0.03 dB/km within an error bound of ±0.04 dB/km as tabulated in Table II. The matched filtered energy data may also be calibrated using the same approach. The corresponding normalized ambiguity surface is plotted in Fig. 14(b) and the results tabulated are in Table II. The difference

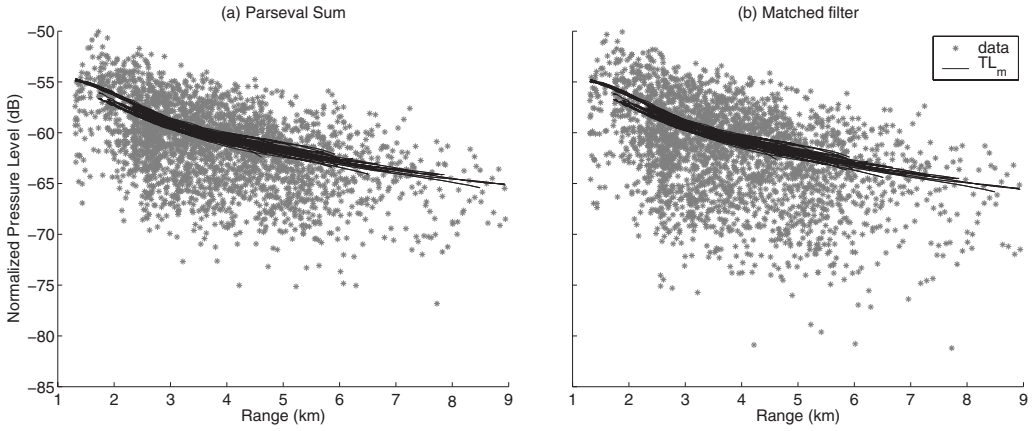


FIG. 15. The normalized expected energy level plotted over the broadband data in frequency range 390–440 Hz, measured with a single desensitized hydrophone, for the (a) PS and (b) MF.

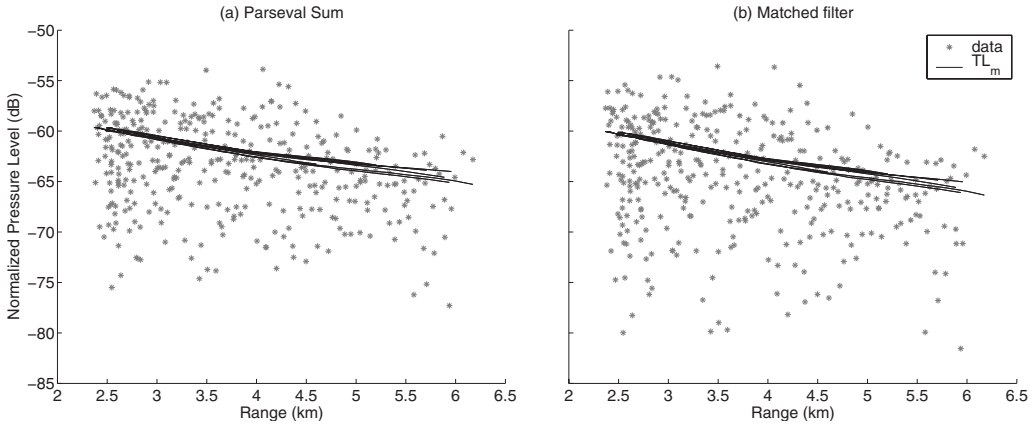


FIG. 16. The normalized expected energy level plotted over the broadband data in frequency range 875–975 Hz, measured with a single desensitized hydrophone, for the (a) PS and (b) MF.

in calibration for the PS and MF data provides a measure of the MF degradation where $c=10\log_{10}(E_{0,\mathrm{PS}}/E_{0,\mathrm{MF}})\approx 0.7$ dB and $m=4.34(a_{\mathrm{PS}}-a_{\mathrm{MF}})\approx 0.05$ dB/km.

The normalized expected pressure level from the calibrated model is plotted along with the 390–440 Hz data for both PS and MF in Figs. 15(a) and 15(b), respectively. The modeled mean level is plotted for each track of the experiment and follows the trend of the data closely at all ranges. Differences in the modeled mean levels for the different tracks occur because the modeled TL takes into account bathymetric variations, and source and receiver depth changes. The calibration results for the higher frequency bands, 875–975 Hz and 1250–1400 Hz, are provided in Table II and plotted in Figs. 16 and 17, respectively.

The proposed approach for calibrating the source level from the measured data makes a number of assumptions. (1) The source level is stationary over the time period of the experiment, and has a variance that is small compared to the variance caused by fluctuations in the environment. (2) The environmental fluctuation is assumed to follow a stationary process over both space and time of the experiment. (3) The initial assumption in Eq. (14) that the TL correction takes on the form $\exp(2a|\boldsymbol{\rho}-\boldsymbol{\rho}_0|)$ may be inaccurate when large attenuation corrections are required. It assumes that the correction in range affects each mode equally. This would not be the case if there is significant energy propagating in the higher order modes or if there is significant mode coupling and mode stripping. For the current data set, the received signal energy is mostly contained in lower order modes beyond 2 km, as discussed in Sec. II C. (4) The corrections to the mean TL should only be applied at propagation distances within the source-receiver ranges where the data were analyzed. For instance, a TL correction of −0.08 dB/km for the MF signal energy should not be extrapolated for ranges too far beyond 10 km range where the data were present.

IV. CONCLUSION

The scintillation statistics of broadband acoustic transmissions in a continental shelf waveguide have been quantified as a function of signal bandwidth B, center frequency f_c, and range. The received signal energy is shown to follow the Gamma distribution implying the instantaneous field is fully saturated from multimodal propagation contributions following the central limit theorem. The Gamma distribution depends on the mean energy, determined as a function of source-receiver separation, and the number of independent statistical fluctuations or coherent cells μ of the received signal determined from the energy standard deviations. The number of coherent cells μ is calculated for several standard

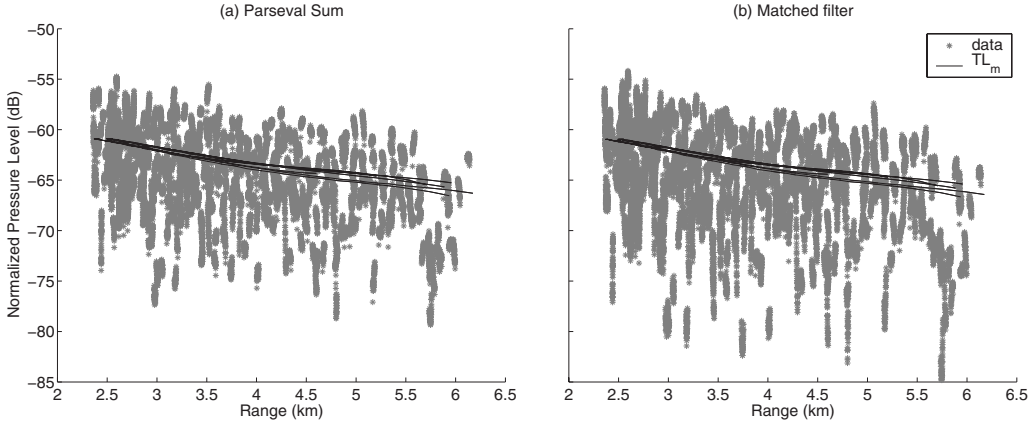


FIG. 17. The normalized expected energy level plotted over the broadband data in frequency range 1250–1400 Hz, measured with an array of 64 hydrophones for the (a) PS and (b) MF.

receivers in ocean acoustics, such as the MF, the PS and the bandpass center frequency. It is found that μ of the received signal is an order of magnitude smaller than the time-bandwidth product TB of the transmitted signal. An empirical relationship is derived showing the monotonically increasing dependence of μ on the relative bandwidth B/fc.

A computationally efficient numerical approach is presented for rapidly estimating the range-dependent mean broadband TL over wide areas of an ocean waveguide. The mean TL is estimated from Monte Carlo simulations using a range-dependent propagation model with only three to five realizations of the acoustic field in an ocean waveguide with random sound speed fluctuations. The approach uses incoherent spatial averaging over both range and depth for the time-harmonic intensity at the center frequency to estimate the mean broadband TL for a statistically saturated environment. A MLE applied to the data is used to calibrate the model for source level and the mean TL decay rate over range. This approach is indispensible for inverting wide-area sonar imagery for scattering properties of distributed targets.

ACKNOWLEDGMENTS

The authors would like express our appreciation toward Nicholas Makris for insightful discussion and clarification on the topic. This research was conducted with funding support from the Office of Naval Research Ocean Acoustics Program, the National Oceanographic Partnership Program, and the Alfred P. Sloan Foundation. This research is a contribution to the Census of Marine Life. The authors would also like to thank Northeastern University's Bernard M. Gordon Center for Subsurface Sensing and Imaging Systems for administrative support.

[1]W. Burdic, *Underwater Acoustic System Analysis*, 2nd ed. (Prentice-Hall, Englewood, NJ, 1990).
[2]N. C. Makris, P. Ratilal, D. T. Symonds, S. Jagannathan, S. Lee, and R. W. Nero, "Fish population and behavior revealed by instantaneous continental shelf-scale imaging," Science **311**, 660–663 (2006).
[3]T. Yang and W.-B. Yang, "Performance analysis of direct-sequence spread-spectrum underwater acoustic communications with low signal-to-noise-ratio input signals," J. Acoust. Soc. Am. **123**, 842–855 (2000).
[4]R. Headrick, J. F. Lynch, J. Kemp, A. Newhall, K. von der Heydt, J. R. Apel, M. Badiey, C.-S. Chiu, S. Finette, M. H. Orr, B. H. Pasewark, A. Turgot, S. Wolf, and D. Tielbuerger, "Acoustic normal-mode fluctuation statistics in the 1995 swarm internal wave scattering experiment," J. Acoust. Soc. Am. **107**, 201–220 (2000).
[5]J. R. Apel, L. A. Ostrovsky, Y. A. Stepanyants, and J. F. Lynch, "Internal solitons in the ocean and their effect on underwater sound," J. Acoust. Soc. Am. **121**, 695–722 (2007).
[6]D. Tielburger, S. Finette, and S. Wolf, "Acoustic propagation through an internal wave field in a shallow water waveguide," J. Acoust. Soc. Am. **101**, 789–807 (1997).
[7]J. Preisig and T. Duda, "Coupled acoustic mode propagation through continental-shelf internal solitary waves," IEEE J. Ocean. Eng. **22**, 256–269 (1997).
[8]S. Ramp, C. Chiu, F. Bahr, Y. Qi, P. Dahl, J. Miller, J. Lynch, R. Zhang, and J. Zhou, "The shelf-edge frontal structure in the central east China sea and its impact on low frequency acoustic propagation," IEEE J. Ocean. Eng. **29**, 1011–1031 (2004).
[9]A. Fredericks, J. A. Colosi, J. Lynch, C. Chiu, and P. Abbot, "Analysis of multipath scintillation from long range acoustic transmissions on the New England continental slope and shelf," J. Acoust. Soc. Am. **117**, 1038–1057 (2005).
[10]P. Worcester, G. Williams, and S. Flatté, "Fluctuations of resolved acoustic multipaths at short range in the ocean," J. Acoust. Soc. Am. **70**, 825–840 (1981).
[11]F. J. Beron-Vera, M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, "Ray dynamics in long range acoustic propagation experiment," J. Acoust. Soc. Am. **114**, 1226–1242 (2003).
[12]R. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1983).
[13]S. Kay, *Fundamentals of Statistical Signal Processing*, Detection Theory Vol. **2** (PTR Prentice-Hall, Englewood, NJ, 1998).
[14]A. Galinde, N. Donabed, M. Andrews, S. Lee, N. C. Makris, and P. Ratilal, "Range-dependent waveguide scattering model calibrated for bottom reverberation in a continental shelf environment," J. Acoust. Soc. Am. **123**, 1270–1281 (2008).
[15]N. C. Makris, "The effect of saturated transmission scintillation on ocean acoustic intensity measurements," J. Acoust. Soc. Am. **100**, 769–783 (1996).
[16]J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985).
[17]S. O. Rice, "A statistical analysis of random noise," Bell Syst. Tech. J. **24**, 46–108 (1945).
[18]N. C. Makris, "Parameter resolution bounds that depend on sample size," J. Acoust. Soc. Am. **99**, 2851–2861 (1996).
[19]G. L. Turin, "An introduction to matched filters," IRE Trans. Inf. Theory **IT-6**, 311–329 (1960).
[20]P. Ratilal, Y. Lai, D. T. Symonds, L. A. Ruhlmann, J. R. Preston, E. K. Scheer, M. T. Garr, C. W. Holland, J. A. Goff, and N. C. Makris, "Long range acoustic imaging of the continental shelf environment: The acoustic clutter reconnaissance experiment 2001," J. Acoust. Soc. Am. **117**, 1977–1998 (2005).
[21]D. Rouseff, "Intersymbol interference in underwater acoustic communications using time-reversal signal processing," J. Acoust. Soc. Am. **117**, 780–788 (2005).
[22]R. Barr, "A design study of an acoustic system suitable for differentiating between orange roughy and other New Zealand deep-water," J. Acoust. Soc. Am. **109**, 164–178 (2001).
[23]E. McDaid and G. Gaunaurd, "Signal processing of ideal echoes resonantly scattered by underwater structures," J. Acoust. Soc. Am. **88**, 2720–2735 (1990).
[24]D. Chu and T. Stanton, "Application of pulse compression techniques to broadband acoustic scattering by live individual zooplankton," J. Acoust. Soc. Am. **104**, 39–55 (1998).
[25]Y. Lai, "Acoustic scattering from stationary and moving targets in shallow water environments: With application of humpback whale detection and localization," Ph.D. thesis, MIT, Cambridge, MA, 2004).
[26]N. Levanon, *Radar Principles* (Wiley, New York, 1988).
[27]G. Bar-Yehoshua, "Quantifying the effect of dispersion in continental shelf sound propagation," MS thesis, MIT, Cambridge, MA, 2002).
[28]Y. H. Goh, P. Gerstoft, W. Hodgkiss, Jr., and C.-F. Huang,"Statistical estimation of transmission loss from geoacoustic inversion using a towed array," J. Acoust. Soc. Am. **122**, 2871–2579 (2007).
[29]D. Tollefsen, S. Dosso, and M. Wilmut, "Matched-field geoacoustic inversion with a horizontal array and low-level source," J. Acoust. Soc. Am. **120**, 221–230 (2006).
[30]L. Sha and L. Nolte, "Effects of environmental uncertainties on sonar detection performance prediction," J. Acoust. Soc. Am. **117**, 1942–1953 (2005).
[31]N. C. Makris, L. Z. Avelino, and R. Menis, "Deterministic reverberation from ocean ridges," J. Acoust. Soc. Am. **97**, 3547–3574 (1995).
[32]M. D. Collins, "Generalization of the split-step pade," J. Acoust. Soc. Am. **96**, 382–385 (1994).
[33]J. A. Goff, D. J. P. Swift, C. S. Duncan, L. A. Mayer, and J. Hughes-Clark, "High resolution swath sonar investigation of sand ridge, dune and ribbon morphology in the offshore environment of the New Jersey margin," Mar. Geol. **161**, 307–337 (1999).
[34]J. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P. S. Golick, and S. Nordfjord, "Seabed characterization on the new jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties," Mar. Geol. **209**, 147–172 (2004).
[35]P. Ratilal and N. C. Makris, "Mean and covariance of the forward field propagated through a stratified ocean waveguide with three-dimensional random inhomogeneities," J. Acoust. Soc. Am. **118**, 3532–3559 (2005).
[36]M. B. Porter and E. L. Reiss, "A numerical method for ocean acoustic normal modes," J. Acoust. Soc. Am. **76**, 244–252 (1984).
[37]J. R. Apel, M. Badiey, C.-S. Chiu, S. Finette, R. Headrick, J. Kemp, J. F. Lynch, A. Newhall, M. H. Orr, B. H. Pasewark, D. Tielbuerger, A. Turgut,

K. von der Heydt, and S. Wolf, "An overview of the 1995 swarm shallow-water internal wave acoustic scattering experiment," IEEE J. Ocean. Eng. **22**, 465–500 (1997).
[38]S. Finette, M. H. Orr, A. Turgut, J. R. Apel, M. Badiey, C.-S. Chiu, R. Headrick, J. Kemp, J. F. Lynch, A. Newhall, K. von der Heydt, B. H. Pasewark, S. N. Wolf, and D. Tielbuerger, "Acoustic field variability induced by time evolving internal wave fields," J. Acoust. Soc. Am. **108**, 957–972 (2000).
[39]D. Rouseff, A. Turgut, S. Wolf, S. Finette, M. Orr, B. Pasewark, J. Apel, M. Badiey, C.-S. Chiu, R. Headrick, J. Lynch, J. Kemp, A. Newhall, and D. Tielbuerger, "Coherence of acoustic modes propagating through shallow water internal waves," J. Acoust. Soc. Am. **111**, 1655–1666 (2002).
[40]J. Lynch, G. Jin, R. Pawlowicz, D. Ray, A. Plueddeman, C. Chiu, J. Miller, R. Bourke, A. Parsons, and R. Muench, "Acoustic travel-time perturbations due to shallow-water internal waves and internal tides in the Barents sea polar front: Theory and experiment," J. Acoust. Soc. Am. **99**, 803–821 (1996).
[41]T. Duda and J. Presig, "A modeling study of acoustic propagation through moving shallow water solitary wave packets," IEEE J. Ocean. Eng. **24**, 16–32 (1999).
[42]T. Chen, P. Ratilal, and N. C. Makris, "Mean and variance of the forward field propagated through three-dimensional random internal waves in a continental-shelf waveguide," J. Acoust. Soc. Am. **118**, 3560–3574 (2005).
[43]D. Creamer, "Scintillating shallow water waveguides," J. Acoust. Soc. Am. **99**, 2825–2838 (1996).
[44]C. 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).

Cooperative prey herding by the pelagic dolphin, *Stenella longirostris*

Kelly J. Benoit-Bird[a)]
College of Oceanic and Atmospheric Sciences, Oregon State University, 104 COAS Administration Building, Corvallis, Oregon 97330

Whitlow W. L. Au
Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

(Received 7 January 2008; revised 20 June 2008; accepted 23 June 2008)

Sonar techniques were used to quantitatively observe foraging predators and their prey simultaneously in three dimensions. Spinner dolphins foraged at night in highly coordinated groups of 16–28 individuals using strict four-dimensional patterns to increase prey density by up to 200 times. Herding exploited the prey's own avoidance behavior to achieve food densities not observed otherwise. Pairs of dolphins then took turns feeding within the aggregation that was created. Using a proxy estimate of feeding success, it is estimated that each dolphin working in concert has more access to prey than it would if feeding individually, despite the costs of participating in the group maneuvers, supporting the cooperation hypothesis. Evidence of a prey density threshold for feeding suggests that feedback from the environment may be enough to favor the evolution of cooperation. The remarkable degree of coordination shown by foraging spinner dolphins, the very strict geometry, tight timing, and orderly turn taking, indicates the advantage conferred by this strategy and the constraints placed upon it. The consistent appearance of this behavior suggests that it may be a critical strategy for energy acquisition by spinner dolphins in energy poor featureless environments in the tropical Pacific Ocean. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.2967480]

PACS number(s): 43.30.Sf, 43.80.Ka [JAS] Pages: 125–137

I. INTRODUCTION

The pelagic ocean at low latitudes is low in available energy (Longhurst *et al.*, 1995), potentially making it difficult for large animals with high energetic needs to consume enough food to survive and reproduce. One solution to the challenge of finding and capturing dispersed food in this environment is to forage as a group. Searching together can increase a predator's prey detection rate, as shown in gulls (Gotmark *et al.*, 1986; Hoffman *et al.*, 1981) and dolphins (Norris and Prescott, 1961; Saayman *et al.*, 1973). Cooperation, collective action by a group of animals that increases each individual's fitness, in cormorants (Bartholomew, 1942), gulls (Gotmark *et al.*, 1986), and killer whales (Baird and Dill, 1996) increases their prey capture efficiency. Cooperative groups may also actively herd prey, increasing its density and decreasing its escape options. Increased densities of aquatic prey have been associated with groups of river otters (Serfass, 1995), various species of piscivorous marine fish (Hiatt and Brock, 1948; Partridge *et al.*, 1983; Schmitt and Strand, 1982), pelicans (Anderson, 1991; Goss, 1888), killer whales (Nottestad *et al.*, 2002; Similä and Ugarte, 1993), and dolphins (Gallo Reynoso, 1991; Würsig and Würsig, 1979).

Spinner dolphins (*Stenella longirostris*) are primarily pelagic animals that reside in the energy poor waters of the subtropics. They feed at night on small prey including lanternfish, shrimp, and squid that are part of a deep-scattering layer (Norris and Dohl, 1980). Because of the small size of their prey (less than 0.002% of a spinner dolphin's mass), the prey's diel migration behavior, and their own energetic needs, spinner dolphins need to be efficient foragers, each consuming an estimated 1.25 prey items per minute throughout the night (Benoit-Bird, 2004). Spinner dolphins have been observed in groups while foraging (Benoit-Bird and Au, 2003b), which may play a role in maximizing their foraging efficiency. However, two-dimensional data collected in previous field studies of spinner dolphins do not address the behavior of groups of foraging spinner dolphins or their potential function. Dolphin feeding groups were organized, in agreement with predictions for cooperative foraging, encouraging further study.

Most aquatic animals forage beneath the ocean's surface, making it difficult to observe their behavior, particularly for nocturnal predators such as spinner dolphins. While group foraging has been observed in many aquatic species, their relative inaccessibility has made quantitative assessment of interactions between groups of predators and their prey difficult (Schmitt and Strand, 1982; Serfass, 1995; Similä and Ugarte, 1993). One way toward a better understanding of the behaviors and benefits of group foraging in aquatic animals is using data that allow us to place predators and prey in three-dimensional space over time. We used a multibeam echosounder capable of simultaneous high-resolution observations of predator and prey behavior be-

a)Author to whom correspondence should be addressed; electronic mail: kbenoit@coas.oregonstate.edu

neath the water's surface to observe the nocturnal behavior of spinner dolphins off the coast of Oahu, HI in three dimensions. We recorded the raw acoustic data from the sonar and used underwater acoustics techniques for analysis rather than analyzing the visual output of the sonar, as has been done in previous biological studies (Axelsen *et al.*, 2001; Nottestad and Axelsen, 1999; Ridoux *et al.*, 1997). This allowed us to quantify dolphin group size, foraging effort, and potential feeding opportunities while simultaneously measuring changes in prey density. Previous work on these animals has suggested that group foraging may be important; however, the short term behavior of dolphins could not be measured nor could the immediate effects of this behavior on prey distribution (Benoit-Bird and Au, 2003b), making it impossible to determine if cooperation could occur. Our goals in this study were to understand the benefits of group foraging and the mechanisms of prey aggregation to ultimately assess the potential importance of group foraging in predator survival.

II. METHODS

A. Survey design

To measure the distribution of spinner dolphins and the scattering layer that serves as their prey, two sites along the leeward coast of Oahu were surveyed with a 200 kHz Kongsberg-Simrad-Mesotech MS 2000 multibeam echosounder. The sonar provided three-dimensional information on the distribution of dolphins and their prey while the vessel was moving. This frequency was selected because of its high spatial resolution and its low likelihood of detection by a *Stenella* species. While an audiogram for *Stenella longirostris* is not available, the hearing of *Stenella coeruleoalba* shows a very steep roll-off in sensitivity, as is typical for mammals, above about 120 kHz (Kastelein *et al.*, 2003). This suggests that a *Stenella* sp. is unlikely to hear a 200 kHz signal with an approximately 7 kHz bandwidth.

Surveys were conducted between 2100 and 0300 h local time for ten nights in May and June of 2003. The moon was less than one-quarter illuminated during all sampling so the ambient light level was very low. Sampling was conducted around two areas known for the presence of dolphins, covering approximately 11 km surrounding Makua Bay and 10 km surrounding Kahe Point Beach, about half of the entire leeward (Waianae) coast of Oahu. Sampling was conducted along linear transects running primarily parallel to the shoreline at a vessel speed of approximately 1.5–2.0 m/s (3–4 kn) with occasional stops for other sampling (Benoit-Bird and Au, in press). These transects were located between 1.0 and 2.2 km from the shoreline, in water depths between 60 and 200 m. A total of 186 km was surveyed.

B. Multibeam echosounder

The head of the 200 kHz multibeam echosounder was mounted off the starboard side of the vessel, approximately 0.5 m below the water's surface, at a 45° angle relative to the direction of travel and 30° up from the vertical plane perpendicular to the bottom. This permitted us to observe animals from the side, rather than having to drive directly over them, limiting potential impacts of the survey on the dolphins. The multibeam system used a 150 μs long outgoing pulse transmitted at a rate of 5.6 per second with a source level of 215 dB re 1 μPa. The system had a resolution of 0.22 m in the direction 30° from vertical, resulting in a true depth resolution of 0.19 m. The system had 120, 1.5° × 20° beams that overlapped by 0.25° in the narrower nearly across-track direction, providing an angular coverage of 120° with 1° resolution in this direction. Data were taken using the external imaging transducer of the sonar, thus forming a Mills Cross to provide the greatest spatial resolution, giving a received beam width of 1.5° in the direction 45° from along-track. The sonar system was set to cover a range of 65 m. In the geometry used, the instrument thus had a maximum range of 37.5 m from the vessel horizontally and 53 m vertically, giving a swath width of 113 m at the furthest point in its range. Further technical details about the sonar can be found in Cochrane *et al.* (2003).

The multibeam echosounder system was calibrated in the Applied Physics Laboratory, University of Washington seawater calibration facility using tungsten carbide spheres with diameters of 20 and 38.1 mm following the protocols established by Foote *et al.* (2005). The system showed a nearly linear gain function, stable performance, and beam patterns consistent with those predicted by the system geometry.

During field sampling, raw complex echo signals from each of the 80 elements of the multibeam echosounder were saved to hard disk with 12 bit digitization, resulting in a dynamic range of 72 dB. The data were beamformed with Chebyshev-type amplitude weighting and the amplitude and range of echoes in each beam were then extracted using a custom MATLAB program. The range versus azimuth data were transformed into a rectangular coordinate system where one axis is along the direction of travel, the second is across the direction of travel, and the third is the depth. To obtain three-dimensional information, six successive sonar returns, or about 1 s of data, were combined. All position data were corrected using coupled differential Global Positioning System (GPS) to eliminate the movement of the vessel. GPS readings were taken with each sonar signal at a rate of 5.6 per second. Correction for vessel pitch and roll could not be conducted; however, all sampling was conducted with sea states below Beaufort 2 (waves of less than 0.2 m, winds below 6 kn), often with completely glassy conditions, limiting undesirable vessel movement.

C. Dolphin position analysis

The transformed, averaged, and position corrected data were used to identify large targets and to create isoscattering surfaces. Dolphins were identified from the data based on their unique set of acoustic scattering characteristics as identified by Benoit-Bird and Au (2003b), most notably, a very strong echo that is caused by their air filled lungs (Au, 1996) and possibly their nasal passages that is surrounded by remarkably low-amplitude echoes presumably by their blubber which is nearly impedance matched to seawater. Echoes from spinner dolphins identified during daylight surfacings

were taken nearly simultaneously with a calibrated 200 kHz echosounder (Simrad EK60) and the SM2000. Echoes from spinner dolphins SM2000 had values between −25.9 and −28.2 while the values measured with the EK60 were within 0.3 dB of −27.1 dB. The SM2000 echoes showed more variance in their amplitude likely because of the increase in the variability in insonified angle relative to the narrow beam EK60. In both systems, the presumptive lung echo provided an extremely high target strength within this field of ~ −27 dB echoes. The length of animals measured with the multibeam was consistent with the size for adult spinner dolphins in Hawaiian waters.

Other large targets, e.g., those with high target strengths and of measurable size with the multibeam's 1.5° resolution, were commonly observed along sampling transects. These targets were not consistent with the measured characteristics of known spinner dolphins and were never observed near targets consistent with spinner dolphins and thus were not considered in the analysis since sections of transects between dolphin observations were not utilized here. It is unlikely that similarly sized species of dolphins (e.g., the spotted dolphin *S. attenuata*) could be differentiated using the sonar. However, no other similarly sized species in this habitat are known to be foraging at depth at night and thus other species are not likely a large source of error.

Multiple echoes were simultaneously detected in multiple beams and range bins from a single spinner dolphin, so for the analysis of dolphin position, the single strongest echo from each animal was utilized, probably from its lungs, creating a standardized method of locating the animal in space and measuring distance between animals. However, all echoes matching the characteristics of an individual dolphin were utilized to estimate the length of each dolphin target by adding the width of the beam for each echo at the range it was detected together. The lengths of targets that were not perpendicular to the beam were corrected by estimating the angle of the target through differences in the range between adjacent beams and using basic trigonometric relationships to convert apparent to actual length.

The sonar system has a resolution of 1.5° in the azimuth direction and 9.6 cm in the range direction. To increase the resolution of measurement of animal spacing, information from six successive pings was used to calculate interanimal distances. Changes in spacing in this 1 s time period were assumed to be negligible. The movement of the sonar relative to the animals during this time, however, decreased the minimum interspacing distance that could be measured in all dimensions to approximately 10 cm.

Individual dolphins could not be tracked for extended periods using the multibeam echosounder because the echoes measured from each individual, while distinguishable in space, do not bear individual features that allow them to be identified from other similar echoes at a later time period. However, individual animals could be measured in several successive signals. The movement of each individual dolphin was measured by mapping the position of its strong dolphin air-passage echo in successive pings, permitting estimates of swimming speed to be made over short time intervals.

D. Prey distribution analysis

After removal of large individually identified targets, volume scattering was thresholded at a value of −70 dB. The scattering of the prey layer in each m^3 within the sampling volume was then normalized to the overall average scattering per m^3 for all observations of the scattering layer for statistical analysis. Spatial statistics (Benoit-Bird and Au, 2003c) were utilized to determine the location of significant changes in relative scattering in three-dimensional space. Each 3 dB m^{-3} change in scattering strength, representing a doubling in prey density [see Eq. (1)], was found to represent a significant difference

$$\Delta\text{prey density} = 10\hat{}(\Delta\text{prey scattering strength}/10) \quad (1)$$

Linear interpolation between points of the same relative scattering value created isodensity volumes in the prey field. The size of prey isodensity volumes and their location relative to dolphins could then be analyzed.

III. RESULTS

We made 267 observations of groups of dolphins in the upper 50 m of the water column. Observation duration ranged from 22 to 257 s with a mean duration of 123 s. These observations are longer than expected based on the vessel speed for dolphins swimming in a random direction. Most of the observations of directionally swimming dolphins were of animals swimming along isobath, parallel to the transects, permitting longer observations when animals were swimming in the same direction as the vessel. The longest observations were only possible when the vessel happened to be making a turn about a group of animals and observed them from multiple angles. Indications of changes in behavior—e.g., increases in audible surfacing rate, bowriding, playing at the surface, and swimming around the boat—were noticed during a few times when the vessel was moving unpredictably for reasons unassociated with sampling. However, these types of behaviors were not detected during normal transect sampling, suggesting limited effects of the vessel or the instruments on the dolphins during sampling.

During night-time surveys, apparently foraging dolphins were found in groups in distinctive geometries, defined here as foraging stages. The split-beam echosounder, which could sample the entire water column throughout the study area, confirmed that the scattering layer in the sampling area did not extend past the 50 m vertical limit of the multibeam sonar. The depth of dolphins was always within the depth range of the detected scattering layer, except during brief surfacings, suggesting foraging activity. The depth of dolphins relative to the scattering layer is shown in Fig. 1. The results show that detected dolphin groups covered nearly the entire vertical range of the scattering layer, missing an average of 1.5 m at the top and 3.5 m at the bottom of the layer.

A. Dolphin groups

Groups of dolphins ranged in size from 16 to 28 with a bimodal distribution (modes=20,24). All 267 groups that we observed were comprised of distinct pairs of animals where one was ahead and slightly above the other dolphin in the

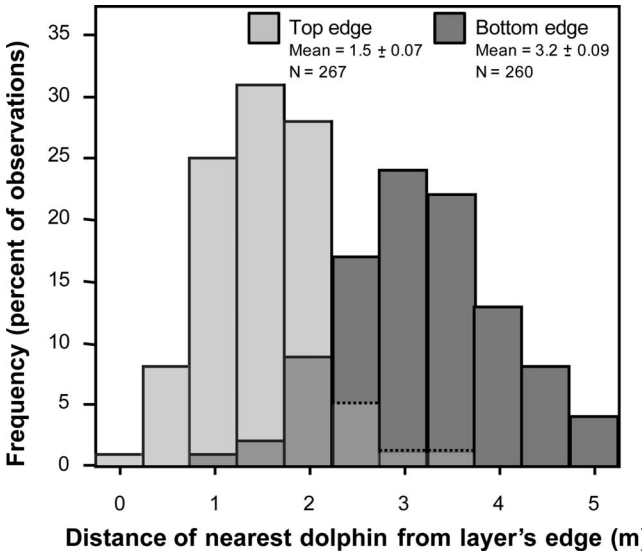


FIG. 1. Depth between the dolphin nearest the edge of the upper and lower edges of the scattering layer and the layer's edge. Foraging dolphins were within the scattering lower so dolphins were deeper than the upper edge of the layer and shallower than the lower edge.

pair (Fig. 2). Consequently, all foraging groups were made up of even numbers of dolphins. This is consistent with results from a previous study (Benoit-Bird and Au, 2003b) though measurements of animal spacing could not be measured from those results. Mean distances between individuals in the pair, corrected for a non-normal distribution using a square-root transformation, are 0.91 ± 0.02 m (95% CI) in the X-Y plane and 1.1 ± 0.01 m (95% CI) in the Z plane. Repeated measures analyses of variance (ANOVAs) on transformed intrapair distance data showed no significant effect of the geometry of the group termed the foraging stage on intrapair distances (X-Y plane: total $N=163,589$, $F_3=1.34$, and $p=0.38$, $1-\beta=0.67$; depth (Z): total $N=163,589$, $dF=3$, $F_3=1.08$, $p=0.73$, and $1-\beta=0.77$).

An ANOVA revealed that group size varied significantly with the initial prey relative density defined as the density of prey measured just prior to the dolphins arrival at that location ($F_{11}=23.11$, $N=96$, and $p<0.001$). These estimates could only be made when the area was observed prior to the movement of a line of dolphins into the area, resulting in a limited sample size for these estimates. Group size tended to be larger when initial prey density was lower and smaller when prey density was higher. A regression showed that dolphin group size was predictably related to prey density (Fig. 3, $r^2=0.40$, $F=173.02$, and $P<0.001$). Dolphin group size was also significantly positively related to increases in variance in scattering strength, a measure of patchiness ($r^2=0.54$, $F=193.66$, and $P<0.001$). The smallest groups of 16 animals were in areas of the scattering layer with σ^2 values of about 2000 while groups of 28 animals were in areas of

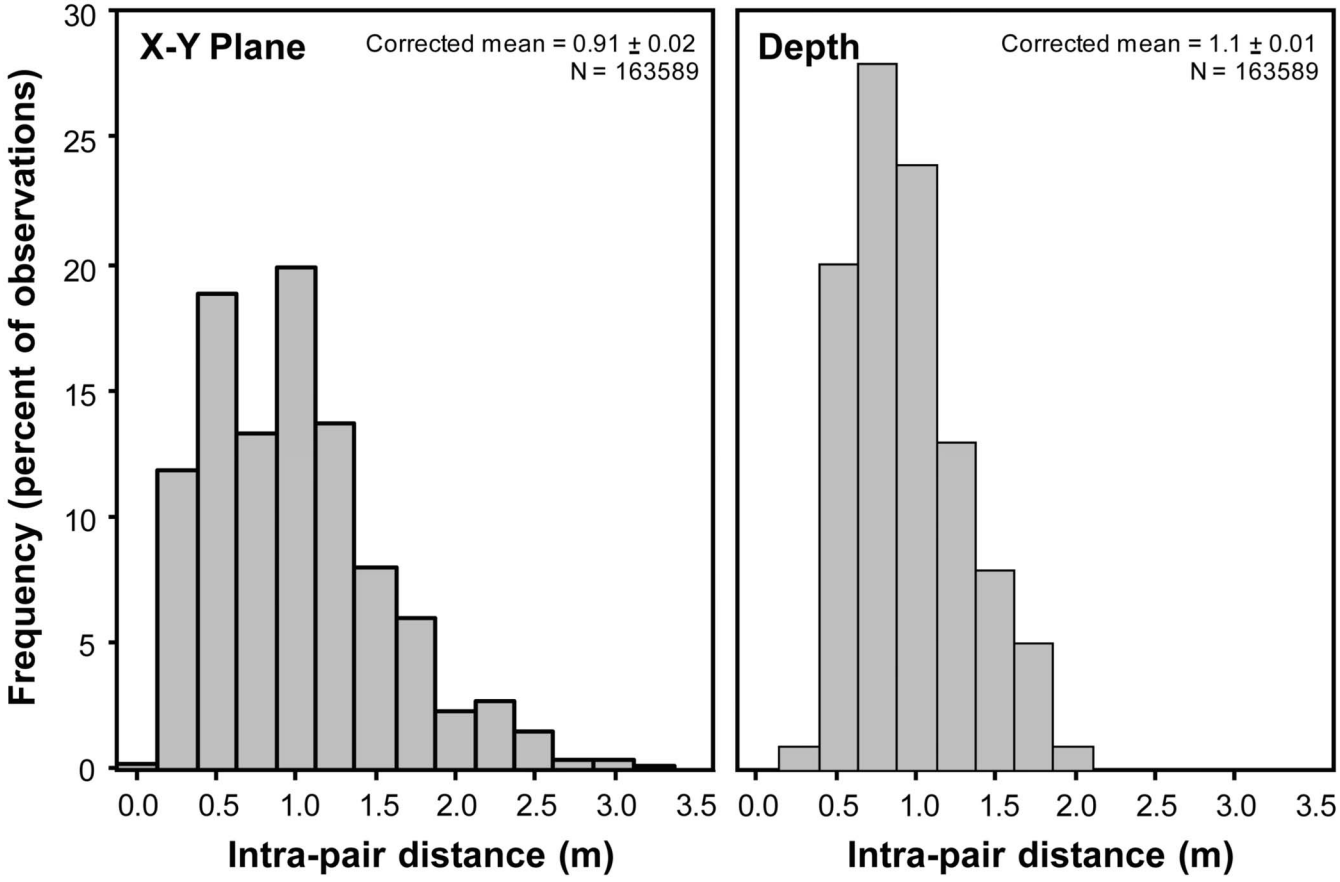


FIG. 2. Distance between individual dolphins within a pair in the X-Y plane (left) and with depth (right). The mean value shown on each graph has been corrected using a square-root transformation of the data which produced a distribution not significantly different from normal. The corrected 95% confidence interval is also shown. Repeated measures ANOVAs on transformed data showed no significant effect of foraging stage on intrapair distances in either plane.

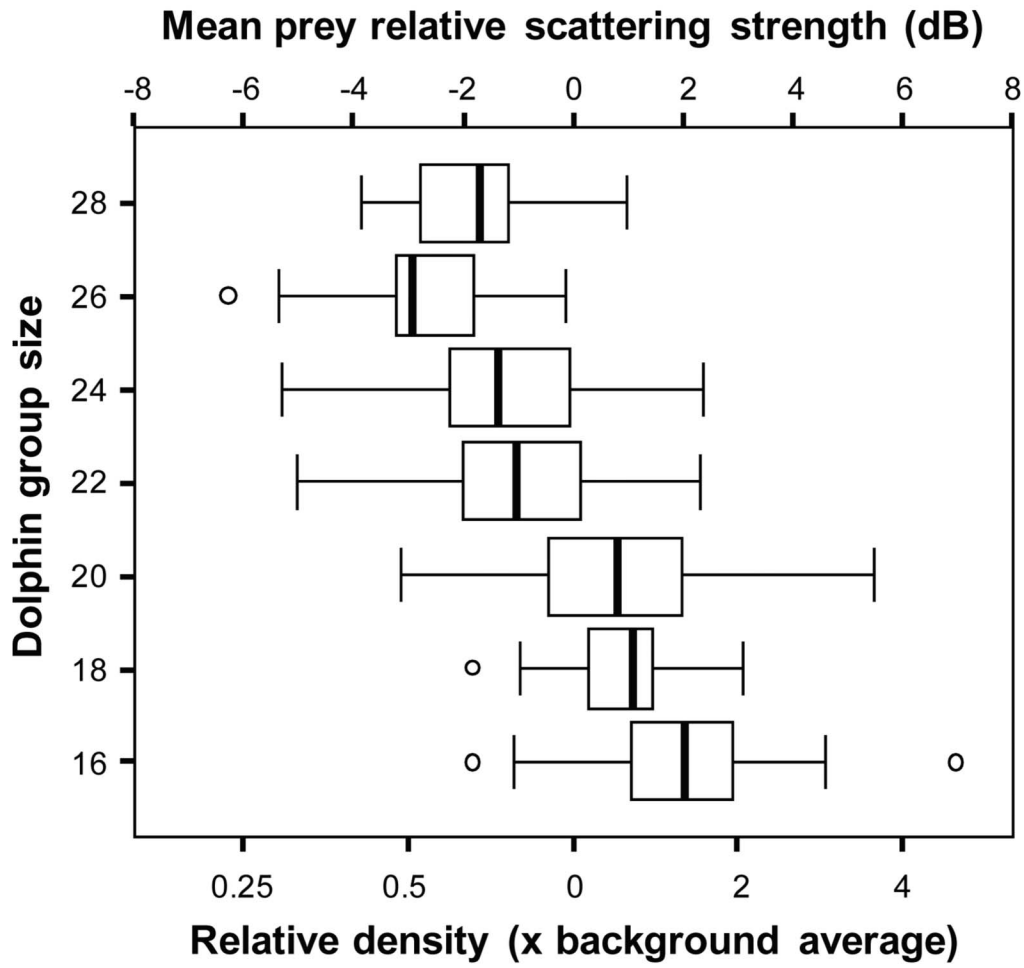


FIG. 3. Dolphin group size as a function of mean relative scattering of prey at the beginning of the observation. Black bars show the median, white boxes show one interquartile range, the error bars show the 95% confidence interval, and dots show outliers.

the scattering layer with σ^2 values of just over 5000. While there was no affect of variance in scattering strength on interpair spacing with depth ($r^2=0.06$, $F=2.37$, and $P>0.05$), the average spacing in the XY plane between pairs of dolphins swimming in a circle was significantly affected by variance in scattering strength ($r^2=0.40$, $F=175.66$, and $P<0.001$). Spacing between animals decreased by up to 1.6 m as variance increased. However, prey relative scattering strength observed within the circle of dolphins, an estimate of the potential benefits of the behavior, did not vary significantly with dolphin group size ($F_{11}=1.57$, $N=267$, $p>0.05$, and $1-\beta=0.76$).

B. Foraging stages

There were four clear geometries or stages of foraging observed in spinner dolphins (Fig. 4). A brief summary will help follow the remaining results and the detailed methods of their determination. Animals were found to swim in a line running perpendicular to the shoreline, termed "wide line." Next, the spacing between pairs in the line decreased dramatically while animals continued to swim forward. The animals in this line then formed a 28–40 m circle, closing the circle from offshore. The final stage of foraging before dolphins surfaced is termed "inside circle." Pairs of dolphins at opposite sides of the circle moved inside the circle formed by the other dolphins. In all foraging stages, dolphins were distributed over nearly the entire depth range of the scattering layer. Each pair of animals appeared to remain at a single depth through an entire foraging bout so that pairs are offset around the circle, forming a cylinder of prey as they repeatedly swim around it.

The beginning and end of foraging stages and their transitions were defined by changes in the geometry of foraging. The beginning of the transition from wide line to tight line was defined as the point when the mean interpair spacing decreased by 10% and the animals were determined to have achieved the tight line formation when the mean interpair spacing for all pairs was within 10% of the minimum of interpair spacing observed in that group later in the observation. The transition to circling from the tight line was determined to have begun when the distance between any two animals perpendicular to the plane of the line was greater than 7 m, slightly greater than the maximum interpair distance measured parallel to the line. The complete formation of the circle was determined as the point when the nearest-neighbor distance between pairs of animals in the group was less than 7 m. The transition to circling to moving inside the circle was defined as the point when the distance between any opposing pairs of animals in the circular formation, the radius of the circle, was 4 m less than the mean radius of the

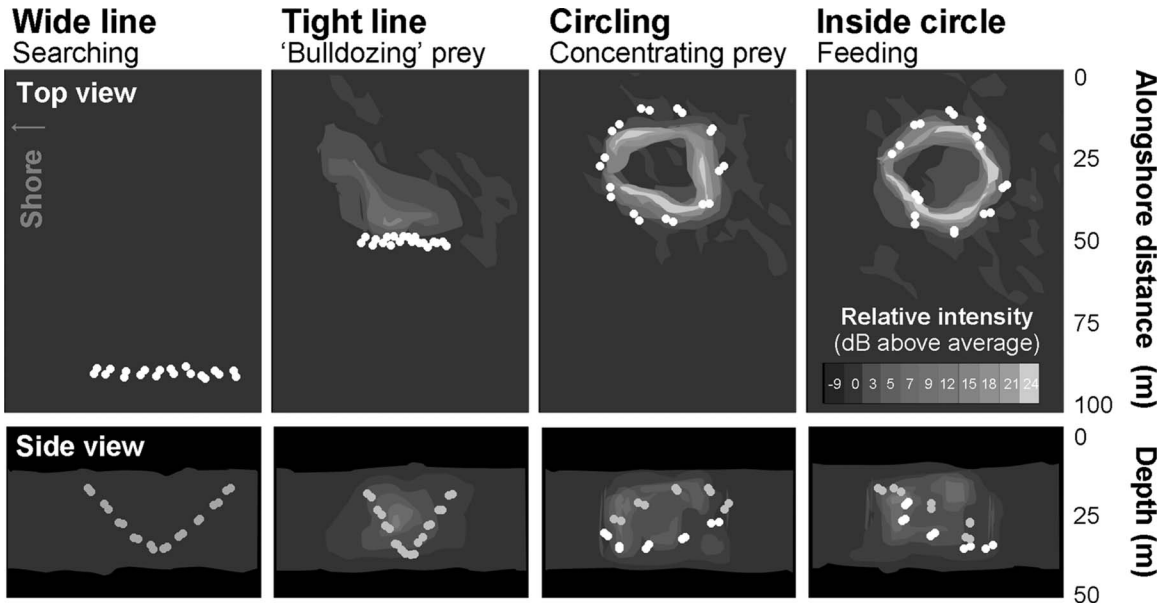


FIG. 4. Top and side view examples of dolphins in each of the four stages of foraging observed. For the top view, scattering was integrated through the entire water column for each square meter. For the side view, scattering was integrated for the 20 m immediately in front of the dolphin group. The relative scattering intensity of prey is indicated by grayscale. Prey density can be related to the scattering intensity with each 3 dB increase in scattering representing a doubling of prey density. The position of the air passage echoes from dolphins is shown as dots. In the side view, the dolphins behind the integration plane are shown as a darker shade.

circle comprised of all the other animals. The interpair spacing for animals clearly within each of the four defined stages is shown in Fig. 5.

Complete observations of entire foraging bouts could not be made from the data; however, estimates of stage duration and stage order could be made using data that contained complete foraging stages, defined as those that contained at least some part of the stage before and after. Combining these data gives us an estimate of the order and duration a complete foraging bout. Figure 6 shows the mean duration of each foraging stage, the 95% confidence interval of the mean, and the number of times each stage was observed completely. Combining all the means gives us an estimate of the duration of a complete foraging bout of 4 min

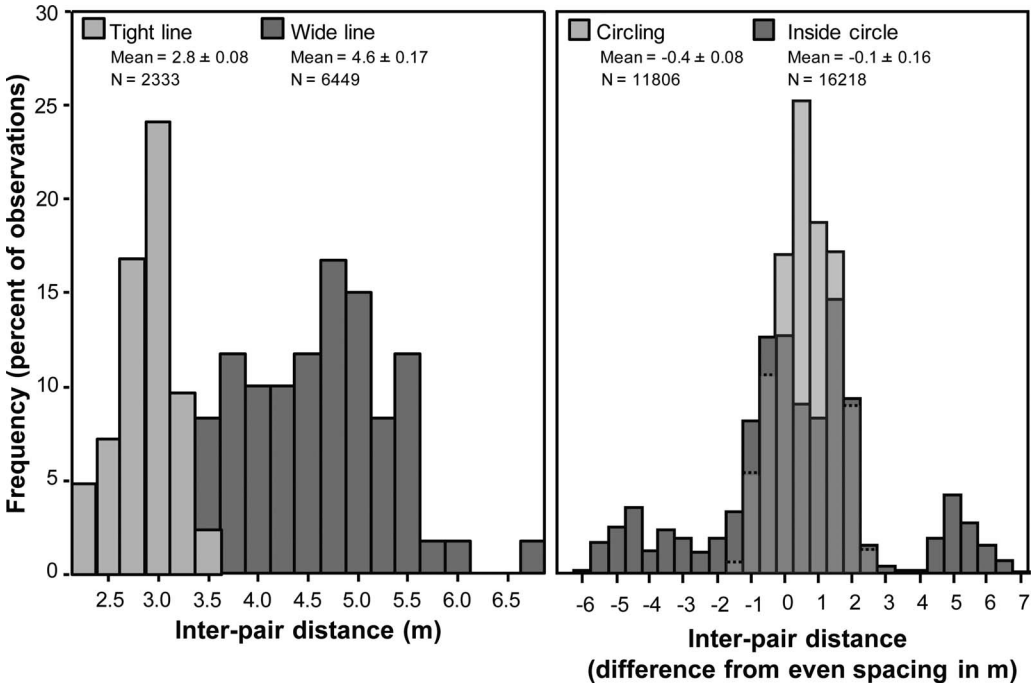


FIG. 5. (Left) Distance in the *x*-*y* plane between pairs of dolphins in a foraging group for the first two foraging stages identified. The distances in line abreast swimming animals were consistent regardless of the size of the group. (Right) Distance in the *x*-*y* plane between pairs of dolphins in a foraging group for the circling foraging stages. Because the distances are significantly affected by the size of the group, distances are shown as the difference between evenly spaced animals in each group, estimated by dividing the circle's diameter by the number of pairs of dolphins in the group, and the actual spacing. Note that the sample sizes represent the number of measurements between pairs, leading to a much greater sample size than the number of groups detected.

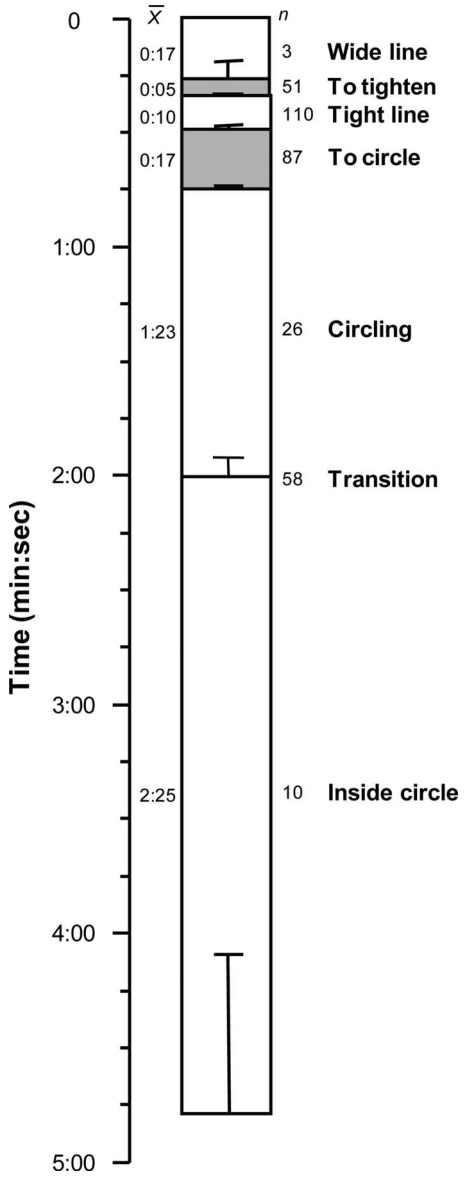


FIG. 6. Mean duration of each foraging stage with 95% confidence intervals indicated by the error bars. The number of complete observations for each stage is shown.

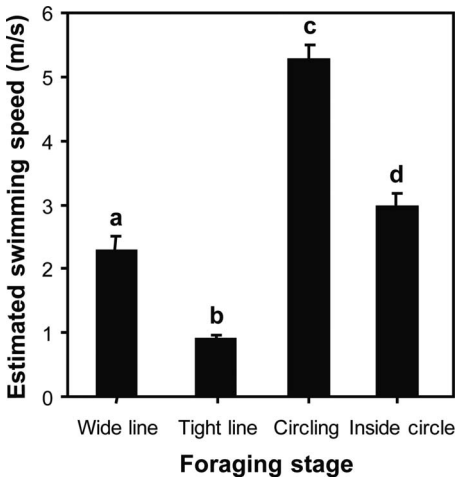


FIG. 7. Mean swimming speed measured from successive detections of dolphins is shown with 95% confidence intervals. An analysis of variance and *post hoc* tests revealed that the swimming speeds for each foraging stage were significantly different.

and 38 s with a total 95% CI for 1 min. The relatively high sample sizes of observed transition between stages, e.g., recording some part of the stage before and some part of the stage after, provide confidence in the interpreted order of behaviors and their consistency. No observations of targets consistent with the echo characteristics of spinner dolphins were made that moved between these stages in a different order from the proposed model of foraging.

There were significant differences in measured swimming speeds between the various defined foraging stages (Fig. 7, ANOVA $F_3=267.96$, and $p<0.0001$). There were no significant effects of time of night or group size on swimming speed ($F_5=0.745$, $P=0.59$; $F_6=2.0$, and $P=0.08$, respectively). A Dunnett's C *post hoc* comparison showed that swimming speed at each stage was significantly different from the speed of every other stage ($p<0.05$). The swimming speeds measured here compare well with the 0.26–4.32 m s^{-1} estimates of bottlenose dolphins speeds made from a video sonar (Ridoux *et al.*, 1997) with the highest speeds observed when animals were swimming in a circle before animals began to move into the circle when their swimming speed decreased by nearly half.

There were significant changes in prey relative scattering strength between and within foraging stages. An ANOVA showed that there was a significant effect of foraging stage on prey relative scattering strength ($F_5=244.45$ and $P<0.0001$). *Post hoc* Dunnett's C comparisons showed that the prey relative scattering strength of each stage was significantly different from that of every other stage ($P<0.05$). Figure 8 shows prey relative scattering strength as a function of time within each foraging stage. Regression analyses were used to observe these changes. See Fig. 8 for all R^2 and p values. There was no significant change in prey scattering while the dolphins were in their wide line, the transition from a line to a circle, or for the first 75 s when dolphins were inside the circle after which there was a significant decrease in prey scattering. There was a significant increase in prey relative scattering strength in the 5 m immediately in front of the dolphin group when the animals were tightening their line, were in the tight line, and were circling. Putting all the separate observations together, the foraging behavior observed correlates with an increase in the mean relative scattering intensity of 9.6 dB from the wide line until the end of circling, representing an increase in the prey density of about 11-fold. The maximum density of the scattering layer over this same observation period increases by an average of 60 times up to a maximum of approximately 200 times.

C. Foraging stage possible functions

The role of each stage of foraging was investigated by simultaneously observing dolphin behavior and changes in the prey field. The mean relative scattering strength of prey as a function of distance from the dolphins is shown in Fig. 9. All of the observations shown were aligned using the time at which animals were first observed to tighten their line abreast formation. This figure only shows the 30 s surround-

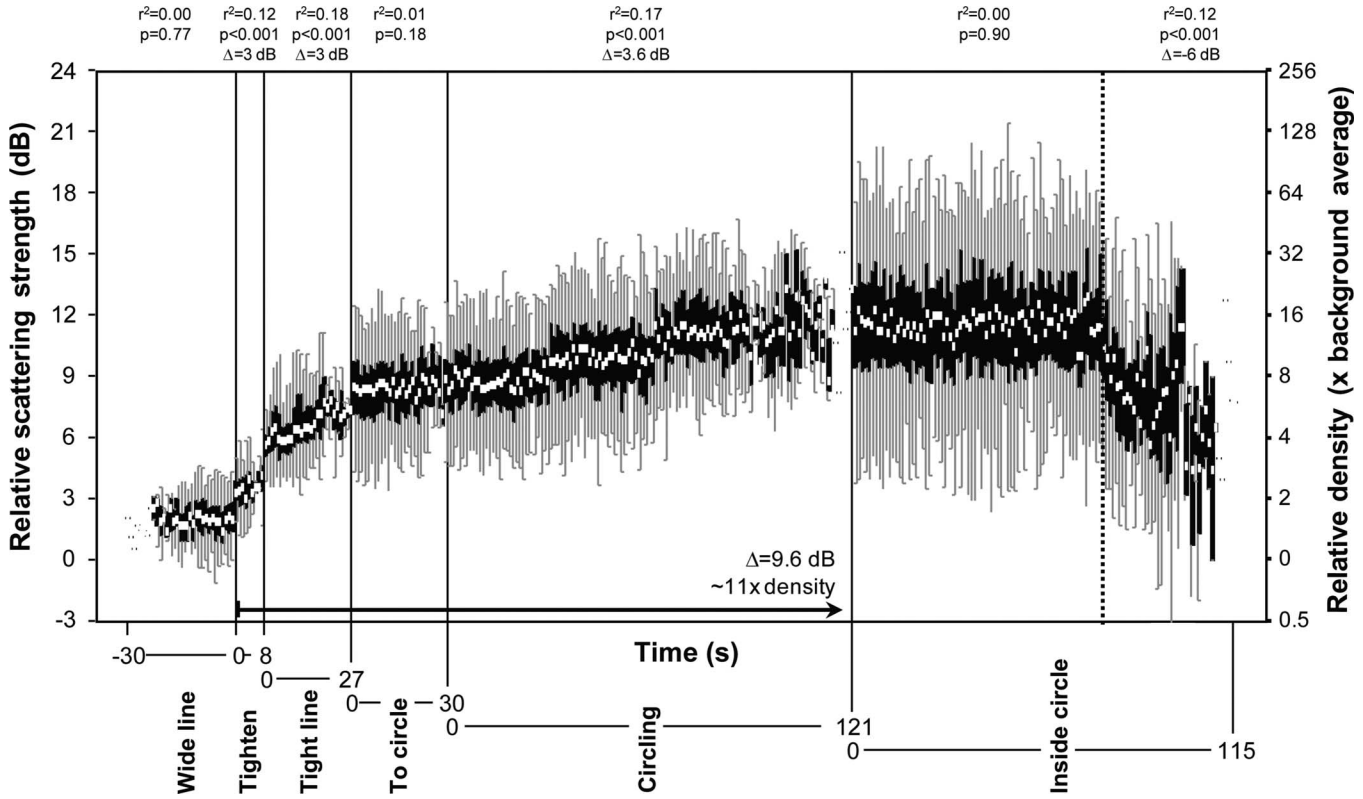


FIG. 8. Mean prey relative scattering strength per m^3 for the 5 m in front of dolphins as a function of time in each foraging stage for all 267 observations. White bars show the median, black boxes show one interquartile range, and the error bars show the 95% confidence interval. Times with single samples are indicated by black points. The inside circle stage is broken up into two parts by a dotted line at 75 s into the behavior. Results of regression and *F*-tests for scattering strength within each foraging stage are shown at the top of the figure.

ing this event. While swimming in the wide line, dolphins appeared to approach an existing patch of increased scattering (shown in light gray) estimated to have a numerical density twice the average background prey density (Fig. 9, first row). This patch was observed up to 25 m ahead of the dolphins. When dolphins were observed to move to within 5 m of this patch, their interpair spacing began to decrease, with the final tight line formation achieved as they reach the edge of the patch (Fig. 9, second row). When groups of dolphins were coincident with the edge of this patch, the swimming speed of the dolphins was observed to decrease by about 60% (see Fig. 7), approximately matching the maximum sustained swimming speed for their prey (Reid, 1994; Sambilay, 1990). During this time, dolphins were also observed to swim in a sinusoidlike pattern, moving up and down by several meters. During this phase of foraging, the density of prey 1 m in front of the dolphins was measured to increase relative to the original patch (shown in dark gray), as the prey apparently tried to swim away from the dolphins, while the density immediately behind the dolphins was observed to decrease (Fig. 9, last two rows).

While the dolphins were observed to be circling, the density of the patch measured as an increase in volume scattering strength was measured to increase further by an average of another 2.5 times to about 11–12 times the background (Fig. 8). In addition, the distribution of prey was observed to change dramatically during this phase. Figure 10 shows the percent of peak prey scattering, or increased prey density, on the side of the circle from which the line of dolphins approached from. There was a significant effect of time on this value ($r^2=0.60$, $F=7166.08$, and $p<0.001$). As the mean duration of the circling stage (1:23 min) is approached, the prey becomes close to evenly distributed on both sides of the circle even though the strong scattering or high prey density was initially on only one side of the circle. Figure 11 shows that the size of the circle is significantly affected by the size of the group ($r^2=0.38$, $F=165.30$, and $p<0.001$) with larger groups in larger diameter circles despite the fact that larger groups had significantly tighter spacing in the *XY* plane.

During the final stage of foraging observed, dolphins moved into the circle in sets of two pairs that were opposite or nearly opposite each other in the circle. The average number of movements into the circle calculated as the total number of observed movements into the circle divided by the number of pairs in the group was 3.7 per pair ± 1.3. Individual pairs spent 10.5 ± 1.1 s in the circle before falling back into their place in the circle. In all observations, the foraging that took place was very orderly with the pairs of dolphins immediately behind the initial pairs in the circle moving into the circle next and so on. The average number of movements into the circle by each pair of dolphins per foraging bout calculated as the total number of observed movements into the circle divided by the number of pairs in the group was not affected by group size (ANOVA $N=154$, $p>0.05$, and $1-\beta=0.69$) which means that more movements into the circle were conducted in larger groups. Unlike all the other foraging geometries, when dolphins were ob-

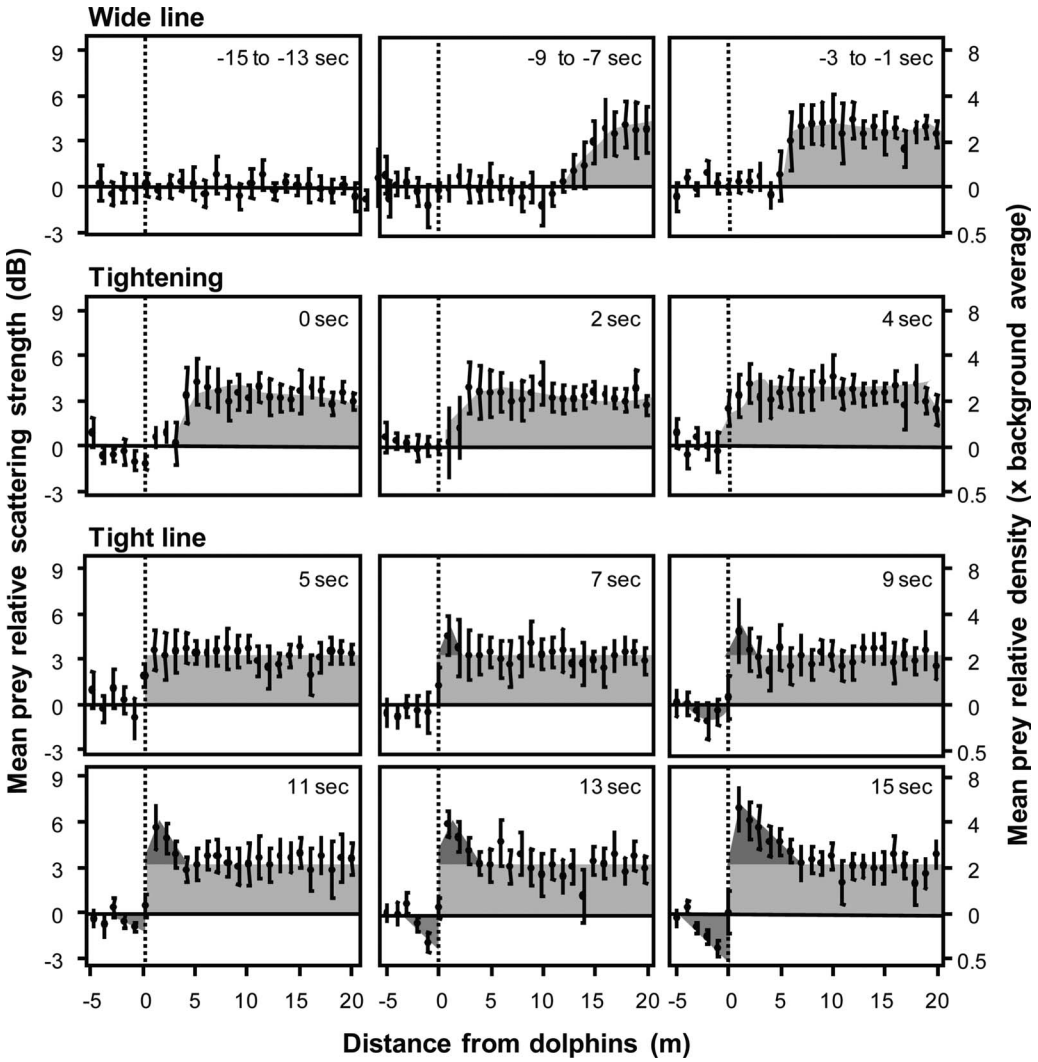


FIG. 9. Mean relative scattering strength of prey as a function of distance from dolphins at each time interval shown. Time zero for all observations was the first time that tightening of dolphins swimming line abreast was observed. Error bars represent the standard error of the mean. Light fill represents the density difference of an existing prey patch relative to the background prey density. Dark fill represents the prey density change hypothesized to have been caused by dolphins herding the prey.

served in the circle, no further increase in the prey density was observed (see Fig. 8), indicating a very different function for this foraging stage from the others. Figure 12 shows that dolphins moving into the circle were moving an average of 5.1 m into the circle, to the location where the prey has a relative scattering strength an average of 17.7 dB and up to 23.1 dB higher than the background scattering by the layer. That represents an average prey density increase of 60 times up to a maximum of 200 times over the average background prey density. Dolphins were strongly associated with the location of highest density of their prey while inside the circle. Near the edge of the circle of dolphins, volume scattering and thus estimated prey density were not significantly higher than the background. This was measured using a paired t-test comparing the average volume scattering near each pair of dolphins in the circle with the average volume scattering 5 m away from the pair toward the outside of the circle ($N=154$, $p \gg 0.05$, and $1-\beta=0.99$).

IV. DISCUSSION

We present a model of spinner dolphin foraging based on observations of dolphins and their prey using a multibeam echosounder. Four distinct geometries were consistently observed in the underwater nocturnal behavior of spinner dolphins feeding on a deep-scattering layer near the apex of its diel vertical migration. No single observation encompassed all stages because of limits in observation time from the moving vessel. To understand how these patterns fit into the overall foraging behavior of spinner dolphins, observations of transitions between these four stages were used to determine stage order and overall time course. Sample sizes of observations of these foraging stage transitions were relatively high (see Fig. 6); however, this approach is biased toward shorter foraging stages which can be seen in the sample sizes. Consistent progression in the foraging stage was observed throughout, suggesting that the model, presented as a time course in Fig. 6, is not often varied under the foraging conditions measured, validating further interpretation.

The model of dolphin foraging presented is a compilation of many observations that are assumed to be independent. We cannot know if individual dolphins were detected more than once in a single night since no method currently

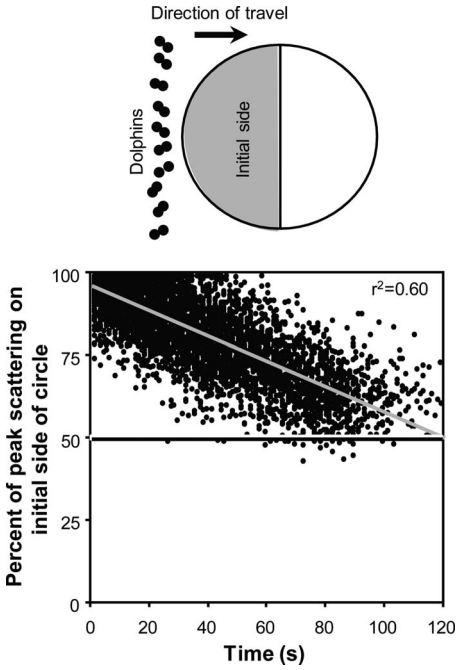


FIG. 10. The initial side of the dolphin foraging circle, the side from which the line of dolphins came from, is shown in the top panel. The percent of the volume of the foraging cylinder that had prey scattering within 6 dB of the maximum for that observation as a function of time from the inception of circling by the dolphins is shown in the bottom. The 50% line indicates the level where prey would be evenly distributed on both sides of the circle.

exists to identify individuals at night. However, using the best population estimates for spinner dolphins, we can estimate the population of spinner dolphins along the area we surveyed to be at least 120 and possibly up to 750 animals (Lammers, 2004; Mobley *et al.*, 2000). Given population estimates, our coverage of the area (~50%), and our average group sizes in the low to mid-20s, it is possible that 20 or more groups may have been present within the general area surveyed. Along individual parallel to shore transects, the number of groups sometimes reached an average of 4/km. During one 8 km long survey at a vessel speed of ~4 kn (faster than the directional movement observed in the dolphins), 18 distinct groups of dolphins were detected. So

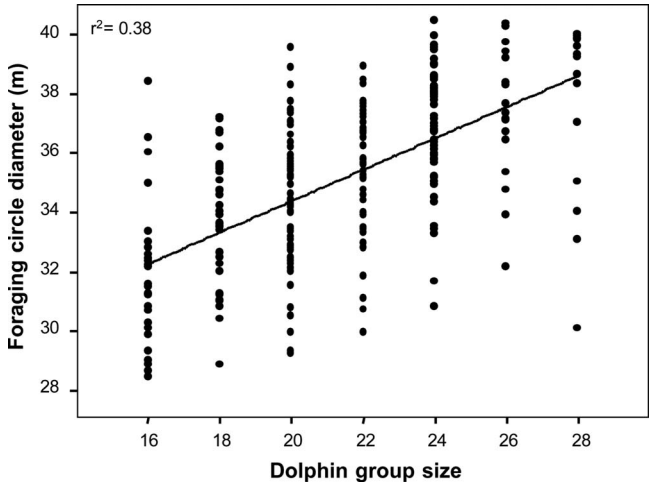


FIG. 11. Foraging circle diameter as a function of dolphin group size

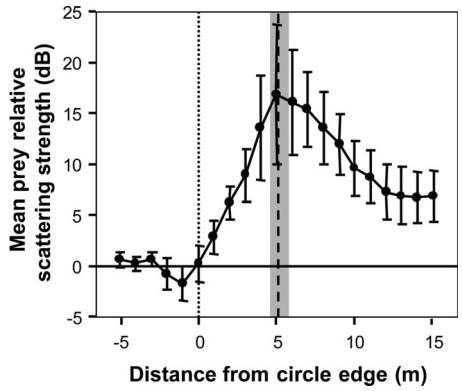


FIG. 12. Mean relative scattering strength of prey as a function of distance from the circle of dolphins. Positive values represent positions inside the circle and negative values represent positions outside the circle. Error bars represent the 95% confidence interval for the mean. The mean for the maximum movement of each pair of dolphins into the circle is shown by the dotted line near 5 m, along with gray bar representing the 95% confidence interval.

while it is possible that some groups may have been detected more than once over the course of a single night's survey, given the mean number of observations each night (~27), the likelihood of significant pseudoreplication is low.

Observations of individual foraging stages suggest that swimming in a wide line, tight line, and then a circle allowed dolphins to increase the density of their prey. Putting it all together, dolphins first found and approached an existing patch that was about twice the density of the background. Then, dolphins used their prey's own avoidance behavior to push it forward as they continued to swim, piling the prey up on itself like snow in front of a plow. Dolphins isolated this prey patch by encircling it. The final patch was, on average, 11–12 times higher in density than the background prey with a peak density that was an average of 60 times higher than the background. During the last observed stage of foraging, dolphins moved into the circle and appeared to actively feed. Inside the circle was where they most closely overlapped the maximum density regions in their prey while dolphins swimming at the edges of the circle did not experience prey densities that were significantly increased over the background. Presumably feeding dolphins moved into the circle with a high degree of coordination not only with their pair partners but also with other pairs. In all observations, two pairs were observed to move into the circle simultaneously from nearly opposite sides of the circle's perimeter. When these two pairs moved back to their places on the circle's edge, another two pairs of dolphins were observed to move into the circle from the next position behind them in the circle.

Group foraging can allow access to prey that would otherwise be difficult or unprofitable for the predator to find or consume alone. In terrestrial predators, this usually means obtaining access to hidden prey (Jones, 1999) or prey that are large relative to the predator (Estes and Duggins, 1995; MacDonald, 1983; Moehlman, 1986; Traniello, 1987). In pelagic aquatic predators, as in aerial predators (Brown *et al.*, 1991), cooperation most often involves finding or creating aggregations of small prey (Anderson, 1991; Bartholomew, 1942; Gallo Reynoso, 1991; Goss, 1888; Gotmark *et al.*, 1986;

Hiatt and Brock, 1948; Hoffman *et al.*, 1981; Norris and Prescott, 1961; Nottestad *et al.*, 2002; Partridge *et al.*, 1983; Saayman *et al.*, 1973; Schmitt and Strand, 1982; Serfass, 1995; Similä and Ugarte, 1993; Würsig and Würsig, 1979). Herding of prey seems to occur most often in markerless environments, like the midwater habitat of the prey of spinner dolphins, where prey do not naturally aggregate around other structures. Group foraging allows aquatic predators to exploit groups of small prey that would not be profitable prey if dispersed. However, unlike other species of marine animals that have been observed to herd prey (Evans, 1987; Gallo Reynoso, 1991; Serfass, 1995; Similä and Ugarte, 1993), spinner dolphins were not observed to bring their prey up to the surface or toward shore. They did not use these boundaries as walls against which to push their prey but rather swam continuously at relatively high speeds which appeared to create a barrier of predators to contain their prey. This makes the behavior difficult to observe and limits the amount of time that dolphins can maintain the patch before needing to surface to take a breath, about 5 min. Similar herding of prey at depth in other marine species has yet to be observed.

Comparing groups of foraging dolphins, we found that the dolphin group size was higher at lower background prey densities. However, the density at which dolphins began to move inside the circle was not affected by background density or group size. This suggests a threshold of prey density at which feeding becomes profitable. A possible threshold mechanism may be the cessation of avoidance of dolphins by masses of prey at the high densities herded prey reach after the circling behavior. This lack of avoidance, observed by measuring the volume scattering 5 m inside the circle of dolphins before and after the movement of dolphins into the circle, could be a result of decreased space for each individual prey. Estimates of the density of micronekton made with a calibrated sonar system in both this and previous studies reached 1800 m^{-3} (Benoit-Bird and Au, 2003a; Benoit-Bird and Au, 2004). This would put the prey, which average 7.5 cm in length (Benoit-Bird and Au, 2006), about 8 cm apart. Because these fish are not observed to swim in a polarized fashion (Benoit-Bird and Au, 2006; Reid, 1994), these densities would severely limit their flight distance, potentially causing great confusion of the prey while under attack. The increase in group size may be necessary to achieve the feeding threshold, whatever its mechanism of action, at lower prey densities.

The observations here lead to the question "is group foraging behavior in spinner dolphins cooperative?" Cooperation can be defined as an outcome that, despite potential individual costs, increases the fitness of the members of a group of two or more individuals, and whose achievement requires collective action (Dugatkin *et al.*, 1992). Coordination of movement, "collective action," of the hunting dolphins is strongly supported by the data presented but that alone is not evidence of cooperation. For us to identify cooperation, we must measure the outcome of the behavior, a difficult task. So in this, as in most studies, we must rely on proxy measures. It was not possible to measure the prey capture success of individual dolphins. However, because prey density is the factor affected by group behavior, we can use prey density, a measure of immediate prey availability as a proxy measure for feeding success. For the purposes of this discussion we will use a simple, one to one linear increase in foraging success with increasing density.

To assess cooperation, we must also estimate the costs of feeding in a group. Based on modeling of spinner dolphin energy needs (Benoit-Bird, 2004) and the density and size of the average patch, we estimate that the prey within at least some single patches of prey could feed the entire group of dolphins for a single day if they could exploit it effectively. Competition between individuals in the group is then not for the food itself but in time to feed. Group feeding reduces the time to feed because of the time it takes to create and maintain a dense prey patch, something we can estimate. Our observations suggest that, on average, each pair of dolphins is inside the circle where feeding is hypothesized to take place 3.7 times/foraging bout for an average of 10.5 s/movement within an average feeding bout of 4 min and 38 s. There is no evidence for significant variation in this time between different pairs of animals within the group. This means that individual dolphins are feeding about 14% of an average foraging bout. Short surfacing times detected in this study support the assumption that animals begin another foraging bout nearly immediately after a previous effort. So, we can use the value of 14% as an estimate of potential foraging time in which dolphins within groups are feeding. A solitary animal would potentially be able to feed 100% of foraging time.

Dolphins in groups experienced a mean prey density increase of about 5.5 times and an increase in maximum density of 30–100 times over the initially selected prey patch, something a solitary individual could also likely select. Looking at the mean density values, the prey capture success measured as density multiplied by access time was 0.75 times greater for animals in groups compared with those alone selecting natural patches in the environment. If maximum density is the important feature, then dolphins in groups could have feeding success rates of 4–14 times higher. Based on this assessment, animals feeding inside these highly dense dolphin-enhanced patches would indeed have higher per-capita feeding success than those attempting to forage alone but only in the high-density areas inside the circle of herding dolphins. The analysis supports the hypothesis that cooperation, rather than just coordinated foraging, occurs and suggests that feedback from the environment alone may be enough to discourage solitary foraging. It is important to note here that targets consistent with individual dolphins were not observed during the study.

The analysis presented supports a positive outcome of the observed collective action over solitary feeding, supporting the hypothesis that this behavior is cooperative. However, cooperation can only evolve if cheating is not a beneficial strategy. We made no observations of animals moving into the circle when it was not their "turn" to do so. No echoes consistent with dolphins were ever observed alone or in small groups. Targets consistent with dolphins were not observed engaging in behaviors that did not fit into the "foraging states" described here that would suggest stalking or

sneaking into a group nor were any observations made of animals spending significantly longer inside the circle than others. This suggests that cheating under the conditions of this study was absent or rare.

Our data show that spinner dolphins worked collectively to achieve densities of prey that did not occur in the habitat in the absence of this dolphin behavior, likely resulting in increased feeding success despite the individual costs of herding and maintaining a prey patch and other potential costs of social feeding. There was no evidence of cheating by animals either within groups or from outside groups. All of this, when considered together, strongly supports cooperative rather than simply group foraging. The remarkable degree of coordination shown by foraging spinner dolphins, the very strict geometry, tight timing, and orderly turn taking, indicates the advantage conferred by this strategy and the constraints placed upon it. The consistent appearance of this behavior shown here and indicated in previous studies (Benoit-Bird and Au, 2003b) suggests that it may be a critical strategy for energy acquisition in Hawaii by spinner dolphins. The number of pelagic species in which cooperative herding has been initially reported (Anderson, 1991; Axelsen *et al.*, 2001; Gallo Reynoso, 1991; Goss, 1888; Hiatt and Brock, 1948; Nottestad *et al.*, 2002; Partridge *et al.*, 1983; Schmitt and Strand, 1982; Serfass, 1995; Similä and Ugarte, 1993; Würsig and Würsig, 1979) suggests that prey herding may be an important strategy for food acquisition in energy poor relatively featureless environments like the tropical Pacific Ocean.

ACKNOWLEDGMENTS

Jeff Condiotty of Simrad, Lynnwood, WA generously allowed us to utilize the multibeam echosounder system. Christopher Jones provided MATLAB code that served as the basis for the custom programs used in this study. Marc Lammers provided assistance in preparing for and conducting the research. Dieter Lammers designed and constructed the multibeam transducer mount and served as our vessel's skipper. Christopher Bird repaired the sonar computer's Windows operating system. The technical staff at Kongsberg-Simrad-Mesotech provided emergency assistance with the multibeam. Michiel Schotten, Aran Mooney, Kristen Taylor, and Stuart Ibsen provided assistance in the field. Alex De Robertis, Christopher Bird, Paul Frost, and Mark Scheuerell commented on the manuscript. This work was funded by the National Science Foundation, Grant No. 0205752, the Office of Naval Research, Grant No. N00014-02-1-0968, as well the National Oceanic and Atmospheric Administration, Project No. R/FM-7, which is sponsored by the University of Hawaii Sea Grant College Program, SOEST, under the Institutional Grant No. NA16RG2254 from NOAA Office of Sea Grant, Department of Commerce. The work presented here represents the views of the authors and not necessarily those of funding agencies. This work was conducted under U.S. National Marine Fisheries Service Permit No. 1000–1617.

Anderson, J. G. T. (**1991**). "Foraging behavior of the American white pelican (*Pelecanus erythrorhyncos*) in western Nevada," Colonial Waterbirds **14**, 166–172.

Au, W. W. L. (**1996**). "Acoustic reflectivity of a dolphin," J. Acoust. Soc. Am. **99**, 3844–3848.

Axelsen, B. E., Anker, N. T., Fossum, P., Kvamme, C., and Nøttestad, L. (**2001**). "Pretty patterns but a simple strategy: Predator-prey interactions between juvenile herring and Atlantic puffins observed with multibeam sonar," Can. J. Zool. **79**, 1586–1596.

Baird, R. W., and Dill, L. M. (**1996**). "Ecological and social determinants of group size in transient killer whales," Behav. Ecol. **7**, 408–416.

Bartholomew, G. A. (**1942**). "The fishing activities of double crested cormorants on San Francisco Bay," Condor **44**, 13–21.

Benoit-Bird, K. J. (**2004**). "Prey caloric value and predator energy needs: Foraging predictions for wild spinner dolphins," Mar. Biol. (Berlin) **145**, 435–444.

Benoit-Bird, K. J., and Au, W. W. L. (**2003a**). "Echo strength and density structure of Hawaiian mesopelagic boundary community patches," J. Acoust. Soc. Am. **114**, 1888–1897.

Benoit-Bird, K. J., and Au, W. W. L. (**2003b**). "Prey dynamics affect foraging by a pelagic predator (*Stenella longirostris*) over a range of spatial and temporal scales," Behav. Ecol. Sociobiol. **53**, 364–373.

Benoit-Bird, K. J., and Au, W. W. L. (**2003c**). "Spatial dynamics of a nearshore micronekton sound-scattering layer," ICES J. Mar. Sci. **60**, 899–913.

Benoit-Bird, K. J., and Au, W. W. L. (**2004**). "Diel migration dynamics of an island-associated sound-scattering layer," Deep-Sea Res. **51**, 707–719.

Benoit-Bird, K. J., and Au, W. W. L. (**2006**). "Extreme diel horizontal migrations by a tropical nearshore resident micronekton community," Mar. Ecol.: Prog. Ser. **319**, 1–14.

Benoit-Bird, K. J., and Au, W. W. L., (**2009**), "Phonation behavior of cooperatively foraging spinner dolphins," J. Acoust. Soc. Am. **124**, 539–546.

Brown, C. R., Brown, M. B., and Shaffer, M. L. (**1991**). "Food-sharing signals among socially foraging cliff swallows," Anim. Behav. **42**, 551–564.

Cochrane, N. A., Li, Y., and Melvin, G. D. (**2003**). "Quantifications of a multibeam sonar for fisheries assessment applications," J. Acoust. Soc. Am. **114**, 745–758.

Dugatkin, L. A., Mesterton-Gibbons, M., and Houston, A. I. (**1992**). "Beyond the prisoners-dilemma-toward models to discriminate among mechanisms of cooperation in nature," Trends Ecol. Evol. **7**, 202–205.

Estes, J. A., and Duggins, D. O. (**1995**). "Sea otters and kelp forests in Alaska: Generality and variation in a community ecology paradigm," Ecol. Monogr. **65**, 75–100.

Evans, D. (**1987**). "Dolphins as beaters for gulls?" Bird Behavior **7**, 47–48.

Foote, K. G., Chu, D., Hammar, T. R., Baldwin, K. C., Mayer, L. A., Hufnagle, L. C., Jr., and Jech, J. M. (**2005**). "Protocols for calibrating multibeam sonar," J. Acoust. Soc. Am. **117**, 2013–2027.

Gallo Reynoso, J. P. (**1991**). "Group behavior of common dolphins *Delphinus delphis* during prey capture," An. Inst. Biol. Univ. Nac. Auton. Mex. Ser. Zool. **62**, 253–262.

Goss, N. S. (**1888**). "Feeding habits of *Pelecanus erythrorhynchos*," Auk **5**, 25–27.

Gotmark, F., Winkler, D. W., and Andersson, M. (**1986**). "Flock-feeding on fish schools increases individual success in gulls," Nature (London) **319**, 589–591.

Hiatt, R. W., and Brock, V. E. (**1948**). "On the herding of prey and the schooling of the black skipjack, *Euthynnus yaito* Kishinouye," Pac. Sci. **2**, 297–298.

Hoffman, W., Heinemann, D., and Wieings, J. A. (**1981**). "The ecology of seabird feeding flocks in Alaska," Auk **98**, 437–456.

Jones, J. (**1999**). "Cooperative foraging in the Mountain Caracara in Peru," Wilson Bull. **111**, 437–439.

Kastelein, R. A., Hagedoorm, M., Au, W. W. L., and De Haan, D. (**2003**). "Audiogram of a striped dolphin (*Stenella coeruleoalba*)," J. Acoust. Soc. Am. **113**, 1130–1137.

Lammers, M. O. (**2004**). "Occurence and behavior of Hawaiian spinner dolphins (*Stenella longirostris*) along Oahu's leeward and south shores," Aquat. Mamm. **30**, 237–250.

Longhurst, A. R., Sathyendranath, S., Platt, T., and Caverhill, C. (**1995**). "An estimate of global primary production in the ocean from satellite radiometer data," J. Plankton Res. **17**, 1245–1271.

MacDonald, D. W. (**1983**). "The ecology of carnivore social behaviour," Nature (London) **301**, 379–384.

Mobley, J., Spitz, S. S., A, K. K., and Grotefendt, R. (**2000**). "Distribution and abundance of odontocete species in Hawai'ian waters: Preliminary results of 1993-98 aerial surveys," Southwest Fisheries Science Center,

Administrative Report No. LJ-00-14C, National Marine Fisheries Service, Silver Spring, MD.
Moehlman, P. D. (**1986**) "Ecology of cooperation in canids," in *Ecological Aspects of Social Evolution in Birds and Mammals*, edited by D. I. Rubenstein and R. W. Wrangham (Princeton University Press, Princeton, NJ), pp. 452–470.
Norris, K. S., and Dohl, T. P. (**1980**). "Behavior of the Hawaiian spinner dolphin, *Stenella longirostris*," Fish. Bull. **77**, 821–849.
Norris, K. S., and Prescott, J. H. (**1961**). "Observations of cetaceans of Californian and Mexican waters," Univ. Calif. Publ. Zool. **63**, 291–402.
Nottestad, L., and Axelsen, B. E. (**1999**). "Herring schooling manoeuvres in response to killer whale attacks," Can. J. Zool. **77**, 1540–1546.
Nottestad, L., Ferno, A., and Axelson, B. E. (**2002**). "Digging in the deep: Killer whales' advanced hunting tactic," Polar Biol. **25**, 939–941.
Partridge, B. L., Jaohansson, J., and Kalish, J. (**1983**). "The structure of schools of giant bluefin tuna in Cape Cod Bay," Environ. Biol. Fishes **9**, 253–262.
Reid, S. B. (**1994**). "Spatial structure of the mesopelagic fish community in the Hawaiian boundary region," Ph.D. thesis, University of Hawaii, Honolulu, HI.
Ridolix, V., Guinet, C., Liret, C., Creton, P., Steensrup, R., and Beauplet, G. (**1997**). "A video sonar as a new tool to study marine mammals in the wild: measurements of dolphin swimming speed," Marine Mammal Sci. **13**, 196–206.
Saayman, G. S., Tayler, C. K., and Brower, D. (**1973**). "Diurnal activity cycles in captive and free ranging Indian Ocean bottlenose dolphins *Tursiops truncatus* Ehrenberg," Behaviour **44**, 212–233.
Sambilay, V. C., Jr. (**1990**). "Interrelationships between swimming speed, caudal fin aspect ratio and body length of fishes," Fishbyte **83**, 16–20.
Schmitt, R. J., and Strand, S. W. (**1982**). "Cooperative foraging by yellowtail, *Seriola lalandei* (Carangidae), on two species of fish prey," Copeia **3**, 714–717.
Serfass, T. L. (**1995**). "Cooperative foraging by North American river otters, *Lutra canadensis*," Can. Field Nat. **4**, 458–459.
Similä, T., and Ugarte, F. (**1993**). "Surface and underwater observations of cooperatively feeding killer whales in northern Norway," Can. J. Zool. **71**, 1494–1499.
Traniello, J. (**1987**). "Comparative foraging ecology of north temperate ants—the role of worker size and cooperative foraging in prey selection," Insectes Soc. **34**, 118–130.
Würsig, B., and Würsig, M. (**1979**). "Behavior and ecology of bottlenose dolphin *Tursiops truncatus*, in the South Atlantic," Fish. Bull. **77**, 399–412.

Fuzzy structure theory modeling of sound-insulation layers in complex vibroacoustic uncertain systems: Theory and experimental validation

Charles Fernandez and Christian Soize[a)]
Laboratoire Modelisation et Simulation Multiechelle FRE3160 CNRS, Universite Paris-Est, 5 Boulevard Descartes, 77454 Marne-la-Vallee, France

Laurent Gagliardini
PSA Peugeot-Citroen, CTV, route de Gisy, 78943 Velizy-Villacoublay Cedex, France

(Received 29 January 2008; revised 11 September 2008; accepted 27 October 2008)

The fuzzy structure theory was introduced 20 years ago in order to model the effects of complex subsystems imprecisely known on a master structure. This theory was only aimed at structural dynamics. In this paper, an extension of that theory is proposed in developing an elastoacoustic element useful to model sound-insulation layers for computational vibroacoustics of complex systems. The simplified model constructed enhances computation time and memory allocation because the number of physical and generalized degrees of freedom in the computational vibroacoustic model is not increased. However, these simplifications introduce model uncertainties. In order to take into account these uncertainties, the nonparametric probabilistic approach recently introduced is used. A robust simplified model for sound-insulation layers is then obtained. This model is controlled by a small number of physical and dispersion parameters. First, the extension of the fuzzy structure theory to elastoacoustic element is presented. Second, the computational vibroacoustic model including such an elastoacoustic element to model sound-insulation layer is given. Then, a design methodology to identify the model parameters with experiments is proposed and is experimentally validated. Finally, the theory is applied to an uncertain vibroacoustic system. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3035827]

PACS number(s): 43.40.At [DF] Pages: 138–153

I. INTRODUCTION

This paper deals with the construction of a robust model of sound-insulation layers required by computational vibroacoustic simulation of complex systems in low- and medium-frequency ranges. The sound-insulation layer is assumed to behave as a resonant continuous dynamical system in the frequency band of interest. In this paper, we will not consider the high-frequency range where more elastic modes may appear. For such a modeling in the low- and medium-frequency ranges, a usual approach consists in modeling a sound-insulation layer as a poroelastic medium using the Biot theory; the finite element method is then classically used to solve the associated boundary value problem. In this case, both vibroacoustic system and sound-insulation layers are modeled by the finite element method (see, for instance, Refs. 1–7). When the first thickness eigenfrequencies belong to the frequency band of analysis, as assumed here, such a finite element model of sound-insulation layer introduces a large number of physical degrees of freedom (DOF) in the computational model as well as numerous elastic modes in the band. The size of the associated reduced computational model is then amply increased. For instance, in a car booming noise analysis (frequency range [20,200] Hz later referred as low-frequency range), the finite element model may involve up to 2×10^6 DOFs for the structure and the reduced model requires about 1000 elastic modes. If the sound-insulation layers were modeled by the finite element method, an additional number of about 5×10^6 DOFs would be necessary. Twenty thousand additional elastic modes then appear in the reduced computational model exceeding the limits of current computational resources. Consequently, there is a great interest to construct simplified sound-insulation layer models without adding neither physical DOF nor generalized DOF. Representing the sound-insulation layer by an adapted wall impedance can be a way to avoid the increase in DOF number (see, for instance, Refs. 8–12 for the notion of wall impedance in vibroacoustics). A great number of publications have been devoted to this subject in the past three decades. It should be noted that the largest part of these works deals with the medium- and high-frequency ranges, where the sound-insulation layer behavior differs from the lower-frequency range that is investigated here. Since the objective of this paper is not to give a state of the art on this particular topic, we simply refer to a few papers such as Refs. 13–23.

In this paper, an alternate construction to the finite element approach or to the usual wall impedances is presented. First, this construction does not increase the number of physical and generalized DOFs in the computational vibroacoustic model. Second, it does not involve the poroelastic equations because the sound-insulation layers considered here have a rather simple dynamic behavior, which does not require advanced material modeling. As explained above, in the frequency band of analysis, the simplified model can be

[a)]Electronic mail: christian.soize@univ-paris-est.fr

 0001-4966/2009/125(1)/138/16/$25.00

originated from a single DOF dynamical system. Due to the large variability of geometry, materials properties and connections to the master structure—mainly induced by the industrial process as well as vehicle diversity in the automotive industry, for example— the sound-insulation layer is considered as complex and, therefore, a statistical description of its internal dynamical DOF is proposed. We are naturally led to use the fuzzy structure theory, which fits this framework and has already been validated. A representation of the sound-insulation layer based on the fuzzy structure theory benefits from an understandable interpretation of complex dynamical systems' behavior. It is simply characterized by a few physical parameters: the participating mass, the modal density, and the internal damping rate. The fuzzy structure theory was introduced 20 ago in order to model the effects on a master structure of complex subsystems imprecisely known (see Refs. 24–28). This theory is developed in the context of the probability theory, which is well adapted to this kind of problem that carries many uncertainties (geometry, material, and boundary conditions). Many other works have been developed in this field, completing the initial construction (see Refs. 29–36). Most of these developments are related to complex structural subsystems coupled to a master vibroacoustic system. No attempt has been performed to develop a specific sound-insulation layer model using the fuzzy structure theory. The known results have to be extended in order to build an elastoacoustic element. That is the aim of this paper. However, such a simplified model introduces model uncertainties. In the original theory introduced in Ref. 24, the system parameters' uncertainties were already taken into account with a parametric probabilistic approach. Today, it is well understood that the parametric probabilistic approach can only address system parameters' uncertainties but not model uncertainties. Recently, a nonparametric probabilistic approach has been introduced (see Refs. 37–40), allowing both system parameters' uncertainties and model uncertainties to be taken into account in the computational model. The use of a nonparametric probabilistic approach to take into account model uncertainties in the sound-insulation layer simplified model constitutes a new extension of the fuzzy structure theory with respect to the model uncertainties' problems. In addition, a complete design methodology is experimentally validated; this two-step methodology can easily be implemented in the current engineering process of mechanical devices.

In order to help the understanding, we summarize below the modeling strategy used in this paper. As explained above, a complete three-dimensional modeling of the sound-insulation layer would consist in introducing a poroelastic medium and such an approach would lead us to introduce a large additional number of DOF in the computational model. A way to construct a simplified model could have consisted in introducing an empirical simplified model for which the identification would have been equivalent to a "curve fitting." Such a way is not used here. In opposite, the proposed simplified model is constructed using the fuzzy structure theory for which the main dynamical physical phenomena are captured and are taken into account. In the frequency band of interest, the major phenomena are due to the internal resonances of the sound-insulation layer. The fuzzy structure theory allows hidden dynamical DOF effects on the master structure to be taken into account in the sense of statistical averaging. With such a theory, the power flow between the fuzzy subsystem (the sound-insulation layer) and the master vibroacoustic system is mainly controlled by the resonances of the sound-insulation layer, by the participating dynamical mass, and by the internal damping. The corresponding parameters of the model are, as explained above, the mean modal density of the internal resonances, the coefficient of participating mass defined with respect to the total physical mass, and finally, the damping rate. It should be noted that the mean modal density, which strongly depends on the geometry and on the materials, is unknown but can be estimated with a usual computational model of the sound-insulation layer as proposed in this paper. Note that such computations are usually performed in industrial engineering processes. Such an estimation is then realistic for practical applications of complex systems. The total mass of the sound-insulation layer is known but the coefficient of participating mass is generally unknown for frequencies larger that the cut-off frequency (which is the known first eigenfrequency of the sound-insulation layer) and has to be identified, for instance, using experiments. This coefficient is less sensitive to the geometry than the modal density. In engineering practice, when this coefficient has been identified for given materials and a given thickness, the identified value can be reused for similar sound-insulation layers without performing a new identification. Clearly, an error is introduced but this error is taken into account by the probabilistic model of uncertainties. So, as soon as a class of sound-insulation layers has been identified, the model can be used for any geometry because for each fixed geometry of a sound-insulation layer, its cut-off frequency and modal density are computed with a usual finite element model. This methodology shows that the proposed model of the sound-insulation layer does not correspond to the curve fitting of an empirical model.

Section II is devoted to the construction of the sound-insulation layer simplified mean model using the fuzzy structure theory. As far as this model is part of a complex vibroacoustic system, the complete vibroacoustic model is presented. In this context, the nominal model will also be called the "mean model" and it refers to the deterministic model, which does not take into account uncertainties. Section III deals with the finite element approximation, which allows a computational vibroacoustic mean model to be constructed. In Sec. IV, we present the first step of the methodology performing the experimental identification of the mean parameters of the sound-insulation layer simplified model. Section V is devoted to the construction of the computational stochastic model for the uncertain vibroacoustic system including the uncertain sound-insulation layer. Some aspects relative to the stochastic solver are also presented in this section. Section VI addresses the second step of the methodology where the experimental identification of the dispersion parameters of the probabilistic model is performed. The identification of the dispersion parameters relative to the structure is performed using an experimental database, while

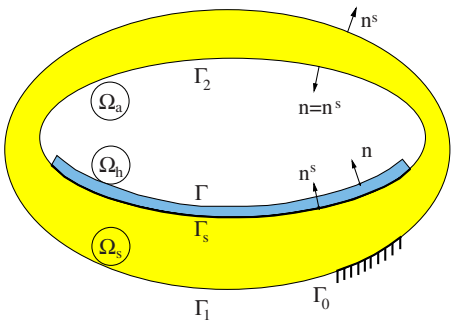


FIG. 1. (Color online) Structural-acoustical problem.

a reference numerical solution is used to identify the dispersion parameters relative to the probabilistic model of the sound-insulation layer. Finally, Sec. VII deals with the prediction of the system's vibroacoustic response using the previously identified computational stochastic model.

II. CONSTRUCTION OF A SIMPLIFIED MEAN MODEL OF THE SOUND-INSULATION LAYER USING THE FUZZY STRUCTURE THEORY IN A VIBROACOUSTIC SYSTEM

A. Definition of the vibroacoustic system

The physical space $\mathbb{R}^3$ is referred to a Cartesian system for which the generic point is denoted by $\boldsymbol{x}=(x_1,x_2,x_3)$. The Fourier transform with respect to time t is denoted by $u(\omega)=\int_{\mathbb{R}} e^{-i\omega t}u(t)dt$. The vibroacoustic system is analyzed in the frequency band $\mathbb{B}=[\omega_{\min},\omega_{\max}]$. The structure occupies a three-dimensional bounded domain Ω_s and is modeled by a nonhomogeneous anisotropic viscoelastic material. The boundary of Ω_s is written as $\partial\Omega_s=\Gamma_s\cup\Gamma_0\cup\Gamma_1\cup\Gamma_2$ (see Fig. 1) and the outward unit normal to $\partial\Omega_s$ is $\boldsymbol{n}^s(\boldsymbol{x})$. The structure is fixed on Γ_0, a surface force field $\boldsymbol{g}^{\text{surf}}(\boldsymbol{x},\omega)$ is given on Γ_1, and a body force field $\boldsymbol{g}^{\text{vol}}(\boldsymbol{x},\omega)$ is given on Ω_s. The acoustic cavity Ω_a is filled with a dissipative acoustic fluid, its boundary is written as $\partial\Omega_a=\Gamma\cup\Gamma_2$, and the inward unit normal to $\partial\Omega_a$ is $\boldsymbol{n}(\boldsymbol{x})$ [note that $\boldsymbol{n}(\boldsymbol{x})=\boldsymbol{n}^s(\boldsymbol{x})$ on Γ_2 and that without a sound-insulation layer, the coupling surface $\partial\Omega_a$ between the structure and the cavity would be $\Gamma_s\cup\Gamma_2$]. A sound-insulation layer, which occupies a bounded domain Ω_h with boundary $\partial\Omega_h=\Gamma\cup\Gamma_s$, is attached to the part Γ_s of the boundary of the structure (see Fig. 1). Let $\boldsymbol{x}\mapsto\boldsymbol{u}^s(\boldsymbol{x},\omega)=(u_1^s(\boldsymbol{x},\omega),u_2^s(\boldsymbol{x},\omega),u_3^s(\boldsymbol{x},\omega))$ be the structural displacement field defined on Ω_s with values in $\mathbb{C}^3$ and which is equal to zero on Γ_0. Let $\boldsymbol{x}\mapsto p(\boldsymbol{x},\omega)$ be the pressure field inside Ω_a with value in $\mathbb{C}$ for which the value on $\partial\Omega_a=\Gamma\cup\Gamma_2$ is still denoted by $p(\boldsymbol{x},\omega)$. Let $\boldsymbol{x}\mapsto\boldsymbol{u}^h(\boldsymbol{x},\omega)=(u_1^h(\boldsymbol{x},\omega),u_2^h(\boldsymbol{x},\omega),u_3^h(\boldsymbol{x},\omega))$ be the sound-insulation layer displacement field defined on Ω_h with values in $\mathbb{C}^3$ whose value on interface Γ is still denoted by $\boldsymbol{x}\mapsto\boldsymbol{u}^h(\boldsymbol{x},\omega)$. Finally, we need to introduce the admissible spaces for the three fields of the problem. Let $\mathcal{C}_0^s$ be the space of the admissible displacement fields of the structure such that $\boldsymbol{u}^s=0$ on Γ_0, let $\mathcal{C}^a$ be the space of the admissible pressure fields in the acoustic cavity, and $\mathcal{C}^h$ be the space of the admissible displacement fields of the sound-insulation layer.

B. Coupling force fields

The coupling force field on boundary Γ_s that the structure exerts on the sound-insulation layer is denoted by $\boldsymbol{x}\mapsto\boldsymbol{f}^s(\boldsymbol{x},\omega)=(f_1^s(\boldsymbol{x},\omega),f_2^s(\boldsymbol{x},\omega),f_3^s(\boldsymbol{x},\omega))$ and can be written for all $\boldsymbol{x}$ fixed in Γ_s as $\boldsymbol{f}^s(\boldsymbol{x},\omega)=f^s(\boldsymbol{x},\omega)\boldsymbol{n}^s(\boldsymbol{x})+\boldsymbol{f}_{\text{tang}}^s(\boldsymbol{x},\omega)$. It is assumed that the tangential force field $\boldsymbol{f}_{\text{tang}}^s(\boldsymbol{x},\omega)$ exerted by the structure on the sound-insulation layer is equal to zero. This hypothesis is reasonable in vibroacoustics for the majority of the cases met in the technologies such as the ones used in the automotive industry. Nevertheless, this hypothesis is not perfectly satisfied for real complex systems and induces uncertainties in the vibroacoustic model. This is a reason why a model of uncertainties will be introduced. Consequently, we have

$$\boldsymbol{f}^s(\boldsymbol{x},\omega)=f^s(\boldsymbol{x},\omega)\boldsymbol{n}^s(\boldsymbol{x}). \tag{1}$$

Note that dimension of $f_i^s(\boldsymbol{x},t)$ is $[M][L]^{-1}[T]^{-2}$. The coupling force field $\boldsymbol{x}\mapsto\boldsymbol{f}^p(\boldsymbol{x},\omega)=(f_1^p(\boldsymbol{x},\omega),f_2^p(\boldsymbol{x},\omega),f_3^p(\boldsymbol{x},\omega))$ on boundary $\partial\Omega_a=\Gamma\cup\Gamma_2$ that the acoustic fluid exerts on the structure (interface Γ_2) and the sound-insulation layer (interface Γ) is written as

$$\boldsymbol{f}^p(\boldsymbol{x},\omega)ds(\boldsymbol{x})=-p(\boldsymbol{x},\omega)\boldsymbol{n}(\boldsymbol{x})ds(\boldsymbol{x}), \tag{2}$$

in which ds is the surface element on $\partial\Omega_a$. The equations of the boundary value problem for the vibroacoustic system made up of the structure coupled with internal acoustic cavity and with the sound-insulation layer are given in Appendix A. Equations (A4)–(A6) are the equations for the structure, the acoustic cavity, and the sound-insulation layer. There are coupling terms in these three equations. In particular, the coupling term between the sound-insulation layer and the structure in Eq. (A4) of the structure is represented by the term $c_{\Gamma_s}(\boldsymbol{\delta u}^s;\omega)$; the coupling term between the sound-insulation layer and the acoustic cavity in Eq. (A5) of the acoustic cavity is represented by the term $c_\Gamma(\boldsymbol{u}^h,\delta p)$. The principle of construction of the simplified model for the sound-insulation layer consists in replacing Eq. (A6) with a simplified model obtained in using the fuzzy structure theory[24,26,27] for which a synthesis is given in Ref. 12. This means that the two coupling terms $c_{\Gamma_s}(\boldsymbol{\delta u}^s;\omega)$ and $c_\Gamma(\boldsymbol{u}^h,\delta p)$ have to be expressed as a function of $\boldsymbol{u}^s$ and p using the fuzzy structure theory. This theory consists in (1) introducing an underlying deterministic dynamical model (see Sec. II C), (2) introducing a probabilistic model of the eigenfrequencies of this dynamical model (see Sec. II D), and (3) performing a statistical averaging (see Sec. II E).

C. Definition of the underlying deterministic model for the fuzzy structure

We introduce the following hypothesis for the sound-insulation layer (see Fig. 2): The surfaces Γ and Γ_s are assumed to be geometrically equivalent and, consequently, for all $\boldsymbol{x}$ in $\Gamma_s\simeq\Gamma$, $\boldsymbol{n}^s(\boldsymbol{x})\simeq\boldsymbol{n}(\boldsymbol{x})$. The normal component to Γ_s of the structural displacement field is

$$\boldsymbol{n}^s(\boldsymbol{x})\cdot\boldsymbol{u}^s(\boldsymbol{x},\omega)=w^s(\boldsymbol{x},\omega),\quad \boldsymbol{x}\in\Gamma_s. \tag{3}$$

The normal component to Γ of the displacement field of the sound-insulation layer is

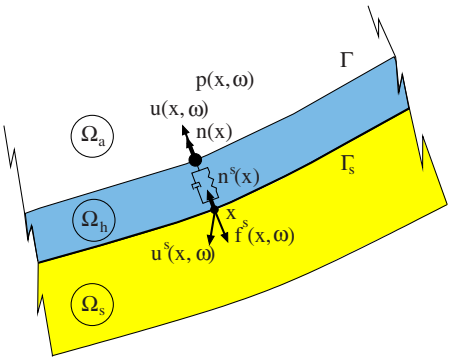


FIG. 2. (Color online) Sound-insulation layer modeling.

$$\boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{u}^h(\boldsymbol{x},\omega) = w(\boldsymbol{x},\omega), \tag{4}$$

in which $\boldsymbol{n}(\boldsymbol{x}) \simeq \boldsymbol{n}^s(\boldsymbol{x})$ for $\boldsymbol{x} \in \Gamma_s \simeq \Gamma$. Using the first step of the fuzzy structure theory and taking into account the hypothesis introduced in Sec. II B, the underlying deterministic model is made up of a density of damped linear oscillators acting in the normal direction to Γ. For a fixed frequency ω and for a fixed $\boldsymbol{x}$ in Γ_s, the displacement of the base of an oscillator is $w^s(\boldsymbol{x},\omega)$ and the displacement of its mass $\mu(\boldsymbol{x},\omega) > 0$ is $w(\boldsymbol{x},\omega)$. The mass density $\mu(\boldsymbol{x},\omega)$ ($[M][L]^{-2}$) is distributed on Γ. The corresponding stiffness density $k(\boldsymbol{x},\omega)$ associated with this oscillator is $k(\boldsymbol{x},\omega) = \mu(\boldsymbol{x},\omega)\omega_p^2(\boldsymbol{x},\omega)$ where $\omega_p(\boldsymbol{x},\omega) > 0$ is the eigenfrequency (rad s^{-1}) of the undamped linear oscillator with a fixed base. The damping rate of this oscillator is denoted by $\xi(\boldsymbol{x},\omega)$. Let $\boldsymbol{f}^s(\boldsymbol{x},\omega)$ be the force applied to the base of this oscillator and corresponding to the force density induced by the structure on the sound-insulation layer [see Eq. (1)]. Let $\boldsymbol{f}^p(\boldsymbol{x},\omega)$ be the force applied to the mass of the oscillator and corresponding to the force density induced by the acoustic pressure $p(\boldsymbol{x},\omega)$ on the sound-insulation layer [see Eq. (2)]. Removing $\boldsymbol{x}$ and ω for brevity, the equation of this oscillator is written as

$$\mu \begin{bmatrix} -\omega^2 + 2i\omega\xi\omega_p + \omega_p^2 & -2i\omega\xi\omega_p - \omega_p^2 \\ -2i\omega\xi\omega_p - \omega_p^2 & 2i\omega\xi\omega_p + \omega_p^2 \end{bmatrix} \times \begin{bmatrix} w(\boldsymbol{x},\omega) \\ w^s(\boldsymbol{x},\omega) \end{bmatrix} = \begin{bmatrix} -p(\boldsymbol{x},\omega) \\ f^s(\boldsymbol{x},\omega) \end{bmatrix}. \tag{5}$$

For all ω in $\mathbb{B}$, from Eq. (5), it can be deduced that

$$w(\boldsymbol{x},\omega) = a^c(\boldsymbol{x},\omega)w^s(\boldsymbol{x},\omega) + \frac{1}{\omega^2}a^a(\boldsymbol{x},\omega)p(\boldsymbol{x},\omega), \tag{6}$$

$$f^s(\boldsymbol{x},\omega) = a^s(\boldsymbol{x},\omega)w^s(\boldsymbol{x},\omega) + a^c(\boldsymbol{x},\omega)p(\boldsymbol{x},\omega), \tag{7}$$

in which

$$a^s(\boldsymbol{x},\omega) = \frac{-\omega^2\mu(\boldsymbol{x},\omega)(2i\omega\xi(\boldsymbol{x},\omega)\omega_p(\boldsymbol{x},\omega) + \omega_p(\boldsymbol{x},\omega)^2)}{-\omega^2 + 2i\omega\xi(\boldsymbol{x},\omega)\omega_p(\boldsymbol{x},\omega) + \omega_p(\boldsymbol{x},\omega)^2}, \tag{8}$$

$$a^a(\boldsymbol{x},\omega) = \frac{-\omega^2/\mu(\boldsymbol{x},\omega)}{-\omega^2 + 2i\omega\xi(\boldsymbol{x},\omega)\omega_p(\boldsymbol{x},\omega) + \omega_p(\boldsymbol{x},\omega)^2}, \tag{9}$$

$$a^c(\boldsymbol{x},\omega) = \frac{2i\omega\xi(\boldsymbol{x},\omega)\omega_p(\boldsymbol{x},\omega) + \omega_p(\boldsymbol{x},\omega)^2}{-\omega^2 + 2i\omega\xi(\boldsymbol{x},\omega)\omega_p(\boldsymbol{x},\omega) + \omega_p(\boldsymbol{x},\omega)^2}. \tag{10}$$

As explained in Sec. II B, we have to express the two terms $c_\Gamma(\boldsymbol{u}^h,\delta p)$ and $c_{\Gamma_s}(\boldsymbol{\delta u};\omega)$. Using $\Gamma \simeq \Gamma_s$, substituting Eq. (4) into $c_\Gamma(\boldsymbol{u}^h,\delta p)$ defined by Eq. (A2), substituting Eq. (6) again into Eq. (A2), and using Eq. (3) yield

$$\omega^2 c_\Gamma(\boldsymbol{u}^h,\delta p) = \omega^2 \int_{\Gamma_s} a^c(\boldsymbol{x},\omega)\boldsymbol{n}^s(\boldsymbol{x}) \cdot \boldsymbol{u}^s(\boldsymbol{x},\omega)\,\delta p(\boldsymbol{x}) ds(\boldsymbol{x}) + \int_{\Gamma_s} a^a(\boldsymbol{x},\omega)p(\boldsymbol{x},\omega)\,\delta p(\boldsymbol{x}) ds(\boldsymbol{x}). \tag{11}$$

Substituting Eq. (1) into $c_{\Gamma_s}(\boldsymbol{\delta u};\omega;\omega)$ defined by Eq. (A3), substituting Eq. (7) again into Eq. (A3), and using Eq. (3) yield

$$c_{\Gamma_s}(\boldsymbol{\delta u}^s;\omega) = \int_{\Gamma_s} a^s(\boldsymbol{x},\omega)(\boldsymbol{n}^s(\boldsymbol{x}) \cdot \boldsymbol{u}^s(\boldsymbol{x},\omega)) \times (\boldsymbol{n}^s(\boldsymbol{x}) \cdot \boldsymbol{\delta u}^s(\boldsymbol{x}))ds(\boldsymbol{x}) + \int_{\Gamma_s} a^c(\boldsymbol{x},\omega)p(\boldsymbol{x},\omega)(\boldsymbol{n}^s(\boldsymbol{x}) \cdot \boldsymbol{\delta u}^s(\boldsymbol{x}))ds(\boldsymbol{x}). \tag{12}$$

D. Probabilistic model of the eigenfrequency of the oscillators

The second step of the fuzzy structure consists in modeling $\omega_p(\boldsymbol{x},\omega)$ by a random variable $\Omega_p(\boldsymbol{x},\omega)$. In this section, we then introduce the random bilinear form associated with $c_\Gamma(\boldsymbol{u}^h,\delta p)$ and the random linear form associated with $c_{\Gamma_s}(\boldsymbol{\delta u}^s;\omega)$ defined by Eqs. (11) and (12). For all ω in $\mathbb{B}$, we choose to represent $\mu(\boldsymbol{x},\omega)$ and $\xi(\boldsymbol{x},\omega)$ by their mean values $\mu(\boldsymbol{x},\omega) = \underline{\mu}(\omega) > 0$ and $\xi(\boldsymbol{x},\omega) = \underline{\xi}(\omega)$ where $\omega \mapsto \underline{\mu}(\omega)$ and $\omega \mapsto \underline{\xi}(\omega)$ are two deterministic functions independent of $\boldsymbol{x}$ with $0 < \underline{\xi}(\omega) < 1$. The mean participating mass can be written[24,26,27] as $\underline{\mu}(\omega) = \underline{\nu}(\omega)m_{\text{tot}}/|\Gamma_s|$ where $0 \le \underline{\nu}(\omega) \le 1$ is the mean coefficient of participating mass, m_{tot} is the total mass of the density of oscillators, and $|\Gamma_s|$ is the measure of surface Γ_s. It should be noted that if there are several sound-insulation layers with different values of parameters $\underline{\mu}$ and $\underline{\xi}$, domain Ω_h is subdivised into several subdomains and, thus, their parameters have to be constant with respect to $\boldsymbol{x}$ in every subdomain. For all $\boldsymbol{x}$ fixed in Γ_s and ω fixed in $\mathbb{B}$, the eigenfrequency $\omega_p(\boldsymbol{x},\omega)$ is modeled by a positive random variable $\Omega_p(\boldsymbol{x},\omega)$ whose probability distribution $P_{\Omega_p(\boldsymbol{x};\omega)} \times (d\omega_p,\omega)$ is assumed to be independent of $\boldsymbol{x}$ and is defined by the probability density function $p_{\Omega_p(\boldsymbol{x};\omega)}(\omega_p,\omega)$ with respect to $d\omega_p$, such that[24,26]

$$p_{\Omega_p(\boldsymbol{x};\omega)}(\omega_p,\omega) = \ell(\omega)\mathbb{1}_{[a(\omega),b(\omega)]}(\omega_p), \tag{13}$$

with $\mathbb{1}_B(\boldsymbol{x}) = 1$ if $\boldsymbol{x} \in B$ and $=0$ if $\boldsymbol{x} \notin B$ and where

$$a(\omega) = \sup\left\{0, \omega - \frac{1}{2\underline{n}(\omega)}\right\}, \tag{14}$$

$$b(\omega) = \omega + \frac{1}{2\underline{n}(\omega)}, \tag{15}$$

$$\ell(\omega) = \frac{1}{b(\omega) - a(\omega)}, \tag{16}$$

in which $\underline{n}(\omega)$ is the mean modal density of the sound-insulation layer. In order to better explain the meaning of parameters $\underline{n}(\omega)$ and $\underline{\mu}(\omega)$, we define them in the simplest case for which the fuzzy structure Ω_h would be made up of N_{osc} oscillators uniformly distributed in the frequency band $\mathbb{B}$ and uniformly distributed on surface Γ_s. In this case, the mass of each oscillator would be m_{osc}. Consequently, the total mass of the fuzzy structure would be $m_{\text{tot}}=N_{\text{osc}}m_{\text{osc}}$ and for all $\boldsymbol{x}$ in Γ_s, we would have $\mu(\boldsymbol{x},\omega)=\underline{\mu}(\omega) = \sqrt{N_{\text{osc}}}m_{\text{osc}}/|\Gamma_s|$, $\int_{\mathbb{B}}\underline{n}(\omega)d\omega=\sqrt{N_{\text{osc}}}$, and $\int_{\mathbb{B}}\underline{\mu}(\omega)\underline{n}(\omega)d\omega = m_{\text{tot}}/|\Gamma_s|$. Coming back to the general case, for all $\boldsymbol{x}$ fixed in Γ_s and ω fixed in $\mathbb{B}$, the coefficients $a^s(\boldsymbol{x},\omega)$, $a^a(\boldsymbol{x},\omega)$, and $a^c(\boldsymbol{x},\omega)$ defined by Eqs. (8)–(10) become random variables denoted by $A^s(\boldsymbol{x},\omega)$, $A^a(\boldsymbol{x},\omega)$, and $A^c(\boldsymbol{x},\omega)$. For all $\boldsymbol{u}^s$ and $\boldsymbol{\delta u}^s$ in $\mathcal{C}_0^s$ and for all p and δp in $\mathcal{C}^a$, the forms $c_\Gamma(\boldsymbol{u}^h,\delta p)$ and $c_{\Gamma_s}(\boldsymbol{\delta u}^s;\omega)$ defined by Eqs. (11) and (12) become random variables, which are rewritten in terms of $\boldsymbol{u}^s, p, \boldsymbol{\delta u}^s$ and δp as $C_\Gamma(\boldsymbol{u}^s,p,\delta p;\omega)$ and $C_{\Gamma_s}(\boldsymbol{u}^s,p,\boldsymbol{\delta u}^s;\omega)$ and which are such that

$$\omega^2 C_\Gamma(\boldsymbol{u}^s,p,\delta p;\omega) = \omega^2\int_{\Gamma_s} A^c(\boldsymbol{x},\omega)\boldsymbol{n}^s(\boldsymbol{x})\,.\,\boldsymbol{u}^s(\boldsymbol{x},\omega)\,\delta p(\boldsymbol{x})ds(\boldsymbol{x}) + \int_{\Gamma_s} A^a(\boldsymbol{x},\omega)p(\boldsymbol{x},\omega)\,\delta p(\boldsymbol{x})ds(\boldsymbol{x}) \tag{17}$$

and

$$C_{\Gamma_s}(\boldsymbol{u}^s,p,\boldsymbol{\delta u}^s;\omega) = \int_{\Gamma_s} A^s(\boldsymbol{x},\omega)(\boldsymbol{n}^s(\boldsymbol{x})\,.\,\boldsymbol{u}^s(\boldsymbol{x},\omega))(\boldsymbol{n}^s(\boldsymbol{x})\,.\,\boldsymbol{\delta u}^s(\boldsymbol{x}))ds(\boldsymbol{x}) + \int_{\Gamma_s} A^c(\boldsymbol{x},\omega)p(\boldsymbol{x},\omega)(\boldsymbol{n}^s(\boldsymbol{x})\,.\,\boldsymbol{\delta u}^s(\boldsymbol{x}))ds(\boldsymbol{x}). \tag{18}$$

E. Statistical averaging and simplified mean model of the sound-insulation layer

The last step of the fuzzy structure theory consists in defining the simplified mean model taking the statistical averaging of random variables $C_\Gamma(\boldsymbol{u}^s,p,\delta p;\omega)$ and $C_{\Gamma_s}(\boldsymbol{u}^s,p,\boldsymbol{\delta u}^s;\omega)$ defined by Eqs. (17) and (18). As we have explained in Sec. II B, the simplified mean model thus consists in replacing the two coupling terms $c_\Gamma(\boldsymbol{u}^h,\delta p)$ and $c_{\Gamma_s}(\boldsymbol{\delta u}^s;\omega)$ with $\underline{c}_\Gamma(\boldsymbol{u}^s,p,\delta p;\omega)$ and $\underline{c}_{\Gamma_s}(\boldsymbol{u}^s,p,\boldsymbol{\delta u}^s;\omega)$ such that

$$\underline{c}_\Gamma(\boldsymbol{u}^s,p,\delta p;\omega) = \mathcal{E}\{C_\Gamma(\boldsymbol{u}^s,p,\delta p;\omega)\}, \tag{19}$$

$$\underline{c}_{\Gamma_s}(\boldsymbol{u}^s,p,\boldsymbol{\delta u}^s;\omega) = \mathcal{E}\{C_{\Gamma_s}(\boldsymbol{u}^s,p,\boldsymbol{\delta u}^s;\omega)\}, \tag{20}$$

in which $\mathcal{E}$ is the mathematical expectation. Analyzing Eqs. (17) and (18) leads us to introduce the following deterministic bilinear forms $b^s(\boldsymbol{u}^s,\boldsymbol{\delta u}^s)$ on $\mathcal{C}_0^s\times\mathcal{C}_0^s$, $c^s(p,\boldsymbol{\delta u}^s)$ on $\mathcal{C}^a\times\mathcal{C}_0^s$, and $b^a(p,\delta p)$ on $\mathcal{C}^a\times\mathcal{C}^a$,

$$b^s(\boldsymbol{u}^s,\boldsymbol{\delta u}^s) = \int_{\Gamma_s}(\boldsymbol{n}^s(\boldsymbol{x})\cdot\boldsymbol{u}^s(\boldsymbol{x}))(\boldsymbol{n}^s(\boldsymbol{x})\cdot\boldsymbol{\delta u}^s(\boldsymbol{x}))ds(\boldsymbol{x}), \tag{21}$$

$$c^s(p,\boldsymbol{\delta u}^s) = \int_{\Gamma_s} p(\boldsymbol{x})(\boldsymbol{n}^s(\boldsymbol{x})\cdot\boldsymbol{\delta u}^s(\boldsymbol{x}))ds(\boldsymbol{x}), \tag{22}$$

$$b^a(p,\delta p) = \int_{\Gamma_s} p(\boldsymbol{x})\,\delta p(\boldsymbol{x})ds(\boldsymbol{x}). \tag{23}$$

From Eqs. (17) and (18) and using Eqs. (19) and (20) with Eqs. (13)–(16) yield

$$\omega^2\underline{c}_\Gamma(\boldsymbol{u}^s,p,\delta p;\omega) = \omega^2\underline{a}^c(\omega)c^s(\delta p,\boldsymbol{u}^s) + \underline{a}^a(\omega)b^a(p,\delta p), \tag{24}$$

$$\underline{c}_{\Gamma_s}(\boldsymbol{u}^s,p,\boldsymbol{\delta u}^s;\omega) = \underline{a}^s(\omega)b^s(\boldsymbol{u}^s,\boldsymbol{\delta u}^s) + \underline{a}^c(\omega)c^s(p,\boldsymbol{\delta u}^s), \tag{25}$$

in which

$$\underline{a}^s(\omega) = -\omega^2\underline{a}^s_R(\omega) + i\omega\underline{a}^s_I(\omega), \tag{26}$$

$$\underline{a}^a(\omega) = \underline{a}^a_R(\omega) + i\omega\underline{a}^a_I(\omega), \tag{27}$$

$$\underline{a}^c(\omega) = \underline{a}^c_R(\omega) + i\underline{a}^c_I(\omega), \tag{28}$$

with

$$\underline{a}^s_R(\omega) = \underline{\mu}(\omega)\underline{n}(\omega)\left[\frac{1}{\underline{n}(\omega)} - \omega\underline{\lambda}(\omega)\Theta_R(\omega)\right], \tag{29}$$

$$\underline{a}^s_I(\omega) = \underline{\mu}(\omega)\underline{n}(\omega)\omega^2\underline{\lambda}(\omega)\Theta_I(\omega), \tag{30}$$

$$\underline{a}^a_R(\omega) = \omega\underline{n}(\omega)\frac{\underline{\lambda}(\omega)}{\underline{\mu}(\omega)}\Theta_R(\omega), \tag{31}$$

$$\underline{a}^a_I(\omega) = \underline{n}(\omega)\frac{\underline{\lambda}(\omega)}{\underline{\mu}(\omega)}\Theta_I(\omega), \tag{32}$$

$$\underline{a}^c_R(\omega) = 1 - \omega\underline{n}(\omega)\underline{\lambda}(\omega)\Theta_R(\omega), \tag{33}$$

$$\underline{a}^c_I(\omega) = -\omega\underline{n}(\omega)\underline{\lambda}(\omega)\Theta_I(\omega). \tag{34}$$

In these equations, the functions $\underline{\lambda}$, Θ_R, and Θ_I are defined in Appendix B. We then deduce the simplified mean model of the sound-insulation layer: Find $\boldsymbol{u}^s$ in $\mathcal{C}_0^s$ and p in $\mathcal{C}^a$ such that, for all $\boldsymbol{\delta u}^s$ in $\mathcal{C}_0^s$ and δp in $\mathcal{C}^a$, we have

$$-\omega^2 m^s(\boldsymbol{u}^s,\boldsymbol{\delta u}^s) + i\omega d^s(\boldsymbol{u}^s,\boldsymbol{\delta u}^s;\omega) + k^s(\boldsymbol{u}^s,\boldsymbol{\delta u}^s;\omega) + c_{\Gamma_2}(\boldsymbol{\delta u}^s,p) + \underline{a}^s(\omega)b^s(\boldsymbol{u}^s,\boldsymbol{\delta u}^s) + \underline{a}^c(\omega)c^s(p,\boldsymbol{\delta u}^s) = l^s(\boldsymbol{\delta u}^s;\omega), \tag{35}$$

$$-\omega^2 m^a(p,\delta p)+i\omega d^a(p,\delta p;\omega)+k^a(p,\delta p)+\omega^2 c_{\Gamma_2}(\boldsymbol{u}^s,\delta p)+\omega^2 \underline{a}^c(\omega)c^s(\delta p,\boldsymbol{u}^s)+\underline{a}^a(\omega)b^a(p,\delta p)=l^a(\delta p;\omega), \tag{36}$$

in which the bilinear forms b^s, c^s, and b^a are defined by Eqs. (21)–(23) and where m^s, d^s, k^s, m^a, d^a, k^a, and c_{Γ_2} are defined by Eqs. (A7)–(A13).

III. COMPUTATIONAL VIBROACOUSTIC MEAN MODEL

The finite element discretization[12,41] of Eqs. (35) and (36) yields the following matrix equation on $\mathbb{C}^{m_s}\times\mathbb{C}^{m_a}$ defined by Eqs. (17) and (18),

$$\begin{bmatrix} [\underline{\mathbb{A}}^s(\omega)]+\underline{a}^s(\omega)[\underline{\mathbb{B}}^s] & [\underline{\mathbb{C}}]+\underline{a}^c(\omega)[\underline{\mathbb{C}}^s] \\ \omega^2\{[\underline{\mathbb{C}}]^T+\underline{a}^c(\omega)[\underline{\mathbb{C}}^s]^T\} & [\underline{\mathbb{A}}^a(\omega)]+\underline{a}^a(\omega)[\underline{\mathbb{B}}^a] \end{bmatrix} \times \begin{bmatrix} \underline{\boldsymbol{u}}^s(\omega) \\ \underline{\boldsymbol{p}}(\omega) \end{bmatrix} = \begin{bmatrix} \underline{\mathbb{f}}^s(\omega) \\ \underline{\mathbb{f}}^a(\omega) \end{bmatrix}, \tag{37}$$

where $[\underline{\mathbb{A}}^s(\omega)]$ is a complex $(m_s\times m_s)$ matrix such that

$$[\underline{\mathbb{A}}^s(\omega)]=-\omega^2[\underline{\mathbb{M}}^s]+i\omega[\underline{\mathbb{D}}^s(\omega)]+[\underline{\mathbb{K}}(\omega)], \tag{38}$$

in which $[\underline{\mathbb{M}}^s]$, $[\underline{\mathbb{D}}^s(\omega)]$, and $[\underline{\mathbb{K}}^s(\omega)]$ are the mass, damping, and stiffness matrices of the structure *in vacuo*. In Eq. (37), $[\underline{\mathbb{A}}^a(\omega)]$ is a complex $(m_a\times m_a)$ matrix,

$$[\underline{\mathbb{A}}^a(\omega)]=-\omega^2[\underline{\mathbb{M}}^a]+i\omega[\underline{\mathbb{D}}^a(\omega)]+[\underline{\mathbb{K}}^a], \tag{39}$$

in which $[\underline{\mathbb{M}}^a]$, $[\underline{\mathbb{D}}^a(\omega)]$, and $[\underline{\mathbb{K}}^a]$ are the "mass," "damping," and "stiffness" matrices of the acoustic cavity with a rigid wall. The real $(m_s\times m_a)$ matrix $[\underline{\mathbb{C}}]$ is the usual vibroacoustic coupling matrix relative to boundary Γ_2 (which is without a sound-insulation layer). The matrices $[\underline{\mathbb{B}}^s]$, $[\underline{\mathbb{C}}^s]$, and $[\underline{\mathbb{B}}^a]$ correspond to the finite element approximation of the bilinear forms defined by Eqs. (21)–(23), respectively. Using n_s structural elastic modes *in vacuo* and n_a acoustic modes of the cavity with rigid walls including the constant pressure mode, the mean reduced matrix model of the vibroacoustic system is written as

$$\underline{\boldsymbol{u}}^s(\omega)=[\underline{\Phi}^s]\underline{\boldsymbol{q}}^s(\omega),\quad \underline{\boldsymbol{p}}(\omega)=[\underline{\Phi}^a]\underline{\boldsymbol{q}}^a(\omega), \tag{40}$$

$$\begin{bmatrix} [\underline{A}^s(\omega)]+\underline{a}^s(\omega)[\underline{B}^s] & [\underline{C}]+\underline{a}^c(\omega)[\underline{C}^s] \\ \omega^2\{[\underline{C}]^T+\underline{a}^c(\omega)[\underline{C}^s]^T\} & [\underline{A}^a(\omega)]+\underline{a}^a(\omega)[\underline{B}^a] \end{bmatrix} \times \begin{bmatrix} \underline{\boldsymbol{q}}^s(\omega) \\ \underline{\boldsymbol{q}}^a(\omega) \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{f}}^s(\omega) \\ \underline{\boldsymbol{f}}^a(\omega) \end{bmatrix}, \tag{41}$$

in which $[\underline{\Phi}^s]$ is the $(m_s\times n_s)$ real matrix of the structural modes, $[\underline{\Phi}^a]$ is the $(m_a\times n_a)$ real matrix of the acoustic modes, $[\underline{C}]$ is the $(n_s\times n_a)$ generalized vibroacoustic coupling matrix, $\underline{a}^s(\omega)[\underline{B}^s]$ is the $(n_s\times n_s)$ generalized matrix, $\underline{a}^c(\omega)[\underline{C}^s]$ is the $(n_s\times n_a)$ generalized matrix, and $\underline{a}^a(\omega)[\underline{B}^a]$ is the $(n_a\times n_a)$ generalized matrix corresponding to the vibroacoustic coupling induced by the sound-insulation layer. The $(n_s\times n_s)$ matrix $[\underline{A}^s(\omega)]$ and the $(n_a\times n_a)$ matrix $[\underline{A}^a(\omega)]$ are written as

$$[\underline{A}^s(\omega)]=-\omega^2[\underline{M}^s]+i\omega[\underline{D}^s(\omega)]+[\underline{K}^s(\omega)], \tag{42}$$

$$[\underline{A}^a(\omega)]=-\omega^2[\underline{M}^a]+i\omega[\underline{D}^a(\omega)]+[\underline{K}^a], \tag{43}$$

in which $[\underline{M}^s]$, $[\underline{D}^s(\omega)]$, and $[\underline{K}^s(\omega)]$ are the generalized mass, damping, and stiffness matrices of the structure and $[\underline{M}^a]$, $[\underline{D}^a(\omega)]$, and $[\underline{K}^a]$ are the generalized mass, damping, and stiffness matrices of the acoustic cavity.

IV. EXPERIMENTAL IDENTIFICATION OF THE MEAN PARAMETERS OF THE FUZZY STRUCTURE MODEL FOR THE SOUND-INSULATION LAYER: DESIGN METHODOLOGY PART 1

We propose to validate the simplified mean model of the sound-insulation layer by using experiments. The methodology used is the following.

(1) We consider a structure for which the experimental frequency response functions (FRFs) are measured on frequency band $\mathbb{B}$.
(2) A sound-insulation layer is laid on this structure and the experimental FRFs are measured again for the structure coupled with the sound-insulation layer.
(3) A mean computational model of the structure is developed and the model is updated using the experimental FRFs measured in point (1) above.
(4) A mean computational model of the structure coupled with the sound-insulation layer is developed using the updated mean computational model of the structure and the simplified mean model of the sound-insulation layer. This simplified model provided by the fuzzy structure theory depends on unknown parameters $\underline{\xi}(\omega)$, $n(\omega)$, and $\underline{\nu}(\omega)$ that we propose to identify using the experimental FRFs measured in point (2) above. The following methodology is carried out.

(i) Over all the frequency band $\mathbb{B}$, the mean damping rate $\underline{\xi}$ of the fuzzy part is assumed to be independent of ω and is fixed to its estimated value corresponding to the damping rate at the first experimental eigenfrequency.
(ii) The mean modal density $\underline{n}(\omega)$ is obtained by performing a modal analysis with a very fine mesh finite element model of the sound-insulation layer embedded on its base. We then calculate the mean number $\underline{N}(\omega)$ of eigenfrequencies in the frequency band $[0,\omega]$ and then, by a numerical derivative, we deduce the mean modal density $\underline{n}(\omega)$, which is such that $\underline{N}(\omega)=\int_0^\omega \underline{n}(\alpha)d\alpha$. The first eigenfrequency for which the mean modal density is different from zero is defined as the cut-off frequency Ω_C. This cut-off frequency can be viewed as the frequency for which the sound-insulation layer begins to act as a power flow transmitter due to its own thickness resonances (the internal dynamical resonances taken into account by the fuzzy structure theory).
(iii) For $\omega<\Omega_C$, the coefficient of participating mass $\underline{\nu}(\omega)$ is equal to 1 and the mean modal density $\underline{n}(\omega)$ is equal to 0. The three coefficients $\underline{a}^s(\omega)$, $\underline{a}^a(\omega)$, and $\underline{a}^c(\omega)$ [defined by Eqs. (26)–(28)] are then taken as $\underline{a}^s(\omega)=-\underline{\mu}(\omega)\omega^2$, $\underline{a}^a(\omega)=0$, and $\underline{a}^c(\omega)=1$.
(iv) For $\omega\geq\Omega_C$, the parameter $\nu(\omega)$ is either experimentally

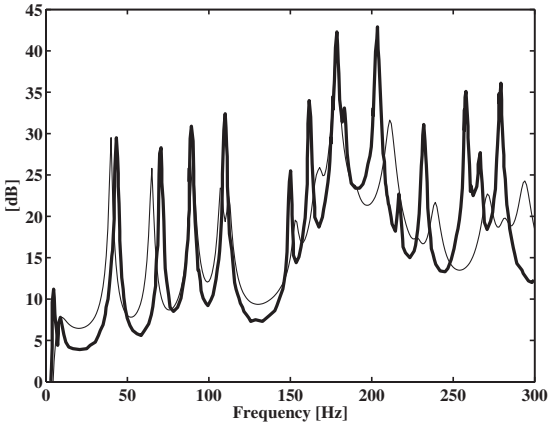


FIG. 3. Structure without the sound-insulation layer: Graphs of $\omega \mapsto r(\omega)$ (thin solid line) and $\omega \mapsto r^{\exp}(\omega)$ (thick solid line).

identified by solving an inverse problem, which is formulated as an optimization problem or is taken from a previous identified database.

A. Experimental configuration and measurements

The experimental configuration is made up of a homogeneous, isotropic, and slightly damped thin plate (steel plate with a constant thickness) connected to an elastic framework on its edges. This structure is set horizontally and is hung up by four soft springs in order to avoid rigid body modes. The highest eigenfrequency of suspension is 9 Hz while the lowest elastic mode of the structure is 43 Hz. The excitation is a point force applied to the framework and excites the dynamical system mainly in bending mode in the frequency band of analysis $\mathbb{B}=]0,300]$ Hz. The number of sampling frequencies is $n_{\text{freq}}=300$. The frequency resolution is $\Delta f=1$ Hz. Sixty FRFs are performed for this structure. The FRFs $\omega \mapsto \gamma_i^{\exp}(\omega)$ are identified on frequency band $\mathbb{B}$ for $n_{\text{obs}}=60$ normal accelerations in the plate measured with a laser velocimeter. We then construct the following experimental FRF: $\omega \mapsto r^{\exp}(\omega)=10\log_{10}(\Sigma_{i=1}^{n_{\text{obs}}}|\gamma_i^{\exp}(\omega)|^2)$. This FRF has been chosen because it gives a robust and simple way to obtain the dynamical behavior of all the structure.

B. Experimental updating of the mean model of the structure without a sound-insulation layer

The mean computational model of the structure is made up of a finite element model having $m_s=57\,768$ structural DOF. The reduced mean computational model is constructed with $n_s=240$ structural modes. The mean computational model has been updated with respect to the Young modulus, the mass density, and the damping rate of the plate and of the elastic framework using the experimental response of the system. The updated mean computational model will simply be called the mean computational model. We introduce the function $\omega \mapsto r(\omega)=10\log_{10}(\Sigma_{i=1}^{n_{\text{obs}}}|\gamma_i(\omega)|^2)$. Figure 3 shows the comparison between the experimental measurements and the updated FRF of the mean computational model for the structure without the sound-insulation layer. We made the choice to update the mean computational model as follows. The level of the FRF has been updated at its experimental value of 29.9 dB at the first eigenfrequency. However, after such an updating, the first eigenfrequency given by the model stays at 42 Hz while the experimental value is at 43 Hz. We have also updated the first torsion eigenfrequency, which is at 178 Hz (due to the framework) and nevertheless, the FRF level at this frequency is 37 dB while the experimental value is 42 dB. This strategy of updating has been chosen because the first elastic mode and the torsion of the structure lightly depend on the sound-insulation layer. The other eigenfrequencies and the corresponding amplitudes of the responses have not specifically been updated. This is a source of uncertainties, which is addressed below.

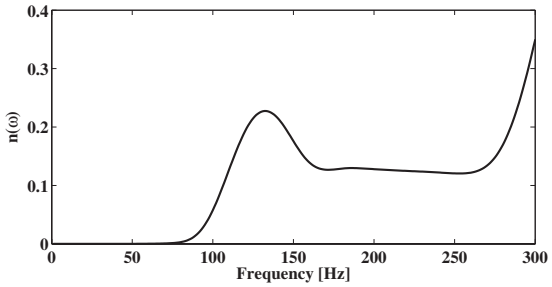


FIG. 4. Graph of $\omega \mapsto \underline{n}(\omega)$.

C. Experimental identification of the parameters of the simplified model of the sound-insulation layer using a design methodology

A similar experimentation to the experiments described in Sec. IV B has been carried out when the structure is coupled with the sound-insulation layer, which is made up of a heterogeneous, anisotropic, poroelastic foam and of a heavily damped septum (loaded polymer). The sound-insulation layer is laid on the plate and is not connected to the elastic framework. The reduced mean computational model is written [see Eq. (41)] as $[\underline{A}^s(\omega)+\underline{a}^s(\omega)[\underline{B}^s]]\underline{\boldsymbol{q}}^s(\omega)=\underline{\boldsymbol{f}}^s(\omega)$. We use this reduced mean computational model to identify the three parameters $\underline{\xi}$, $\underline{n}(\omega)$, and $\underline{\nu}(\omega)$ of the simplified model of the sound-insulation layer for ω in $\mathbb{B}$. The methodology used is the following.

(1) As previously explained, the mean modal density is calculated using a refined finite element model of the sound-insulation layer (33 210 DOFs for the foam and 13 284 DOFs for the septum; there are $N=1900$ elastic modes in the frequency band]0,450] Hz). The cut-off frequency Ω_C and the mean modal density $\underline{n}(\omega)$ are then deduced in the frequency band $\mathbb{B}$. We obtain $\Omega_C=67\times 2\pi$ rad s^{-1} and the graph of the smoothed function $\omega \mapsto \underline{n}(\omega)$ is given in Fig. 4 for frequency band $\mathbb{B}$. It should be noted that the modal density increases in the frequency band [250,300] Hz.

(2) In the frequency band $[\omega_{\min},\Omega_C[$, we remind that the sound-insulation layer is equivalent to an added mass. The expression for the fuzzy coefficient $\underline{a}^s(\omega)$ is then $\underline{a}^s(\omega)=-\underline{\mu}\omega^2$ for $\omega<\Omega_C$. As explained in point (i) above, $\underline{\xi}$ is experimentally identified as the damping rate

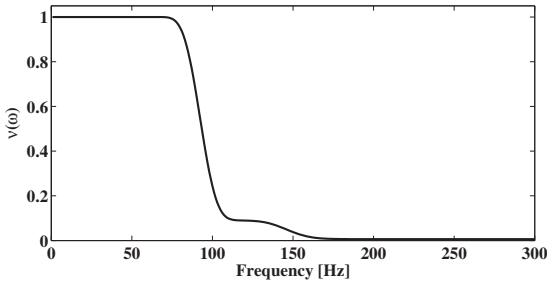


FIG. 5. Graph of $\omega \mapsto \underline{\nu}(\omega)$.

of the first elastic eigenmode of the structure with the sound-insulation layer. The identified values are $\underline{\xi}=0.01$ and $\underline{\mu}=5.9$ kg m^{-2}.

(3) We have chosen to directly identify the function $\omega \mapsto \nu(\omega)$ as explained in points (iii) and (iv) above. The optimization problem consists in minimizing the distance between the model and the experiments for the FRF $\omega \mapsto r(\omega;\nu(\omega))=10\log_{10}(\Sigma_{i=1}^{n_{\text{obs}}}|\gamma_i(\omega;\nu(\omega))|^2)$.

(4) In the frequency band $[\omega_{\min};\Omega_C[$, we have $\underline{\nu}(\omega)=1$ and the modal density $\underline{n}(\omega)$, which is equal to zero. In the frequency band $[\Omega_C,\omega_{\max}]$, we use the calculated mean modal density $\underline{n}(\omega)$ and the mean damping rate $\underline{\xi}$ in order to identify the mean coefficient of the participating mass $\underline{\nu}(\omega)$ as explained in point (iv) above. Figure 5 displays the updated mean coefficient $\underline{\nu}(\omega)$ of the participating mass and Fig. 6 shows the comparison between the experimental FRF and the FRF of the mean computational model for the structure with the sound-insulation layer calculated with the parameters experimentally identified.

(5) The use of the structural part of the vibroacoustic model allows the identification of the mean parameters $\underline{\nu}(\omega)$, $\underline{\xi}$, and $\underline{n}(\omega)$. It should be noted that this identification allows not only the coefficient $\underline{a}^s(\omega)$ to be identified but also the coefficients $\underline{a}^c(\omega)$ and $\underline{a}^a(\omega)$. Therefore, we only need the structural part of the vibroacoustic model to identify all of the mean computational simplified model of the sound-insulation layer (i.e., the structural, the coupling, and the acoustic parts).

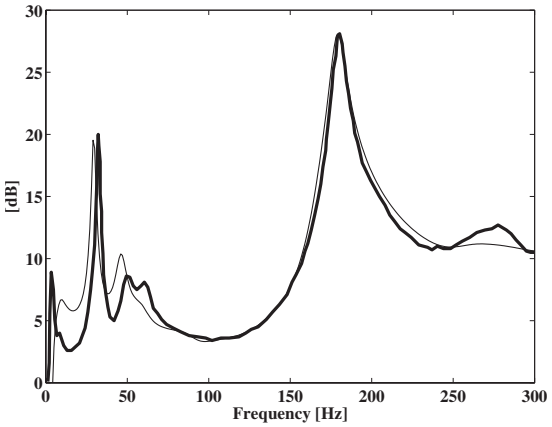


FIG. 6. System with a sound-insulation layer; graphs of $\omega \mapsto r(\omega;\underline{n}(\omega);\underline{\nu}(\omega))$ (thin solid line) and $\omega \mapsto r^{\exp}(\omega)$ (thick solid line).

V. COMPUTATIONAL STOCHASTIC MODEL FOR THE UNCERTAIN VIBROACOUSTIC SYSTEM INCLUDING THE SOUND-INSULATION LAYER

A. Stochastic vibroacoustic system

As we have explained in Sec. I, there are two sources of uncertainties. The first source is relative to the mean computational model of the structure without the sound-insulation layer coupled with the acoustic cavity. Concerning the structure *in vacuo*, Fig. 3 shows that the mean computational model gives a good enough prediction but some significant differences exist between the prediction and the experimental results. Consequently, the quality of the predictions of the mean computational model can be improved in implementing a nonparametric probabilistic model of uncertainties to take into account both system parameters' uncertainties and model uncertainties. Therefore, (1) the mean generalized matrices $[\underline{M}^s]$, $[\underline{D}^s(\omega)]$, and $[\underline{K}^s(\omega)]$ in Eq. (42) relative to the structure are replaced with the random matrices $[\mathbf{M}^s]$, $[\mathbf{D}^s(\omega)]$, and $[\mathbf{K}^s(\omega)]$; (2) the mean generalized matrices $[\underline{M}^a]$, $[\underline{D}^a(\omega)]$, and $[\underline{K}^a]$ in Eq. (43) relative to the acoustic cavity are replaced with the random matrices $[\mathbf{M}^a]$, $[\mathbf{D}^a(\omega)]$, and $[\mathbf{K}^a]$; and (3) the mean generalized coupling matrix $[\underline{C}]$ is replaced with the random matrix $[\mathbf{C}]$. The level of uncertainties of these random matrices is controlled by the dispersion parameters δ_{M^s}, δ_{D^s}, δ_{K^s}, δ_{M^a}, δ_{D^a}, δ_{K^a}, and δ_C, which are independent of the matrix dimension and of the frequency. The second source of uncertainties is introduced by the use of the simplified model of the sound-insulation layer based on the fuzzy structure theory. In fact, if this simplified model has the capability to predict in mean the effects of the sound-insulation layer on the structure and on the acoustic cavity, this simplified model does not describe the fluctuations around the mean value induced by these uncertainties. We then propose to use again the nonparametric probabilistic approach to take into account uncertainties in the simplified model. Consequently, the matrices $[\underline{B}^s]$, $[\underline{C}^s]$, and $[\underline{B}^a]$ of the mean simplified model of the sound-insulation layer are replaced with the random matrices $[\mathbf{B}^s]$, $[\mathbf{C}^s]$, and $[\mathbf{B}^a]$. The level of uncertainties of these random matrices is controlled by the dispersion parameters δ_{B^s}, δ_{C^s}, and δ_{B^a}, which are independent of the matrices' dimension and of the frequency. The development of the construction of the probability model of all these random matrices with a nonparametric approach will not be detailed here. Such an approach is presented in Refs. 37 and 40 and its application to a structural-acoustic system without a sound-insulation layer can be found in Refs. 42 and 43. However, some details are given in Appendix D. Taking into account Eq. (40), the vectors $\mathbf{U}^s(\omega)$ of the random structural displacement and $\mathbf{P}(\omega)$ of the random acoustic fluid pressure are written as

$$\mathbf{U}^s(\omega)=[\underline{\Phi}^s]\mathbf{Q}^s(\omega), \quad \mathbf{P}(\omega)=[\underline{\Phi}^a]\mathbf{Q}^a(\omega). \tag{44}$$

The stochastic reduced computational vibroacoustic model can then be deduced and is written as, for all $\omega \in \mathbb{B}$,

$$\begin{bmatrix} [\mathbf{A}^s(\omega)] + \underline{a}^s(\omega)[\mathbf{B}^s] & [\mathbf{C}] + \underline{a}^c(\omega)[\mathbf{C}^s] \\ \omega^2\{[\mathbf{C}]^T + \underline{a}^c(\omega)[\mathbf{C}^s]^T\} & [\mathbf{A}^a(\omega)] + \underline{a}^a(\omega)[\mathbf{B}^a] \end{bmatrix} \times \begin{bmatrix} \mathbf{Q}^s(\omega) \\ \mathbf{Q}^a(\omega) \end{bmatrix} = \begin{bmatrix} \underline{f}^s(\omega) \\ \underline{f}^a(\omega) \end{bmatrix}, \quad (45)$$

in which the random vectors $\mathbf{Q}^s(\omega)$ and $\mathbf{Q}^a(\omega)$ are the solution of the stochastic computational model and where the random matrices $[\mathbf{A}^s(\omega)]$ and $[\mathbf{A}^a(\omega)]$ are defined by Eqs. (D2) and (D3).

B. Solver for the stochastic reduced computational model

This section is devoted to the construction of the random solution of the stochastic computational model defined by Eq. (45). The stochastic solver is based on the use of the Monte Carlo method. The methodology used is then the following. For ω fixed in $\mathbb{B}$, (1) n_r independent realizations of the random variables $\mathbf{U}^s(\omega) = (U_1^s(\omega), \ldots, U_{m_s}^s(\omega))$ and $\mathbf{P}(\omega) = (P_1(\omega), \ldots, P_{m_a}(\omega))$ are constructed. For each realization $\mathbf{U}^s(\omega;\theta_\ell)$ and $\mathbf{P}(\omega;\theta_\ell)$ of the random vectors $\mathbf{U}^s(\omega)$ and $\mathbf{P}(\omega)$, the realizations of $R^s(\omega;\theta_\ell) = 10\log_{10}(\Sigma_{i=1}^{n_{\text{obs}}}|\omega^2 U_{k_i}^s(\omega;\theta_\ell)|^2)$ and $R^a(\omega;\theta_\ell) = 10\log_{10}(\Sigma_{i=1}^{m_{\text{obs}}}|P_{k_i}(\omega;\theta_\ell)|^2)$ are calculated in which m_{obs} is the number of pressure observations. (2) A convergence analysis is performed with respect to the number n_r of realizations and to the numbers n_s and n_a of modes. The order statistics and the method of quantiles (see Ref. 44) are used to construct an estimation of the confidence regions of the observations.

C. Convergence analysis

For fixed values of the dispersion parameters, the convergence analysis with respect to n_r, n_s, and n_a is carried out in studying the function

$$(n_r, n_s, n_a) \mapsto \text{Conv}^j(n_r, n_s, n_a) = \frac{1}{n_r}\sum_{\ell=1}^{n_r} \|\mathbf{Q}^j(\omega,\theta_\ell)\|^2, \quad (46)$$

with $j = s$ or a.

D. Estimation of the mean value and of the confidence region

Let $W(\omega)$ be the random variable representing $R^s(\omega)$ or $R^a(\omega)$. Let $F_{W(\omega)}(w)$ be the cumulative distribution function (continuous from the right) of the random variable $W(\omega)$, such that $F_{W(\omega)}(w) = P(W(\omega) \le w)$. For $0 < p < 1$, the pth quantile or fractile of $F_{W(\omega)}$ is defined as

$$\zeta(p) = \inf_{F_{W(\omega)}(w) \ge p} \{w\}. \quad (47)$$

Then, the upper envelope $w_+(\omega)$ and the lower envelope $w_-(\omega)$ of the confidence region are defined by

$$w_+(\omega) = \zeta\left(\frac{1+P_c}{2}\right), \quad w_-(\omega) = \zeta\left(\frac{1-P_c}{2}\right), \quad (48)$$

and are such that

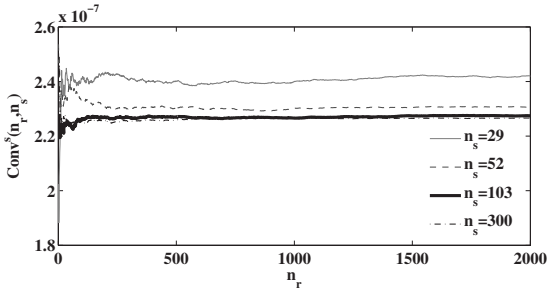


FIG. 7. Graph of $(n_r, n_s) \mapsto \text{Conv}^s(n_r, n_s)$.

$$\mathcal{P}(w_-(\omega) < W(\omega) \le w_+(\omega)) = P_c. \quad (49)$$

The estimation of $w_+(\omega)$ and $w_-(\omega)$ is performed by using the sample quantiles.[44] Let $w_1(\omega) = W(\omega;\theta_1), \ldots, w_{n_r}(\omega) = W(\omega;\theta_{n_r})$ be the n_r independent realizations of the random variable $W(\omega)$ computed as explained in Sec. V. Let $\tilde{w}_1(\omega) < \cdots < \tilde{w}_{n_r}(\omega)$ be the order statistics associated with $w_1(\omega), \ldots, w_{n_r}(\omega)$. Therefore, one has the following estimation:

$$w_+(\omega) \simeq \tilde{w}_{j_+}(\omega), \quad j_+ = \text{fix}(n_r(1+P_c)/2), \quad (50)$$

$$w_-(\omega) \simeq \tilde{w}_{j_-}(\omega), \quad j_- = \text{fix}(n_r(1-P_c)/2), \quad (51)$$

where $\text{fix}(z)$ is the integer part of the real number z.

VI. IDENTIFICATION OF THE DISPERSION PARAMETERS OF THE PROBABILISTIC MODELS USING EXPERIMENTS: DESIGN METHODOLOGY PART 2

A. Convergence of the stochastic solver

In this section, we perform the experimental identification of the dispersion parameters δ_{M_s}, δ_{D_s}, and δ_{K_s} for the structure and δ_{B_s} for the sound-insulation layer. This identification is carried out using the stochastic reduced computational model $[\mathbf{A}^s(\omega) + \underline{a}^s(\omega)[\mathbf{B}^s]]\mathbf{Q}^s(\omega) = \underline{f}^s(\omega)$ for the structure coupled with the sound-insulation layer and not coupled with the acoustic cavity [see Eq. (45)]. Consequently, for the largest possible values fixed at 0.8 of the dispersion parameters δ_{M_s}, δ_{D_s}, δ_{K_s}, and δ_{B_s}, we need to calculate the values of the parameters n_r and n_s in order that the mean-square convergence is reached, i.e., in studying the function $(n_r, n_s) \mapsto \text{Conv}^s(n_r, n_s)$. The graph of this function is shown in Fig. 7. It can be deduced that the convergence is reasonably reached for $n_s = 103$ modes and $n_r = 800$ realizations. Below, in the identification procedure presented, the parameters n_r and n_s are fixed to these values.

B. Description of the dispersion parameters' identification

In a first step, we use the experimental configuration (described in Sec. IV A) in order to identify the dispersion parameters δ_{M_s}, δ_{D_s}, and δ_{K_s} for the structure. The value of the damping dispersion parameter δ_{D_s} is fixed *a priori* to $\delta_{D_s} = 0.3$ according to the conclusion of Ref. 42. In order to verify that the random response is not really sensitive[42] to the

value of the parameter of dispersion δ_{D_s}, we have performed a sensitivity analysis with respect to δ_{D_s} varying in the interval [0.2,0.4] where δ_{M_s} and δ_{K_s} are fixed to the value 0.1 (small value of the dispersion parameter for the structural mass and stiffness matrices). With this sensitivity analysis, we have effectively verified that the influence of this dispersion parameter is negligible. Concerning the identification of the dispersion parameters δ_{M_s} and δ_{K_s}, we use the maximum likelihood method with a statistical reduction method[45] (the method is recalled in Sec. VI C) and we use[42,43] the assumption $\delta = \delta_{M_s} = \delta_{K_s}$. For this first step, Eq. (45) is then replaced with the random equation $[\mathbf{A}^s(\omega)]\mathbf{Q}^s(\omega) = \underline{f}^s(\omega)$ relative to the uncertain structure without a sound-insulation layer and without coupling with the acoustic cavity.

The second step is devoted to the identification of the dispersion parameter δ_{B_s} relative to the stochastic simplified model of the sound-insulation layer. Numerical simulations have shown that the sensitivity of the response of the structure coupled with the sound-insulation layer is smaller than the sensitivity of the response induced by the dispersion parameters of the structure. Consequently, the identification of δ_{B_s} cannot be carried out with the uncertain structure. Parameter δ_{B_s} must thus be identified with a "reference structure" for which there are no uncertainties (note that the sound-insulation layer cannot be analyzed alone and has to be coupled with a structure). The methodology proposed consists in the following.

(1) Defining a reference structure and analyzing the response of this reference structure coupled with the sound-insulation layer. This reference coupling system is analyzed by the finite element method using a fine mesh for the sound-insulation layer and the reference structure. This deterministic computational model allows the responses to be computed. These responses are defined below as the "numerical experiments." Note that this computational model does not represent the experimental configuration, but this choice is completely coherent because the stochastic simplified model of the sound-insulation layer is independent of the choice of the structure. This model is constituted of a thin plate similar to the plate of the experimental configuration presented in Sec. IV B.
(2) Constructing a stochastic computational model constituted of the computational model above for the reference structure and of the stochastic simplified model for the sound-insulation layer, which depends on δ_{B_s}. The sound-insulation layer model is similar to the one presented in Sec. IV B. For this second step, Eq. (45) is then replaced with the random equation $[\underline{A}^s_{\text{ref}}(\omega) + \underline{a}^s(\omega)[\mathbf{B}^s]]\mathbf{Q}^s(\omega) = \underline{f}^s(\omega)$.

The method then consists in minimizing the distance between the numerical experiments and the response of the stochastic reduced computational model. We then use again the maximum likelihood method with a statistical reduction.

C. Statistical reduction and maximum likelihood method

An efficient method has recently been proposed[45] to identify the dispersion parameters with the maximum likelihood method for a stochastic process such as the modulus of a FRF indexed by the frequency ω. This method consists in introducing a statistical reduction of the data and then applying the maximum likelihood method to the reduced random variables.

First, we set the problem for both the steps described in Sec. VI B. Second, we present the results for step 1 above and third, we simply give the results for step 2 above.

In this section, the optimization parameter is denoted by δ and represents the dispersion parameter $\delta_{M_s} = \delta_{K_s}$ or δ_{B_s}. For ω in band $\mathbb{B}$, let $\omega \mapsto W(\delta,\omega)$ be the second-order stochastic process defined on a probabilistic space $(\Theta,\mathcal{T},\mathcal{P})$, which depends on the optimization parameter $\delta \in \Delta \subset \mathbb{R}$ and which represents the FRF $R^s(\delta,\omega)$ either for step 1 or step 2 defined in Sec. VI B. Let $\omega \mapsto W^{\text{exp}}(\omega)$ be the experimental observation corresponding to the random observation $W(\delta,\omega)$ of the stochastic system. It should be noted that $W^{\text{exp}}(\omega)$ represents either the experimental observation $W^{\text{exp}}_{\text{step 1}}(\omega)$ of step 1 or the reference calculation $W^{\text{ref}}_{\text{step 2}}(\omega)$ of step 2 above. Let $\{\omega_1,\ldots,\omega_{n_{\text{freq}}}\}$ be the frequency sampling of band $\mathbb{B}$.

Let $P_{W(\delta,\omega_1),\ldots,W(\delta,\omega_{n_{\text{freq}}})}(dw_1,\ldots,dw_{n_{\text{freq}}},\delta)$ be the joint probability distribution on $\mathbb{R}^{n_{\text{freq}}}$ of random variables $W(\delta,\omega_1),\ldots,W(\delta,\omega_{n_{\text{freq}}})$ depending on the dispersion parameter δ. For δ in Δ, this joint probability is assumed to be written as

$$P_{W(\delta,\omega_1),\ldots,W(\delta,\omega_{n_{\text{freq}}})}(dw_1,\ldots,dw_{n_{\text{freq}}},\delta) = p(w_1,\ldots,w_{n_{\text{freq}}},\delta)dw_1,\ldots,dw_{n_{\text{freq}}}, \tag{52}$$

in which $p(w_1,\ldots,w_{n_{\text{freq}}},\delta)$ is the probability density function on $\mathbb{R}^{n_{\text{freq}}}$ with respect to the volume element $dw_1,\ldots,dw_{n_{\text{freq}}}$. For all δ fixed in Δ and for all $w_1,\ldots,w_{n_{\text{freq}}}$ given in $\mathbb{R}$, the estimation of $p(w_1,\ldots,w_{n_{\text{freq}}},\delta)$ is performed by using the stochastic computational model and the Monte Carlo method (described in Sec. V B) with n_r independent realizations $\{W(\delta,\omega_1,\theta_1),\ldots,W(\delta,\omega_{n_{\text{freq}}},\theta_1)\},\ldots,\{W(\delta,\omega_1,\theta_{n_r}),\ldots,W(\delta,\omega_{n_{\text{freq}}},0a_{n_r})\}$ of the $\mathbb{R}^{n_{\text{freq}}}$-valued random observation $\{W(\delta,\omega_1),\ldots,W(\delta,\omega_{n_{\text{freq}}})\}$ with $\theta_1,\ldots,\theta_{n_r}$ in Θ. In practice, the sampling $\{\omega_1,\ldots,\omega_{n_{\text{freq}}}\}$ of $\mathbb{B}$ is used and the experimental observation is

$$\{W^{\text{exp}}(\omega_1),\ldots,W^{\text{exp}}(\omega_{n_{\text{freq}}})\}. \tag{53}$$

The problem to be solved then is to find the optimal value δ^{opt} of the dispersion parameter δ of the stochastic computational model using the experimental values defined by Eq. (53), such that

$$\delta^{\text{opt}} = \arg\max_{\delta\in\Delta} \mathcal{L}(\delta), \tag{54}$$

with

$$\mathcal{L}(\delta) = \log_{10} p(W^{\exp}(\omega_1), \ldots, W^{\exp}(\omega_{n_{\text{freq}}}), \delta), \tag{55}$$

because there is only one experiment. Since n_{freq} is large (for instance, $n_{\text{freq}}=300$), the numerical cost to solve Eq. (54) is very high. A usual possible approximation consists in replacing the likelihood function $\mathcal{L}(\delta)$ with the following $\tilde{\mathcal{L}}(\delta) = \sum_{k=1}^{n_{\text{freq}}} \log_{10} p_{W(\delta,\omega_k)}(W^{\exp}(\omega_k), \delta)$ in which $p_{W(\delta,\omega_k)}$ is the probability density function of the random variable $W(\delta,\omega_k)$. This approximation is not efficient and gives an overestimation of δ^{opt} due to the statistical dependence of $W(\delta,\omega_1), \ldots, W(\delta,\omega_{n_{\text{freq}}})$. The method proposed[45] consists in the following methodology. First, a statistical reduction of information using a principal component analysis is performed and second, the maximum likelihood method is applied in the space of the uncorrelated random variables related to the reduced statistical information. For all δ in Δ, we define the random vectors $\mathbb{W}(\delta)=(W(\delta,\omega_1), \ldots, W(\delta,\omega_{n_{\text{freq}}}))$, $\mathbb{m}(\delta)=\mathcal{E}\{\mathbb{W}(\delta)\}$, and $\mathbb{W}^{\exp}=(W^{\exp}(\omega_1), \ldots, W^{\exp}(\omega_{n_{\text{freq}}}))$. Let $[C_{\mathbb{W}}(\delta)]$ be the $(n_{\text{freq}} \times n_{\text{freq}})$ covariance matrix defined by

$$[C_{\mathbb{W}}(\delta)] = \mathcal{E}\{(\mathbb{W}(\delta) - \mathbb{m}(\delta))(\mathbb{W}(\delta) - \mathbb{m}(\delta))^T\}, \tag{56}$$

where $\mathbb{m}(\delta)$ and $[C_{\mathbb{W}}(\delta)]$ are estimated with the Monte Carlo method. We introduce the eigenvalue problem

$$[C_{\mathbb{W}}(\delta)]\mathbb{x}(\delta) = \lambda(\delta)\mathbb{x}(\delta), \tag{57}$$

for which the positive eigenvalues are such that $\lambda_1(\delta) \geq \lambda_2(\delta) \geq \cdots \geq \lambda_{n_{\text{freq}}}(\delta)$. The corresponding vectors $\mathbb{x}^1(\delta), \mathbb{x}^2(\delta), \ldots, \mathbb{x}^{n_{\text{freq}}}(\delta)$ are orthonormal in $\mathbb{R}^{n_{\text{freq}}}$ and are written as $\mathbb{x}^\alpha(\delta)=(x^{\alpha 1}(\delta), \ldots, x^{\alpha n_{\text{freq}}}(\delta))$. Let N_{red} be an integer lesser than n_{freq}. We can then introduce the approximation $\mathbb{W}^{N_{\text{red}}}(\delta)$ of $\mathbb{W}(\delta)$ defined by

$$\mathbb{W}^{N_{\text{red}}}(\delta) = \mathbb{m}(\delta) + \sum_{\alpha=1}^{N_{\text{red}}} \sqrt{\lambda_\alpha(\delta)}\, Y_\alpha(\delta)\mathbb{x}^\alpha(\delta), \tag{58}$$

in which $Y_1(\delta), \ldots, Y_{N_{\text{red}}}(\delta)$ are N_{red} real-valued uncorrelated random variables such that, for all $\alpha=1, \ldots, N_{\text{red}}$,

$$Y_\alpha(\delta) = \frac{1}{\sqrt{\lambda_\alpha(\delta)}} \langle \mathbb{W}(\delta) - \mathbb{m}(\delta), \mathbb{x}^\alpha(\delta) \rangle_{\mathbb{B}}, \tag{59}$$

where $\langle .,. \rangle_{\mathbb{B}}$ is the Euclidian inner product on $\mathbb{R}^{n_{\text{freq}}}$ and $\|\cdot\|_{\mathbb{B}}$ is its associated norm. Let $|||\cdot|||$ be the norm defined by $|||W|||^2 = \mathcal{E}\{\|W\|_{\mathbb{B}}^2\}$. The order N_{red} of the statistical reduction is calculated to get an approximation with a given accuracy ε independent of N_{red} and δ, such that $\max_\delta\{|||\mathbb{D} - \mathbb{D}^{\text{red}}|||^2 / |||\mathbb{D}|||^2\} \leq \varepsilon$ with $\mathbb{D}=\mathbb{W}(\delta)-\mathbb{m}(\delta)$ and $\mathbb{D}^{\text{red}}=\mathbb{W}^{N_{\text{red}}}(\delta) - \mathbb{m}(\delta)$. The value of N_{red} has then to be such that

$$\max_{\delta \in \Delta} \left(1 - \frac{\sum_{\alpha=1}^{N_{\text{red}}} \lambda_\alpha(\delta)}{\text{tr}[C_{\mathbb{W}}(\delta)]} \right) \leq \varepsilon. \tag{60}$$

The statistical reduction is efficient when $N_{\text{red}} \ll n_{\text{freq}}$.

Let $(y_1, \ldots, y_{N_{\text{red}}}) \mapsto p_{Y_1(\delta),\ldots,Y_{N_{\text{red}}}(\delta)}(y_1, \ldots, y_{N_{\text{red}}}, \delta)$ be the probability density function on $\mathbb{R}^{N_{\text{red}}}$ with respect to $dy_1, \ldots, dy_{N_{\text{red}}}$ of the uncorrelated (but not independent) random variables $Y_1(\delta), \ldots, Y_{N_{\text{red}}}(\delta)$. From Eq. (59), it can be deduced that, for all α in $\{1, \ldots, N_{\text{red}}\}$, the experimental realization $Y_\alpha^{\exp}(\delta)$, which now depends on δ, is given, for all $\alpha=1, \ldots, N_{\text{red}}$, by

$$Y_\alpha^{\exp}(\delta) = \frac{1}{\sqrt{\lambda_\alpha(\delta)}} \langle \mathbb{W}^{\exp} - \mathbb{m}(\delta), \mathbb{x}^\alpha(\delta) \rangle_{\mathbb{B}}. \tag{61}$$

Let $\tilde{\mathcal{L}}^{\text{red}}$ be the following approximation of the reduced log-likelihood function, which is defined, for all δ fixed in Δ, by

$$\tilde{\mathcal{L}}^{\text{red}}(\delta) = \sum_{\alpha=1}^{N_{\text{red}}} \log_{10} p_{Y_\alpha(\delta)}(Y_\alpha^{\exp}(\delta), \delta), \tag{62}$$

in which $y \mapsto p_{Y_\alpha(\delta)}(y, \delta)$ is the probability density function on $\mathbb{R}$ of the real-valued random variable $Y_\alpha(\delta)$. This approximation would be exact if the random variables $Y_1(\delta), \ldots, Y_{N_{\text{red}}}(\delta)$ were mutually independent and so the joint probability density function could be written as the product of the marginal probability density functions, which is not true in the present case. Nevertheless, this approximation is reasonably good because the centered random variables $Y_1(\delta), \ldots, Y_{N_{\text{red}}}(\delta)$, although they are mutually dependent, are uncorrelated. The optimization problem to be solved is then given by the following:

$$\delta^{\text{opt}} = \arg \max_{\delta \in \Delta} \tilde{\mathcal{L}}^{\text{red}}(\delta). \tag{63}$$

D. Summary of the identification method

From the experimental measurements $W^{\exp}(\omega_1), \ldots, W^{\exp}(\omega_{n_{\text{freq}}})$ and using Eq. (61) yield the values $Y_1^{\exp}(\delta), \ldots, Y_{N_{\text{red}}}^{\exp}(\delta)$ for all δ in Δ. The use of the stochastic computational model and the Monte Carlo method allows to compute independent realizations of the dependent random variables $W(\delta,\omega_1), \ldots, W(\delta,\omega_{n_{\text{freq}}})$ for all δ in Δ. For a fixed accuracy parameter ε, the smallest value of $N_{\text{red}} \leq n_{\text{freq}}$ is calculated with the use of Eq. (60). The use of Eq. (59) then allows the independent realizations of the dependent but uncorrelated random variables $Y_1(\delta), \ldots, Y_{N_{\text{red}}}(\delta)$ to be calculated and to deduce estimations of the marginal probability density function $p_{Y_\alpha(\delta)}(Y_\alpha^{\exp}(\delta), \delta)$. Using Eqs. (62) and (63) yields the optimized dispersion parameter δ^{opt}.

E. Results

This section is devoted to the identification of the dispersion parameters δ_{M_s} and δ_{K_s} of the structure according to step 1 of Sec. VI B. We then perform the identification of the dispersion parameter δ_{B_s} for the sound-insulation layer according to step 2 of Sec. VI B. First, Eq. (60) is used with $\Delta=[0.1, 0.95]$ to compute the value of N_{red} and yields $N_{\text{red}}=100$ for $\varepsilon=0.02$. Figure 8 displays the graph of the function $\delta \mapsto \tilde{\mathcal{L}}^{\text{red}}(\delta)$ for $N_{\text{red}}=100$.

The maximum of this function is reached for $\delta_{M_s}^{\text{opt}} = \delta_{K_s}^{\text{opt}} = \delta^{\text{opt}} = 0.3$. Figure 9 displays the graph of the FRF $\omega \mapsto R^s(\omega)$ defined in Sec. V B for the uncertain structure without the sound-insulation layer and with $\delta_{D_s}=0.3$.

The identification of the dispersion parameter δ_{B_s} is performed as explained in Sec. VI B. For the sake of brevity, we

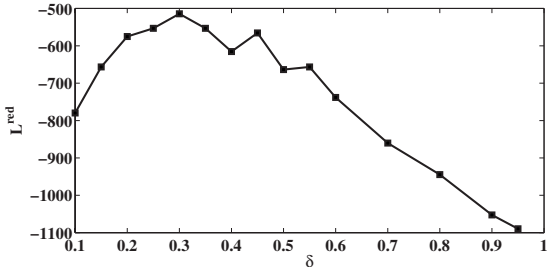


FIG. 8. Graph of the function $\delta \mapsto \tilde{\mathcal{L}}^{\mathrm{red}}(\delta)$ for $N_{\mathrm{red}}=100$.

only give the result obtained from the optimization problem, which is $\delta_{B_s}^{\mathrm{opt}}=0.6$. Figure 10 displays the graph of the FRF $\omega \mapsto R^s(\omega)$ defined in Sec. V B for the uncertain structure coupled with the uncertain sound-insulation layer and with $\delta_{D_s}=0.3$.

The confidence region is computed with a probability level of 95% and then the probability that the experiments be outside the confidence region is nonzero. Certainly, this prediction could be improved both in increasing the number of realizations and in improving the underlying deterministic model of the structure without the sound-insulation layer.

VII. PREDICTION OF THE VIBROACOUSTIC RESPONSES WITH THE IDENTIFIED COMPUTATIONAL STOCHASTIC MODEL

In this section, we use the identified computational stochastic model to predict a vibroacoustic response. The response of the identified model is compared to a reference solution calculated with a commercial software. The structure and the sound-insulation layer models are defined in Sec. IV. The sound-insulation layer is coupled with a parallelepiped acoustic cavity ($m_a=23\,354$ DOF and $n_a=67$ modes), which is assumed to be without uncertainties and which is filled with air. First, we observe the pressure at $m_{\mathrm{obs}}=120$ points in the acoustic cavity while the excitation force is applied to the elastic framework of the structure as in Sec. IV. We then compute the FRF $\omega \mapsto r^a(\omega)$ relative to the acoustic cavity and Fig. 11 displays its graph for the mean acoustic system with and without the sound-insulation layer.

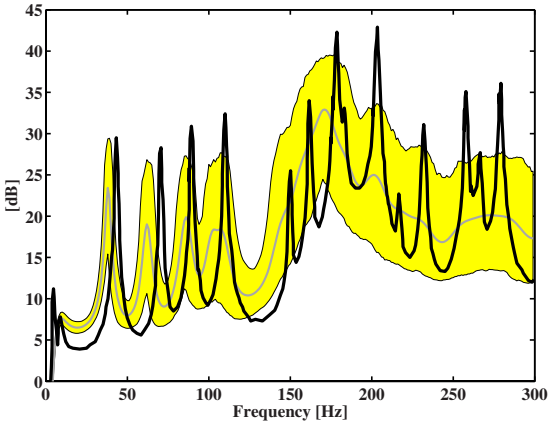


FIG. 9. (Color online) Graph of $\omega \mapsto R^s(\omega)$ for the structure without the sound-insulation layer: measurements (thick black line), stochastic confidence zone (gray envelop), and mean stochastic response (thick dark gray line).

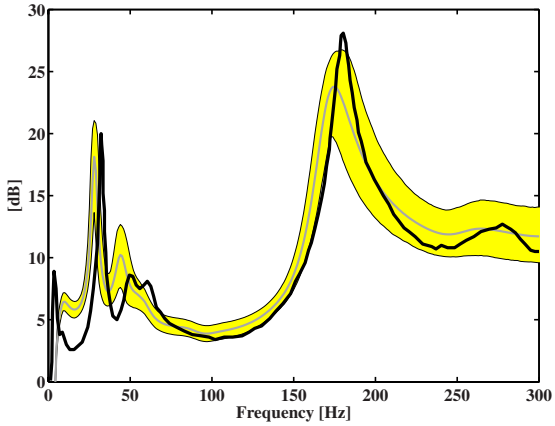


FIG. 10. (Color online) Graph of $\omega \mapsto R^s(\omega)$ for the uncertain structure with the uncertain sound-insulation layer: measurements (thick black line), stochastic confidence zone (gray region), and mean stochastic response (thick dark gray line).

This figure shows the effects of the sound-insulation layers of the vibroacoustic responses of the reference configuration. It can be seen that there is a significant effect on frequency band [200,450] Hz. Second, we use the same observation in the acoustic cavity and the same structural excitation as above. Uncertainties are now taken into account in the structure and in the sound-insulation layer. The random equation, which has to be solved, is then the following:

$$\begin{bmatrix} [\mathbf{A}^s(\omega)] + \underline{a}^s(\omega)[\mathbf{B}^s] & \underline{a}^c(\omega)[\mathbf{C}^s] \\ \omega^2 \underline{a}^c(\omega)[\mathbf{C}^s]^T & [\underline{A}^a(\omega)] + \underline{a}^a(\omega)[\mathbf{B}^a] \end{bmatrix} \times \begin{bmatrix} \mathbf{Q}^s(\omega) \\ \mathbf{Q}^a(\omega) \end{bmatrix} = \begin{bmatrix} \underline{f}^s(\omega) \\ \underline{f}^a(\omega) \end{bmatrix}. \tag{64}$$

The values of the dispersion parameters of the probabilistic model are the values identified in Sec. VI E, i.e., $\delta_{M_s}^{\mathrm{opt}}=\delta_{K_s}^{\mathrm{opt}}=0.3$, $\delta_{D_s}=0.3$, and $\delta_{B_a}^{\mathrm{opt}}=0.6$. It is assumed that the dispersion parameters for the sound-insulation layer are equal, that is to say $\delta_{C_s}^{\mathrm{opt}}=\delta_{B_a}^{\mathrm{opt}}=\delta_{B_s}^{\mathrm{opt}}=0.6$. Figure 12 defined in Sec. V B for the uncertain structure coupled with the uncertain sound-insulation layer and the acoustic cavity. The analysis of this figure shows that there are differences between the reference response (thick solid line) and the response of the mean simplified computational model (these differences can be estimated looking at the statistical mean response given by the mid gray line). Nevertheless, the stochastic simplified computational model allows the prediction to be improved in the

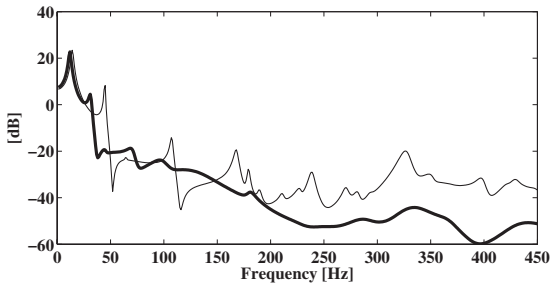


FIG. 11. Graphs of $\omega \mapsto r^a(\omega)$ for the reference solution without a sound-insulation layer (thin solid line) and with a sound-insulation layer (thick solid line).

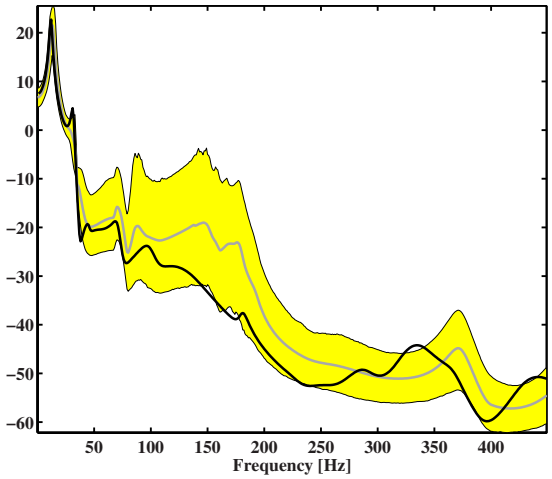


FIG. 12. (Color online) Graph of $\omega \mapsto R^a(\omega)$ for the system with a sound-insulation layer: reference computation (thick black line), stochastic confidence zone (gray region), and mean stochastic response (thick dark gray line).

probability sense. It can be seen that the reference response belongs to the confidence region. First, the two responses coincide in the frequency band [0,40] Hz. Second, it can be seen that the reference solution is close to the lower bound of the confidence region in the frequency band [40,250] Hz. Finally, the reference solution reaches the lower and the upper envelopes of the confidence in the frequency band [250,450] Hz.

VIII. CONCLUSION

In this paper, a new extension of the fuzzy structure theory to elastoacoustic element is presented in order to construct a simplified model of sound-insulation layers. Such a simplified model, based on an extension of the fuzzy structure theory, (1) allows the dynamics of the sound-insulation layer to be taken into account without increasing the number of DOF in the computational vibroacoustic model and (2) allows a representation of the sound-insulation in terms of physical parameters such as its participating mass, its modal density, and its internal damping rate. This approach allows several kinds of sound-insulation layers to be simultaneously taken into account in the computational vibroacoustic model of a complex system such as a car with a very small increase in the computational cost. In addition, taking into account the complexity of the actual sound-insulation layers' design, it is necessary to implement a model of uncertainties. This is a reason why a probabilistic approach of both model and system parameters' uncertainties is introduced in the simplified model of the sound-insulation layer based on the fuzzy structure theory. The complete related developments are given and an experimental validation is presented. Figure 10 (relative to the response of the structure coupled with the sound-insulation layer) and Fig. 12 (relative to the response of the acoustic cavity for the vibroacoustic system constituted of the structure coupled with the sound-insulation layer and with the acoustic cavity) show that the predictions are good from a stochastic point of view. Finally, an efficient design methodology is proposed to identify the parameters of the simplified model of the sound-insulation layer. The mean parameters are identified by solving an inverse problem formulated as an optimization problem using an experimental database. The maximum likelihood method coupled to a statistical reduction of information is performed to obtain the dispersion parameters.

The method proposed in this paper is not intended to give a tool for designing sound-insulation layers (which are designed using high-frequency response considerations). This approach has been developed in the low- and medium-frequency bands in complex structural-acoustics systems such as cars for which, due to industrial processes and vehicle diversity, there are very large variabilities induced by the kinematic conditions at the attachment interface between the master structure and the sound-insulation layers.

ACKNOWLEDGMENT

This research has been supported by the French National Agency for Research and Technology (ANRT).

APPENDIX A: FORMULATION OF THE VIBROACOUSTIC PROBLEM WITH A SOUND-INSULATION LAYER

In this section, we give additional explanations relative to Sec. II useful for the construction of the simplified model of the sound-insulation layer. Let $\mathcal{C}_0^s$ be the space of the admissible displacement fields of the structure, $\mathcal{C}^a$ be the space of the admissible pressure fields in the acoustic cavity, and $\mathcal{C}^h$ be the space of the admissible displacement fields of the sound-insulation layer. For all ω in $\mathbb{B}$, we introduce the following bilinear form defined on $\mathcal{C}_0^s \times \mathcal{C}^a$:

$$c_{\Gamma_2}(\boldsymbol{u}^s, p) = \int_{\Gamma_2} \boldsymbol{u}^s(\boldsymbol{x}) \cdot \boldsymbol{n}^s(\boldsymbol{x}) p(\boldsymbol{x}) ds(\boldsymbol{x}), \tag{A1}$$

the bilinear form defined on $\mathcal{C}^h \times \mathcal{C}^a$:

$$c_{\Gamma}(\boldsymbol{u}^h, p) = \int_{\Gamma} \boldsymbol{u}^h(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}) p(\boldsymbol{x}) ds(\boldsymbol{x}), \tag{A2}$$

and the linear form defined on $\mathcal{C}_0^s$ or $\mathcal{C}^h$:

$$c_{\Gamma_s}(\boldsymbol{u}; \omega) = \int_{\Gamma_s} \boldsymbol{f}^s(\boldsymbol{x}, \omega) \boldsymbol{u}(\boldsymbol{x}) ds(\boldsymbol{x}). \tag{A3}$$

The weak formulation of the vibroacoustic boundary value problem is formulated as follows (see Ref. 12). For all ω in $\mathbb{B}$, find $(\boldsymbol{u}^s, p, \boldsymbol{u}^h)$ in $\mathcal{C}_0^s \times \mathcal{C}^a \times \mathcal{C}^h$ such that for all $(\boldsymbol{\delta u}^s, p, \boldsymbol{\delta u}^h)$ in $\mathcal{C}_0^s \times \mathcal{C}^a \times \mathcal{C}^h$, we have, for the structure,

$$-\omega^2 m^s(\boldsymbol{u}^s, \boldsymbol{\delta u}^s) + i\omega d^s(\boldsymbol{u}^s, \boldsymbol{\delta u}^s; \omega) + k^s(\boldsymbol{u}^s, \boldsymbol{\delta u}^s; \omega) + c_{\Gamma_2}(\boldsymbol{\delta u}^s, p) = -c_{\Gamma_s}(\boldsymbol{\delta u}^s; \omega) + l^s(\boldsymbol{\delta u}^s; \omega), \tag{A4}$$

for the acoustic cavity,

$$-\omega^2 m^a(p, \delta p) + i\omega d^a(p, \delta p; \omega) + k^a(p, \delta p) + \omega^2 \{c_{\Gamma_2}(\boldsymbol{u}^s, \delta p) + c_{\Gamma}(\boldsymbol{u}^h, \delta p)\} = l^a(\delta p; \omega), \tag{A5}$$

and for the sound-insulation layer,

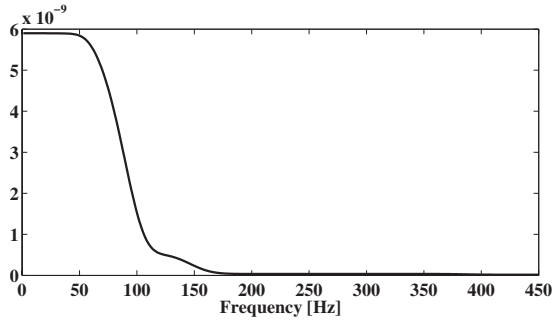


FIG. 13. Graph of $\omega \mapsto \underline{a}_r^s(\omega)$.

$$- \omega^2 m^h(\boldsymbol{u}^h, \boldsymbol{\delta u}^h) + i\omega d^h(\boldsymbol{u}^h, \boldsymbol{\delta u}^h; \omega) + k^h(\boldsymbol{u}^h, \boldsymbol{\delta u}^h; \omega) c_\Gamma(\boldsymbol{\delta u}^h, p) = c_{\Gamma_s}(\boldsymbol{\delta u}^h; \omega). \tag{A6}$$

The bilinear forms m^s, d^s, and k^s [of the acoustic cavity (m^a, d^a, and k^a, respectively) and of the sound-insulation layer (m^h, d^h, and k^h, respectively)] relative to the mass, damping, and stiffness of the structure and the linear forms l^s and l^a related to the structural and acoustical excitations are defined in Ref. 12. For instance, we have

$$m^s(\boldsymbol{u}^s, \boldsymbol{\delta u}^s) = \int_{\Omega_s} \rho^s(\boldsymbol{x}) \boldsymbol{u}^s(\boldsymbol{x}) \cdot \boldsymbol{\delta u}^s(\boldsymbol{x}) d\boldsymbol{x}, \tag{A7}$$

$$d^s(\boldsymbol{u}^s, \boldsymbol{\delta u}^s; \omega) = \int_{\Omega_s} b_{ijkh}(\boldsymbol{x}, \omega) \varepsilon_{kh}(\boldsymbol{u}^s) \varepsilon_{ij}(\boldsymbol{\delta u}^s) d\boldsymbol{x}, \tag{A8}$$

$$k^s(\boldsymbol{u}^s, \boldsymbol{\delta u}^s; \omega) = \int_{\Omega_s} a_{ijkh}(\boldsymbol{x}, \omega) \varepsilon_{kh}(\boldsymbol{u}^s) \varepsilon_{ij}(\boldsymbol{\delta u}^s) d\boldsymbol{x}, \tag{A9}$$

$$m^a(p, \delta p) = \frac{1}{\rho_0 c_0^2} \int_{\Omega_a} p(\boldsymbol{x}) \delta p(\boldsymbol{x}) d\boldsymbol{x}, \tag{A10}$$

$$d^a(p, \delta p; \omega) = \tau(\omega) k^a(p, \delta p), \tag{A11}$$

$$k^a(p, \delta p) = \frac{1}{\rho_0} \int_{\Omega_a} \boldsymbol{\nabla} p \cdot \boldsymbol{\nabla} \delta p d\boldsymbol{x}, \tag{A12}$$

$$c_{\Gamma_2}(\boldsymbol{u}^s, \delta p) = \int_{\Gamma_2} \boldsymbol{u}^s(\boldsymbol{x}) \,.\, \boldsymbol{n}^s(\boldsymbol{x}) \delta p(\boldsymbol{x}) ds(\boldsymbol{x}). \tag{A13}$$

APPENDIX B: FUNCTIONS OF THE FUZZY COEFFICIENTS INTRODUCED IN SECTION IIE

For all $\omega \in [\Omega_C, \omega_{\max}]$,

$$\Theta_R(\omega) = \frac{1}{4\sqrt{1 - \underline{\xi}(\omega)^2}} \ln\left\{ \frac{N^+(\tilde{b}(\omega), \underline{\xi}(\omega)) N^-(\tilde{a}(\omega), \underline{\xi}(\omega))}{N^-(\tilde{b}(\omega), \underline{\xi}(\omega)) N^+(\tilde{a}(\omega), \underline{\xi}(\omega))} \right\}, \tag{B1}$$

$$\Theta_I(\omega) = \frac{1}{2\sqrt{1 - \underline{\xi}(\omega)^2}} [\Lambda(\tilde{b}(\omega), \underline{\xi}(\omega)) - \Lambda(\tilde{a}(\omega), \underline{\xi}(\omega))], \tag{B2}$$

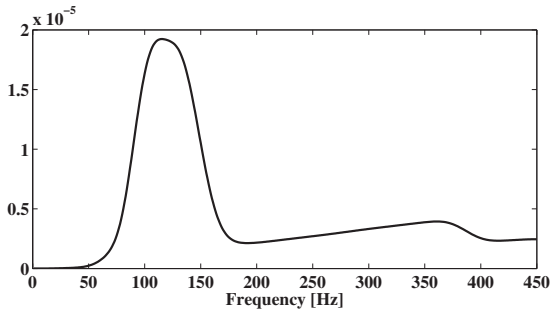


FIG. 14. Graph of $\omega \mapsto \underline{a}_i^s(\omega)$.

$$N^\pm(u, \xi) = u^2 \pm 2u\sqrt{1 - \xi^2} + 1, \tag{B3}$$

$$\Lambda(u, \xi) = \arctan\left\{ \frac{u^2 + 2\xi^2 - 1}{2\xi\sqrt{1 - \xi^2}} \right\}, \tag{B4}$$

$$\tilde{a}(\omega) = \frac{1}{\omega} a(\omega), \tag{B5}$$

where a is defined in Eq. (14),

$$\tilde{b}(\omega) = \frac{1}{\omega} b(\omega), \tag{B6}$$

where b is defined in Eq. (15),

$$\tilde{\ell}(\omega) = \frac{1}{\tilde{b}(\omega) - \tilde{a}(\omega)}, \quad \underline{\lambda}(\omega) = \frac{\tilde{\ell}(\omega)}{\omega \underline{n}(\omega)}. \tag{B7}$$

APPENDIX C: GRAPHS OF THE FUZZY COEFFICIENTS

Figures 13–18 display the graphs of the fuzzy coefficients defined by Eqs. (29)–(34) with Eqs. (B1)–(B7) for $\underline{\xi} = 0.01$ and $\underline{n}(\omega)$ given by Fig. 4 and $\underline{\nu}(\omega)$ given by Fig. 5.

APPENDIX D: COMPLEMENTS RELATIVE TO THE NONPARAMETRIC PROBABILISTIC APPROACH

This section deals with complements relative to the probabilistic nonparametric approach, which allows both model and system parameters' uncertainties to be taken into account in the computational model of the structure, in the simplified model of the sound-insulation layer, and in the computational model of the acoustic cavity. The random matrices associated with $[\underline{M}^s]$, $[\underline{D}^s(\omega)]$, $[\underline{K}^s(\omega)]$, $[\underline{C}]$, $[\underline{M}^a]$,

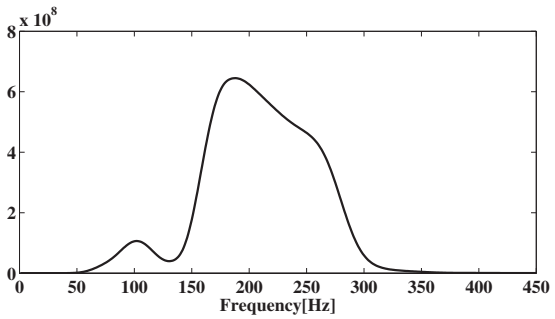


FIG. 15. Graph of $\omega \mapsto \underline{a}_r^a(\omega)$.

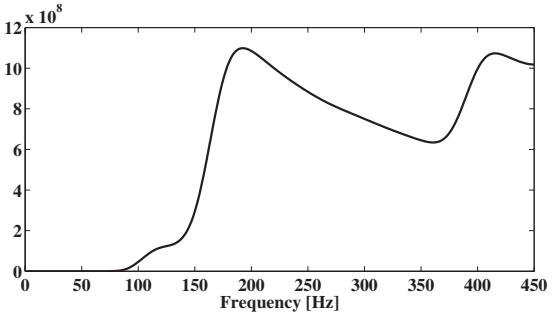


FIG. 16. Graph of $\omega \mapsto \underline{a}_i^a(\omega)$.

$[\underline{D}^a(\omega)]$, $[\underline{K}^a]$, $[\underline{B}^s]$, $[\underline{B}^a]$, and $[\underline{C}^s]$ are constructed as explained in Ref. 40. In this construction, the complex coefficients $\underline{a}^s(\omega)$, $\underline{a}^a(\omega)$, and $\underline{a}^c(\omega)$ are deterministic. For example, the random matrix associated with $\underline{a}^s(\omega)[\underline{B}^s]$ is $\underline{a}^s(\omega)[\mathbf{B}^s]$ in which $[\mathbf{B}^s]$ is the random matrix associated with $[\underline{B}^s]$. For ω fixed in $\mathbb{B}$, we then introduce the matrices $[\mathbf{M}^s]$, $[\mathbf{D}^s(\omega)]$, $[\mathbf{K}^s(\omega)]$, $[\mathbf{C}]$, $[\mathbf{M}^a]$, $[\mathbf{D}^a(\omega)]$, $[\mathbf{K}^a]$, $[\mathbf{B}^s]$, $[\mathbf{C}^s]$, and $[\mathbf{B}^a]$. These matrices are independent second-order random variables. The random matrices $[\mathbf{M}^s]$, $[\mathbf{D}^s(\omega)]$, and $[\mathbf{K}^s(\omega)]$ are with values in $\mathbb{M}_{n_s}^+(\mathbb{R})$; the random matrix $[\mathbf{B}^s]$ is with values in $\mathbb{M}_{n_s}^{+0}(\mathbb{R})$; the random matrix $[\mathbf{M}^a]$ is with values in $\mathbb{M}_{n_a}^+(\mathbb{R})$; the random matrices $[\mathbf{K}^a]$, $[\mathbf{D}^a(\omega)]$, and $[\mathbf{B}^a]$ are with values in $\mathbb{M}_{n_a}^{+0}(\mathbb{R})$; and the random matrices $[\mathbf{C}]$ and $[\mathbf{C}^s]$ are with values in $\mathbb{M}_{n_s,n_a}(\mathbb{R})$. The mean values of these random matrices are such that

$$\mathcal{E}\{[\mathbf{M}^s]\} = [\underline{M}^s], \quad \mathcal{E}\{[\mathbf{D}^s(\omega)]\} = [\underline{D}^s(\omega)],$$

$$\mathcal{E}\{[\mathbf{K}^s(\omega)]\} = [\underline{K}^s(\omega)], \quad \mathcal{E}\{[\mathbf{M}^a]\} = [\underline{M}^a],$$

$$\mathcal{E}\{[\mathbf{D}^a(\omega)]\} = [\underline{D}^a(\omega)], \quad \mathcal{E}\{[\mathbf{K}^a]\} = [\underline{K}^a],$$

$$\mathcal{E}\{\underline{a}^s(\omega)[\mathbf{B}^s]\} = \underline{a}^s(\omega)[\underline{B}^s], \quad \mathcal{E}\{[\mathbf{C}]\} = [\underline{C}],$$

$$\mathcal{E}\{\underline{a}^a(\omega)[\mathbf{B}^a]\} = \underline{a}^a(\omega)[\underline{B}^a], \quad \mathcal{E}\{\underline{a}^c(\omega)[\mathbf{C}^s]\} = \underline{a}^c(\omega)[\underline{C}^s], \tag{D1}$$

where $\mathcal{E}$ is the mathematical expectation. Moreover, those matrices have the required mathematical properties (see Refs. 39 and 40 in which the reader will be able to find all the details). The generalized stiffness random matrices are then written as

$$[\mathbf{A}^s(\omega)] = -\omega^2[\mathbf{M}^s] + i\omega[\mathbf{D}^s(\omega)] + [\mathbf{K}^s(\omega)], \tag{D2}$$

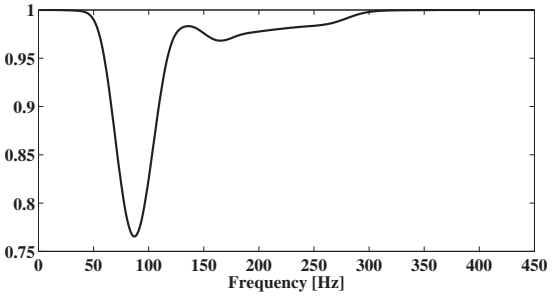


FIG. 17. Graph of $\omega \mapsto \underline{a}_r^c(\omega)$.

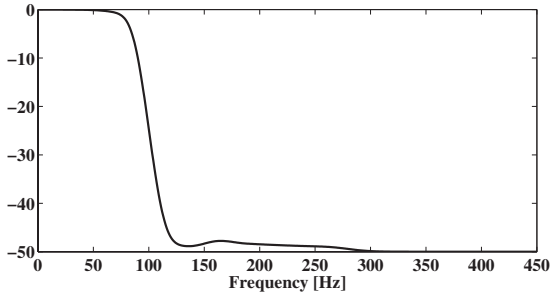


FIG. 18. Graph of $\omega \mapsto \underline{a}_i^c(\omega)$.

$$[\mathbf{A}^a(\omega)] = -\omega^2[\mathbf{M}^a] + i\omega[\mathbf{D}^a(\omega)] + [\mathbf{K}^a], \tag{D3}$$

where $[\mathbf{A}^s(\omega)]$ is a random matrix with values in $M_{n_s}^S(\mathbb{C})$ and where $[\mathbf{A}^a(\omega)]$ is a random matrix with values in $M_{n_a}^S(\mathbb{C})$. For example, we give below the detail of the construction for the random matrix $[\mathbf{K}^s(\omega)]$. The matrix $[\underline{K}^s(\omega)]$ can be written as $[\underline{K}^s(\omega)] = [L_{K^s}(\omega)]^T[L_{K^s}(\omega)]$ corresponding to the Choleski decomposition of the positive-definite matrix $[\underline{K}^s(\omega)]$. We then introduce the random matrix $[\mathbf{K}^s(\omega)] = [L_{K^s}(\omega)]^T[\mathbf{G}_{K^s}] \times [L_{K^s}(\omega)]$ where the random matrix $[\mathbf{G}_{K^s}]$ belongs to the $\mathbf{SG}^+$ ensemble defined in Ref. 40 and is independent of the frequency. The dispersion parameter δ_{K^s} of this random matrix $[\mathbf{K}^s(\omega)]$ is independent of the dimension and of the frequency and is defined by $\delta_{K^s} = (\mathcal{E}\{\|[\mathbf{G}_{K^s}] - [\underline{G}_{K^s}]\|_F^2\} / \|[\underline{G}_{K^s}]\|_F^2)^{1/2}$ in which $\|K\|_F$ is the Frobenius norm defined by $\|K\|_F^2 = \mathrm{tr}(K^T K)$.

[1] Y.-J. Kang and J. S. Bolton, "Finite element modeling of isotropic elastic porous materials coupled with acoustical finite elements," J. Acoust. Soc. Am. **98**, 635–643 (1995).

[2] Y.-J. Kang and J. S. Bolton, "A finite element model for sound transmission through foam-lined double-panel structures," J. Acoust. Soc. Am. **99**, 2755–2765 (1996).

[3] R. Panneton and N. Atalla, "Numerical prediction of sound transmission through finite multilayer systems with poroelastic materials," J. Acoust. Soc. Am. **100**, 346–354 (1996).

[4] R. Panneton and N. Atalla, "An efficient finite element scheme for solving the three-dimensional poroelasticity problem in acoustics," J. Acoust. Soc. Am. **101**, 3287–3298 (1997).

[5] N. Atalla, R. Panneton, and P. Debergue, "A mixed displacement-pressure formulation for poroelastic materials," J. Acoust. Soc. Am. **104**, 1444–1452 (1998).

[6] N. Atalla, M. A. Hamdi, and R. Panneton, "Enhanced weak integral formulation for the mixed (u,p) poroelastic equations," J. Acoust. Soc. Am. **109**, 3065–3068 (2001).

[7] N. Atalla, F. Sgard, and C. K. Amedin, "On the modeling of sound radiation from poroelastic materials," J. Acoust. Soc. Am. **120**, 1990–1995 (2006).

[8] F. J. Fahy, *Sound and Structural Vibration, Radiation, Transmission and Response* (Academic, New York, 1985).

[9] D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (McGraw Hill, NY, 1981).

[10] M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction* (Acoustical Society of America, New York, 1993).

[11] J. F., Allard, *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials* (Chapman and Hall, London, 1994).

[12] R. Ohayon and C. Soize, *Structural Acoustics and Vibration* (Academic, San Diego, CA, 1998).

[13] J. L. Guyader and C. Lesueur, "Acoustic transmission through orthotropic multilayered plates, Part. I. Plate vibration modes," J. Sound Vib. **58**, 51–86 (1978).

[14] J. A. Moore and R. H. Lyon, "Resonant porous material absorbers," J. Acoust. Soc. Am. **72**, 1989–1999 (1982).

[15] J. F. Allard, C. Champoux, and C. Depollier, "Modelization of layered sound absorbing materials with transfer matrices," J. Acoust. Soc. Am. **82**,

1792–1796 (1987).
[16]A. Blaise, C. Lesueur, M. Gotteland, and M. Barbe, "On sound transmission into an orthotropic infinite shell—Comparison with Koval's results and understanding of phenomena," J. Sound Vib. **150**, 233–243 (1991).
[17]J.-F. Allard and G. Daigle, "Propagation of sound in porous media: Modeling sound absorbing materials," J. Acoust. Soc. Am. **95**, 2785 (1994).
[18]C. Lesueur, G., Poumerol, and A. Blaise, "Vibroacoustic response of composite multilayered plate coupled to a rectangular cavity and excited by white noise and a turbulent boundary layer," Acta Acust. **3**, 153–167 (1995).
[19]N. Atalla and R. Panneton, "The effects of multilayer sound-absorbing treatments on the noise field inside a plate backed-cavity," Noise Control Eng. J. **44**, 235–243 (1996).
[20]R. S. Langley and P. Bremner, "A hybrid method for the vibration analysis of complex structural-acoustic systems," J. Acoust. Soc. Am. **105**, 1657–1671 (1999).
[21]J. M. Cuschieri and D. Feit, "Influence of circumferential partial coating on the acoustic radiation from a fluid-loaded shell," J. Acoust. Soc. Am. **107**, 3196–3207 (2000).
[22]B. Faverjon and C. Soize, "Equivalent acoustic impedance model. Part 2: Analytical approximation," J. Sound Vib. **276**, 593–613 (2004).
[23]L. Jaouen, B. Brouard, N. Atalla, and C. Langlois, "A simplified numerical model for a plate backed by a thin foam layer in the low frequency range," J. Sound Vib. **280**, 681–698 (2005).
[24]C. Soize, "Probabilistic structural modeling in linear dynamical analysis of complex mechanical systems. I. Theoretical elements," Rech. Aerosp. **5**, 23–48 (1986).
[25]F. Chabas, A. Desanti, and C. Soize, "Probabilistic structural modeling in linear dynamical analysis of complex mechanical systems. II. Numerical analysis and applications," Rech. Aerosp. **5**, 49–67 (1986).
[26]C. Soize, "A model and numerical method in the medium frequency range for vibroacoustic predictions using the theory of structural fuzzy," J. Acoust. Soc. Am. **94**, 849–865 (1993).
[27]C. Soize, "Estimation of the fuzzy substructure model parameters using the mean power flow equation of the fuzzy structure," ASME J. Vibr. Acoust. **120**, 279–286 (1998).
[28]C. Soize and K. Bjaoui, "Estimation of fuzzy structure parameters for continuous junctions," J. Acoust. Soc. Am. **107**, 2011–2020 (2000).
[29]A. D. Pierce, V. W. Sparrow, and D. A. Russell, "Fundamental structural-acoustic idealizations for structures with fuzzy internals," ASME J. Vibr. Acoust. **117**, 339–348 (1995).
[30]M. Strasberg and D. Feit, "Vibration damping of large structures induced by attached small resonant structures," J. Acoust. Soc. Am. **99**, 335–344 (1996).
[31]G. Maidanik and J. Dickey, "Design criteria for the damping effectiveness of structural fuzzies," J. Acoust. Soc. Am. **100**, 2029–2033 (1996).
[32]R. L. Weaver, "The effect of an undamped finite degree of freedom "fuzzy" substructure: Numerical solution and theoretical discussion," J. Acoust. Soc. Am. **100**, 3159–3164 (1996).
[33]R. L. Weaver, "Mean and mean-square responses of a prototypical master/fuzzy structure," J. Acoust. Soc. Am. **101**, 1441–1449 (1997).
[34]R. L. Weaver, "Multiple-scattering theory for mean responses in a plate with sprung masses," J. Acoust. Soc. Am. **101**, 3466–3474 (1997).
[35]R. L. Weaver, "Mean-square responses in a plate with sprung masses, energy flow and diffusion," J. Acoust. Soc. Am. **103**, 414–427 (1998).
[36]J.-M. Mencik and A. Berry, "A theoretical formulation of the dynamical response of a master structure coupled with elastic continuous fuzzy subsystems with discrete attachments," J. Sound Vib. **280**, 1031–1050 (2005).
[37]C. Soize, "A nonparametric model of random uncertainties for reduced matrix models in structural dynamics," Probab. Eng. Mech. **15**, 277–294 (2000).
[38]C. Soize, "Maximum entropy approach for modeling random uncertainties in transient elastodynamics," J. Acoust. Soc. Am. **109**, 1979–1996 (2001).
[39]C. Soize, "A comprehensive overview of a non-parametric probabilistic approach of model uncertainties for predictive models in structural dynamics," J. Sound Vib. **288**, 623–652 (2005).
[40]C. Soize, "Random matrix theory for modeling uncertainties in computational mechanics," Comput. Methods Appl. Mech. Eng. **194**, 1333–1366 (2005).
[41]O. C. Zienkewicz and R. L. Taylor, *The Finite Element Method*, 4th ed. (McGraw-Hill, New York, 1989).
[42]J.-F. Durand, "Modélisation de véhicules en vibroacoustique numérique avec incertitudes de modélisation et validation expérimentale (Uncertain computational vibroacoustics modeling of vehicles including model uncertainties and experimental validation," Doctoral Thesis, Université de Marne-la-Vallée, France (2007).
[43]J.-F. Durand, C. Soize, and L. Gagliardini, "Structural-acoustic modeling of automotive vehicles in presence of uncertainties and experimental identification and validation," J. Acoust. Soc. Am. **124**, 1513–1525 (2008).
[44]R. Serfling, *Approximations Theorems of Mathematical Statistics* (Wiley, New York, 1980).
[45]C. Soize, E. Capiez-Lernout, J.-F. Durand, C. Fernandez, and L. Gagliardini, "Probabilistic model identification of uncertainties in computational models for dynamical systems and experimental validation," Comput. Methods Appl. Mech. Eng. **198**, 150–163 (2008).

Wave characterization of cylindrical and curved panels using a finite element method

Elisabetta Manconi[a)]
Dipartimento di Ingegneria Industriale, Università degli Studi di Parma, Parma 43100, Italy

Brian R. Mace
ISVR, University of Southampton, Southampton S017 1BJ, United Kingdom

(Received 26 May 2008; revised 9 October 2008; accepted 16 October 2008)

This paper describes a wave finite element method for the numerical prediction of wave characteristics of cylindrical and curved panels. The method combines conventional finite elements and the theory of wave propagation in periodic structures. The mass and stiffness matrices of a small segment of the structure, which is typically modeled using either a single shell element or, especially for laminated structures, a stack of solid elements meshed through the cross-section, are postprocessed using periodicity conditions. The matrices are typically found using a commercial FE package. The solutions of the resulting eigenproblem provide the frequency evolution of the wavenumber and the wave modes. For cylindrical geometries, the circumferential order of the wave can be specified in order to define the phase change that a wave experiences as it propagates across the element in the circumferential direction. The method is described and illustrated by application to cylinders and curved panels of different constructions. These include isotropic, orthotropic, and laminated sandwich constructions. The application of the method is seen to be straightforward even in the complicated case of laminated sandwich panels. Accurate predictions of the dispersion curves are found at negligible computational cost.
 [DOI: 10.1121/1.3021418]

PACS number(s): 43.40.Ey, 43.20.Mv, 43.20.Bi [SFW] Pages: 154–163

I. INTRODUCTION

Cylindrical and curved panels are used in many applications. Examples include pipes, acoustic ducts, aircraft fuselage, and aerospace structures, to name a few. Free and forced vibrations of cylindrical and curved panels can be described in terms of wave components. In particular, the wave approach is valuable at mid to high frequency when the size of the structure is large compared to the wavelength. This is of importance in studies concerning the vibroacoustic behavior of such structures or in studies of nanoscale devices such as nanotubes, where the evaluation of the dynamic behavior at a very high frequency is fundamental. Theoretical understanding of wave propagation in curved structures also provides the background necessary for the utilization and a better implementation of many techniques. Typical applications include, amongst others, the transmission of structure-borne sound, statistical energy analysis, shock response, and nondestructive testing. In many cases, once the characteristics of wave propagation are known, the analysis becomes straightforward. The primary characteristics of these waves are the dispersion relationship, which is the relationship between frequency and wave heading to the wavenumber, and wave modes, which are related to the cross-sectional displacements. In simple cases, such as isotropic thin cylinders, analytical expressions for the dispersion curves can be found.[1] However, the knowledge of wave characteristics requires a comprehensive development of a mathematical model that is difficult at best, especially for complex constructions such as fiber reinforced laminates. As a consequence, dispersion relations and wave modes are frequently unavailable or treated in a simplified manner.

Studies of the dispersion characteristics of cylindrical shells in the framework of three-dimensional elasticity theory have been made by several authors. To cite a few we can list the classical works of Gazis[2,3] and Kumar and Stephens.[4] However, due to the complex algebra involved in the three-dimensional theory of elasticity, curved structures have generally been studied by making assumptions and approximations concerning the stress-strain distributions in the solid.[5] Wave characteristics and dispersion curves of thin cylindrical shells have been obtained by various approximate theories in a number of studies.[6–10] In any event, the analysis proposed is difficult, and the equations lead to complicated dispersion relations, which are nearly intractable. Hence, most of the studies are solely confined to purely real dispersion curves. This is particularly true for laminated structures due to the great amount of algebra involved in enforcing interlaminar continuity when a large number of layers are considered. There are also numerical problems in obtaining the dispersion curves because of convergence difficulties for short wavelengths and large circumferential wavenumbers. Moreover, the accuracy of the solutions depends on whether the assumptions made are valid. Thus, in all such cases alternative numerical methods are potentially beneficial for determining the dispersion properties, wave modes, and group and phase velocities.

[a)]Electronic mail: elisabetta.manconi@unipr.it

The finite element (FE) approach has been used widely to study wave propagation in structures. For example, Nelson *et al.*[11] analyzed wave propagation in laminated orthotropic cylinders, assuming interpolation functions over the thickness of thin laminas. In their approach the cylinder was discretized into a number of thin laminas for which wave forms were considered. The dispersion equation was then obtained by applying Hamilton's principle, and the solution for the real dispersion curves was extracted using an iterative eigensolution technique. Huang and Dong[12] used a similar approach to study waves in isotropic and laminated composite cylinders. A spectral FE model was developed by Mahapatra and Gopalakrishnan[13] to analyze wave propagation in uniform composite thin cylinders using a first order shear theory. There have also been various applications of FEs to the study of wave propagation in spatially periodic structures.[14] One of the first applications was the work of Orris and Petyt,[15] where a FE approach and a receptance method[14] were applied for evaluating the dispersion curves of periodic structures. In his Ph.D. thesis, Abdel-Rahman[16] extended this FE approach to beams on periodic elastic supports, two-dimensional flat plates with periodic flexible stiffeners, and three-dimensional periodic beam systems. Multi-supported curved panels were studied by Pany and Parthan[17] using a similar approach.

The main aim of this work is to describe a wave finite element (WFE) method for cylindrical and curved structures. The method is an extension to axisymmetric structures of the WFE approach for two-dimensional structures of Mace and Manconi.[18] Compared to similar methods, the WFE method proposes a systematic and straightforward approach, which combines the theory of wave propagation in periodic structures with conventional FE analysis. As such, it is an application of FE analysis to periodic structures, although in the cases considered the structures are homogeneous, hence the periodicity of arbitrary length. For one-dimensional waveguides there have been several applications of the WFE method.[19–24] Some of these concern axisymmetric structures.[22–24] However, in these works a one-dimensional approach was used (wave propagation in the axial direction only), and axisymmetry was not exploited. Hence the WFE models were significantly larger, and it was not possible to characterize the motion in terms of circumferential orders *a priori*.

The structures of interest in the present work are homogeneous in the axial and circumferential directions, but their properties can vary in an arbitrary manner in the radial direction. Examples include isotropic, laminated, and sandwich cylindrical and curved panels. The method requires the analysis of just a small segment of the structure, which is discretized using conventional FEs. This involves a low order FE model: just a single rectangular FE or a stack of elements meshed through the cross-section. The mass and stiffness matrices of the FE model are then typically obtained using commercial FE packages. These matrices are subsequently postprocessed using periodicity conditions to obtain an eigenproblem whose solutions provide the dispersion curves and the wave modes. The form of the eigenproblem depends on the nature of the problem at hand. In particular for wave propagation in a closed cylinder, where the wavenumber around the circumference can only take certain discrete values, the eigenproblem is a quadratic eigenproblem. The general approach proposed is in contrast to the spectral finite element method (see, e.g., Ref. 13) in which new elements, with a space-harmonic displacement along the axis of the waveguide, must be derived on a case-by-case basis.

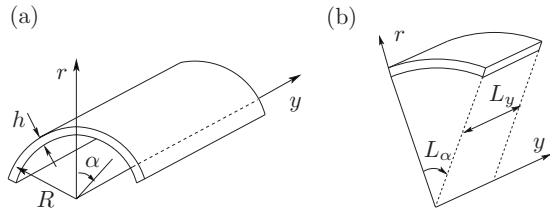


FIG. 1. (a) Axisymmetric structure and cylindrical coordinates; (b) small rectangular prismatic segment of the axisymmetric structure.

In the first part of the paper, the method is presented. Several numerical examples are then shown. The first two examples concern an isotropic cylinder and orthotropic cylindrical and curved panels, while the third example deals with the more significant case of a curved sandwich panel, which is made of two laminated skins sandwiching a soft foam core.

II. WFE FORMULATION FOR AXISYMMETRIC STRUCTURES

A curved axisymmetric structure is shown in Fig. 1(a), where y, r, and α are the cylindrical coordinates, R is the mean radius, and h is the thickness. A time-harmonic disturbance at a frequency ω is assumed to propagate through the structure with a helical pattern so that

$$w(r,\alpha,y,t) = W(r)e^{i(\omega t - k_\alpha \alpha - k_y y)}. \tag{1}$$

In Eq. (1), $W(r)$ is the complex wave amplitude, while k_α and k_y are the projections of the wavenumber k in the circumferential and axial directions. For real wavenumbers, $k_\alpha = k\cos\theta$ and $k_y = k\sin\theta$, where θ is the direction in which the wave propagates.

Exploiting the periodicity of the structure, a prismatic segment of length L_y subtending an angle L_α is taken, as shown in Fig. 1(b). This segment should be meshed in such a way that it has an equal number of nodes in both the α and y directions. If the periodic lengths L_α and L_y are small enough, the simplest way to discretize the segment is obtained either using just one shell element or generally using a stack of solid elements meshed through the cross-section, as shown in Fig. 2(a). The element type and the mesh size of the FE model should be chosen in accordance with standard FE practice for the frequency band of interest. For example, as a guideline there should be at least six elements per wavelength to ensure that FE discretization effects are acceptably small. In particular, a convergence analysis may be required. Emphasis here is placed on four noded shell elements or eight noded solid elements, but the use of other elements is straightforward.[18] The degrees of freedom (DOFs) of the FE model are arranged in a column vector $\mathbf{q}$ as

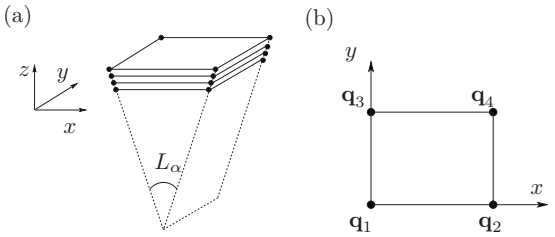


FIG. 2. (a) FE mesh of a small rectangular prismatic segment of the axisymmetric structure; (b) node numbering.

$$\mathbf{q} = [\mathbf{q}_1^T \quad \mathbf{q}_2^T \quad \mathbf{q}_3^T \quad \mathbf{q}_4^T]^T, \tag{2}$$

where the superscript T denotes the transpose and where $\mathbf{q}_j$ is the vector of the nodal DOFs of all the element nodes that lie on the jth corner of the segment, as shown in Fig. 2(b). A similar expression is given for the nodal forces

$$\mathbf{f} = [\mathbf{f}_1^T \quad \mathbf{f}_2^T \quad \mathbf{f}_3^T \quad \mathbf{f}_4^T]^T. \tag{3}$$

The vectors $\mathbf{q}$ and $\mathbf{f}$ are then the concatenation of the nodal DOFs and forces. Internal and edge nodes can also be included[18] but are not considered here.

The local coordinates must be rotated in order to model the desired curvature. A transformation matrix $\mathbf{R}$ is defined so that the mass and stiffness matrices of the "curved" element become

$$\mathbf{M} = \bar{\mathbf{R}}^T \mathbf{M}_{\text{loc}} \bar{\mathbf{R}},$$

$$\mathbf{K} = \bar{\mathbf{R}}^T \mathbf{K}_{\text{loc}} \bar{\mathbf{R}}, \tag{4}$$

where $\mathbf{M}_{\text{loc}}$ and $\mathbf{K}_{\text{loc}}$ are the element's mass and stiffness matrices in local coordinates, that is, the mass and stiffness matrices of the flat FE model. The matrix $\mathbf{R}$ allows for rotation around the y axis through angle L_α between nodes 1 and 2 and between nodes 3 and 4. It contains submatrices of the form

$$\begin{bmatrix} \cos(L_\alpha) & 0 & \mp\sin(L_\alpha) \\ 0 & 1 & 0 \\ \pm\sin(L_\alpha) & 0 & \cos(L_\alpha) \end{bmatrix} \tag{5}$$

for nodal displacements in the x, y, and z directions [Fig. 2(a)].

The equation of motion for the FE model, assuming time-harmonic behavior, is

$$(\mathbf{K} - \omega^2 \mathbf{M})\mathbf{q} = \mathbf{f}. \tag{6}$$

The presence of viscous or structural damping can be included by the addition of viscous or structural damping matrices. The structures of interest for this analysis can be considered as periodic structures in two dimensions. Therefore the free propagating wave can take the form of a Bloch wave,[25] and the propagation of a free wave can be obtained from the propagation constants

$$\mu_\alpha = k_\alpha L_\alpha \quad \text{and} \quad \mu_y = k_y L_y. \tag{7}$$

These propagation constants relate the displacements $\mathbf{q}$ on each side of the periodic element by

$$\mathbf{q}_2 = \lambda_\alpha \mathbf{q}_1, \quad \mathbf{q}_3 = \lambda_y \mathbf{q}_1, \quad \mathbf{q}_4 = \lambda_\alpha \lambda_y \mathbf{q}_1, \tag{8}$$

where

$$\lambda_\alpha = e^{-i\mu_\alpha}, \quad \lambda_y = e^{-i\mu_y}. \tag{9}$$

The nodal DOFs are rearranged to give

$$\mathbf{q} = \Lambda_R \mathbf{q}_1, \quad \Lambda_R = [\mathbf{I} \quad \lambda_\alpha \mathbf{I} \quad \lambda_y \mathbf{I} \quad \lambda_\alpha \lambda_y \mathbf{I}]^T. \tag{10}$$

In the absence of external excitation, equilibrium at node 1 implies that the sum of the nodal forces of all the elements connected to node 1 is zero. Consequently

$$\Lambda_L \mathbf{f} = \mathbf{0}, \quad \Lambda_L = [\mathbf{I} \quad \lambda_\alpha^{-1}\mathbf{I} \quad \lambda_y^{-1}\mathbf{I} \quad (\lambda_\alpha \lambda_y)^{-1}\mathbf{I}]. \tag{11}$$

Substituting Eq. (10) into Eq. (6) and premultiplying both sides of Eq. (6) by Λ_L, the equation of free wave motion takes the form

$$[\bar{\mathbf{K}}(\mu_\alpha, \mu_y) - \omega^2 \bar{\mathbf{M}}(\mu_\alpha, \mu_y)]\mathbf{q}_1 = \mathbf{0}. \tag{12}$$

The eigenvalue problem of Eq. (12) can also be written as

$$\bar{\mathbf{D}}(\omega, \lambda_\alpha, \lambda_y)\mathbf{q}_1 = 0, \tag{13}$$

where $\bar{\mathbf{D}}$ is the reduced dynamic stiffness matrix (DSM). If there are n DOFs per node, the nodal displacement and force vectors are $n \times 1$, the element mass and stiffness matrices are $4n \times 4n$ while the reduced matrices are $n \times n$.

By partitioning the DSM of the element in Eq. (6) as

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} & \mathbf{D}_{13} & \mathbf{D}_{14} \\ \mathbf{D}_{21} & \mathbf{D}_{22} & \mathbf{D}_{23} & \mathbf{D}_{24} \\ \mathbf{D}_{31} & \mathbf{D}_{32} & \mathbf{D}_{33} & \mathbf{D}_{34} \\ \mathbf{D}_{41} & \mathbf{D}_{42} & \mathbf{D}_{43} & \mathbf{D}_{44} \end{bmatrix}, \tag{14}$$

the reduced eigenvalue problem is given by

$$\begin{aligned} &[(\mathbf{D}_{11} + \mathbf{D}_{22} + \mathbf{D}_{33} + \mathbf{D}_{44})\lambda_\alpha \lambda_y + (\mathbf{D}_{12} + \mathbf{D}_{34})\lambda_\alpha^2 \lambda_y \\ &\quad + (\mathbf{D}_{13} + \mathbf{D}_{24})\lambda_\alpha \lambda_y^2 + \mathbf{D}_{32}\lambda_\alpha^2 + \mathbf{D}_{23}\lambda_y^2 + (\mathbf{D}_{21} + \mathbf{D}_{43})\lambda_y \\ &\quad + (\mathbf{D}_{31} + \mathbf{D}_{42})\lambda_\alpha + \mathbf{D}_{14}\lambda_\alpha^2 \lambda_y^2 + \mathbf{D}_{41}]\mathbf{q}_1 = \mathbf{0}. \end{aligned} \tag{15}$$

Since the mass and stiffness matrices in Eq. (6) are real, symmetric, and positive definite or semidefinite, for the partitions of the DSM it follows that $D_{ij} = D_{ij}^T$. Considering the transpose of Eq. (15) divided by $\lambda_\alpha \lambda_y$, it can be proved that the solutions come in pairs, involving λ_α, $1/\lambda_\alpha$, λ_y, and $1/\lambda_y$ for a given real frequency ω. These, of course, represent the same disturbance propagating in the four directions $\pm\theta$ and $\pi \pm \theta$.

Equations (12) and (15) give different forms of the eigenproblem relating λ_α, λ_y, and ω, whose solutions give FE estimates of the wave modes (eigenvectors) and dispersion relations for the continuous structure. The three different forms of the eigenvalue problem, which follow from Eq. (12) or Eq. (15), are described in the following subsections.

A discussion about numerical issues related to the application of the method can be found in Ref. 18.

A. WFE eigenvalue problem for helical waves

If μ_α and μ_y are prescribed and real, a linear eigenvalue problem results in ω whose solutions yield the propagating

waves. For real values of μ_α and μ_y, it can be proved that the reduced matrices in Eq. (12) are positive definite Hermitian matrices. Therefore, for any given value of the propagation constants, there will be n real positive eigenvalues ω^2 for which wave propagation is possible. The corresponding eigenvectors will define the wave modes at these frequencies. Although there are a certain number of solutions, not all of these represent wave motion in the continuous structure.[18] Some of them are, in fact, artifacts of the discretization by FE of the structure.

Secondly, when ω and θ are prescribed and k is regarded as the eigenvalue parameter, the resulting problem is either a polynomial eigenvalue problem or a transcendental eigenvalue problem whose solutions for k may be purely real, purely imaginary, or complex, denoting propagating, evanescent, or decaying waves, respectively. Hence λ_α and λ_y are of the form

$$\lambda_\alpha = e^{-i\mu_\alpha}, \quad \lambda_y = e^{-i\mu_y}, \quad \frac{\mu_y}{\mu_\alpha} = \frac{L_y}{L_\alpha}\tan\theta, \tag{16}$$

where μ_α and μ_y might be complex, but their ratio is real and given.

If the ratio $\mu_y/\mu_\alpha = m_2/m_1$ is rational, m_2 and m_1 being integers with no common divisor, the propagation constants can be written as $\mu_\alpha = m_1\sigma$ and $\mu_y = m_2\sigma$. Putting $\gamma = e^{i\sigma}$, the eigenvalue problem in Eq. (15) becomes

$$[\mathbf{A}_8\gamma^{2m_1+2m_2} + \mathbf{A}_7\gamma^{2m_1+m_2} + \mathbf{A}_6\gamma^{m_1+2m_2} + \mathbf{A}_5\gamma^{m_1+m_2} + \mathbf{A}_4\gamma^{2m_1} + \mathbf{A}_3\gamma^{2m_2} + \mathbf{A}_2\gamma^{m_2} + \mathbf{A}_1\gamma^{m_1} + \mathbf{A}_0]\mathbf{q}_1 = \mathbf{0}. \tag{17}$$

The matrices $\mathbf{A}$ are of order $n \times n$ so that Eq. (17) is a polynomial eigenvalue problem of order $2(m_1+m_2)$, which has $2n(m_1+m_2)$ solutions for γ.

If the ratio of μ_y/μ_α is irrational, Eq. (17) cannot formally be obtained. Hence Eq. (15) is rewritten in the following general form:

$$\mathbf{B}(\lambda_\alpha, \lambda_y)\mathbf{q}_1 = 0. \tag{18}$$

To avoid trivial solutions, $|\mathbf{B}|$ must be equal to zero. The function $|\mathbf{B}|$ is a complete polynomial function in the two complex variables λ_α and λ_y, i.e.,

$$[\mathbf{B}] = \sum_{p=0}^{p=2n}\sum_{q=0}^{q=2n} B_{pq}\lambda_\alpha^p\lambda_y^q. \tag{19}$$

Since

$$\lambda_\alpha = e^{-i\mu_\alpha} = e^{-ikL_\alpha\cos\theta},$$

$$\lambda_y = e^{-i\mu_y} = e^{-ikL_y\sin\theta}, \tag{20}$$

Eq. (19) then becomes a transcendental eigenvalue problem in k for given θ. Solutions of Eq. (19) can be sought by a number of numerical methods for root finding.[26]

B. The WFE eigenvalue problem for closed cylinders

The third form of the eigenvalue problem is associated with wave motion in closed cylinders. Equation (1) corresponds to the general case of a nonclosed axisymmetric structure so that k_α can, in principle, attain any value. In closed structures, the phase change of a wave as it propagates around the circumference must be a multiple of 2π. Therefore the circumferential wavenumber can only take the discrete values $k_\alpha = n$, $n = 0, 1, 2, \ldots$, which defines the circumferential order n of the mode. The modes n are independent and can be analyzed separately. Therefore

$$\lambda_\alpha = e^{-inL_\alpha} \tag{21}$$

is given for a given circumferential order n, and Eq. (13) becomes either a linear eigenproblem in ω^2 for a given λ_y or a quadratic eigenproblem in λ_y for a given ω. In the latter case, the polynomial eigenvalue problem takes the form

$$[\mathbf{A}_2\lambda_y^2 + \mathbf{A}_1\lambda_y + \mathbf{A}_0]\mathbf{q}_1 = 0, \tag{22}$$

where $\mathbf{A}_2 \neq 0$. The following standard linear companion form

$$\mathbf{L}(\lambda_y) = \begin{bmatrix} -\mathbf{A}_2^{-1}\mathbf{A}_1 & -\mathbf{A}_2^{-1}\mathbf{A}_0 \\ \mathbf{I} & \mathbf{0} \end{bmatrix} - \lambda_y \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{23}$$

can be considered, and the eigenvalues and the eigenvectors of Eq. (22) can be recovered from those of Eq. (23) using subroutines for the standard linear eigenproblem. The eigenvalues may be purely real, purely imaginary, or complex. Thus the complex frequency spectrum can be determined for any given circumferential mode order.

III. NUMERICAL EXAMPLES

In this section numerical examples are presented to illustrate the application of the WFE method. Results are obtained by postprocessing FE models found using a commercial FE package.

A. Isotropic cylindrical shell

The material properties for this cylindrical shell are as follows: Young's modulus $E = 192$ GPa, Poisson's ratio $\nu = 0.3$, and density $\rho = 7800$ kg/m^3. The ratio of the thickness-to-mean radius for the cylinder is $h/R = 0.05$, where $R = 1$ m. For this example, the rectangular four-noded element SHELL63 in ANSYS was used. The dimensions in the (x, y) plane were taken as $L_y = L_x = 1$ mm, and results were compared to those derived directly from the Flügge equations of motion[5] to investigate the accuracy of the WFE results. The WFE model results in 6DOFs after the WFE reduction. Figure 3 presents the dispersion curves for $n = 0$ and $n = 1$, breathing and bending modes, respectively. The nondimensional frequency $\Omega = \omega/\omega_r$ is introduced, where $\omega_r = 1/R\sqrt{E/[\rho(1-\nu^2)]}$ is the shell ring frequency evaluated by Love's first order theory.[5] The ring frequency is the first transition frequency for $n = 0$. Critical frequencies can be easily evaluated by the WFE method by solving Eq. (12) for $\mu_y = 0$ and $\mu_\alpha = nL_\alpha$. The three branches shown in Fig. 3 broadly correspond to flat-plate flexural, shear, and extensional waves. This behavior is particularly clear above the ring frequency. It is seen that the results computed from the WFE method agree well with those obtained by the Flügge theory.

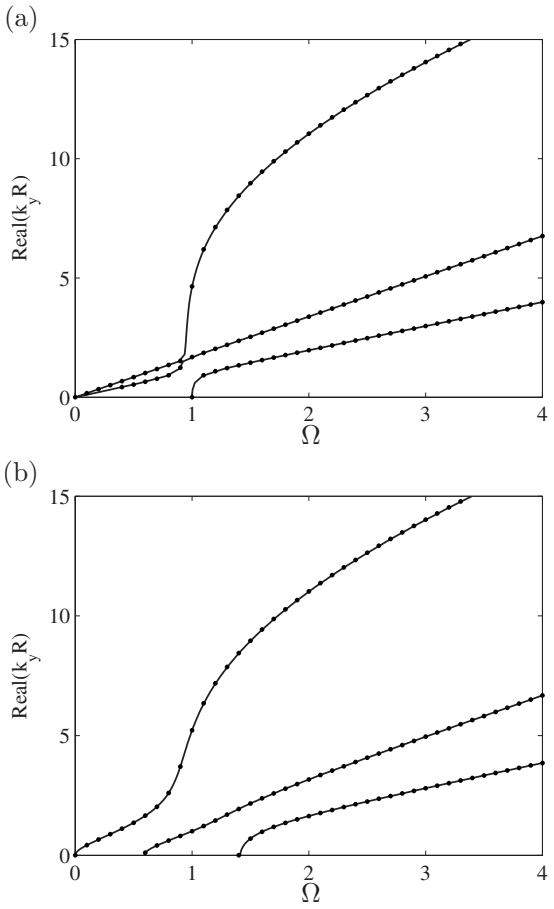


FIG. 3. Dispersion curves for the isotropic cylindrical shell. Circumferential modes: (a) $n=0$ and (b) $n=1$. (……) WFE results; ( ) analytic solution from the Flügge equations of motion.

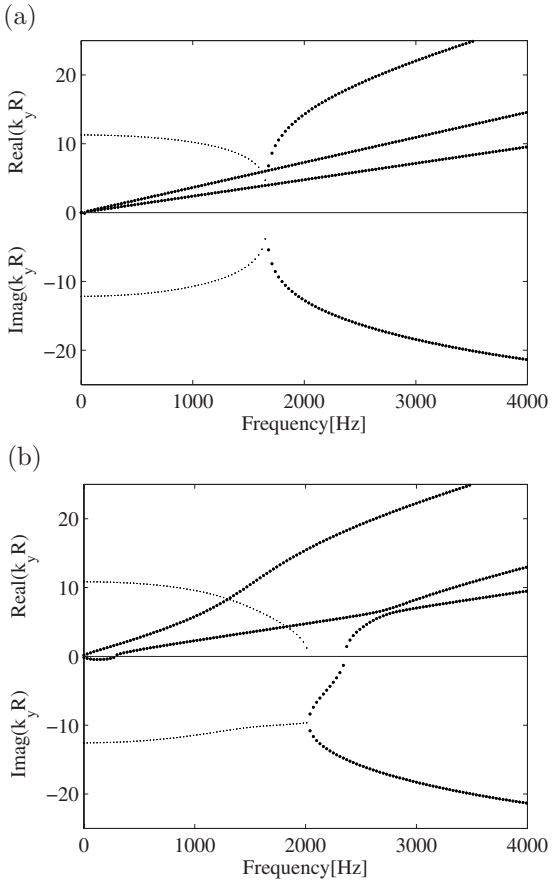


FIG. 4. Dispersion curves for the orthotropic cylindrical shell. Circumferential modes: (a) $n=0$ and (b) $n=1$. (…..) Complex valued wavenumbers; (**…..**) pure real and pure imaginary wavenumbers.

B. Orthotropic cylinder and orthotropic curved panel

The orthotropic panel considered in this section has mean radius $R=1$ m and thickness to mean radius ratio $h/R=0.05$. The material properties are $E_x=144.48$ GPa, $E_y=E_z=9.63$ GPa, $G_{xy}=G_{yz}=G_{xz}=4.128$ GPa, $\nu_{xy}=\nu_{xz}=0.02$, $\nu_{yz}=0.3$, and $\rho=1389$ kg/m^3. The FE model is realized using 20 SOLID45 elements in ANSYS, for which $L_x=L_y=1$ mm.

Figure 4 shows the complex dispersion curves for the breathing and bending modes obtained by solving the WFE eigenproblem in Eq. (22). Propagation in the positive y direction is considered. For $n=0$ [Fig. 4(a)], two waves start propagating at 0 Hz. They represent shear and extensional waves. At the ring frequency a third branch cuts on and the second branch veers, approaching a flexural wave in a flat plate. For $n=1$ [Fig. 4(b)], only one wave propagates from 0 Hz, while the second branch and the third branch cut on, respectively, below and above the ring frequency. Above the ring frequency, these branches approach flexural, shear, and extensional branches in a flat plate. Waves of orders $n>1$ can propagate only above a certain frequency. Figure 4 shows also complex branches, representing a pair of complex conjugate wavenumbers, for which the real and imaginary parts reduce with increasing frequency. In particular, for $n=1$, the corresponding imaginary part of the complex branch bifurcates into a pair of evanescent waves with pure imaginary wavenumbers at the frequency for which the corresponding real part becomes zero. It can be seen from Fig. 4(b) that this bifurcation point appears at a stationary point of the pure imaginary branch. Close to this region, the pure imaginary branch is double valued for a given frequency.

Figure 5 shows, as an example, the real part of the helical wave solutions for the heading directions $\theta=90°$, which corresponds to the circumferential mode order $n=0$, for $\theta=45°$, and for $\theta=0°$. The latter corresponds to wave propagation in the circumferential direction. At higher frequencies, higher order waves start propagating, which represent higher order wave modes across the thickness, e.g., branches 4–8 in Fig. 5(a). These wave modes can be evaluated analytically if the theory of elasticity is considered. The characteristics of the wave modes can be investigated by considering the eigenvectors ($\mathbf{q}_1$ in the WFE eigenproblems), which represent the deformations of the cross-section under the passage of the wave and the kinetic (or strain) energy under the passage of the wave associated with the individual DOFs. The kinetic (and strain) energy Γ follows from the mass (and stiffness) matrices in Eq. (4) and the eigenvectors $\mathbf{q}_1$ as

$$\Gamma = \tfrac{1}{2}[(i\omega\mathbf{V})^*\mathbf{M}(i\omega\mathbf{V})],$$

$$\mathbf{V} = [\mathbf{q}_1{}^T \quad \lambda_\alpha\mathbf{q}_1{}^T \quad \lambda_y\mathbf{q}_1{}^T \quad \lambda_\alpha\lambda_y\mathbf{q}_1{}^T]^T, \tag{24}$$

where the superscript * denotes the transpose conjugation. Figure 6 shows the deformation through the thickness under the passage of the waves in the fourth and fifth pure real

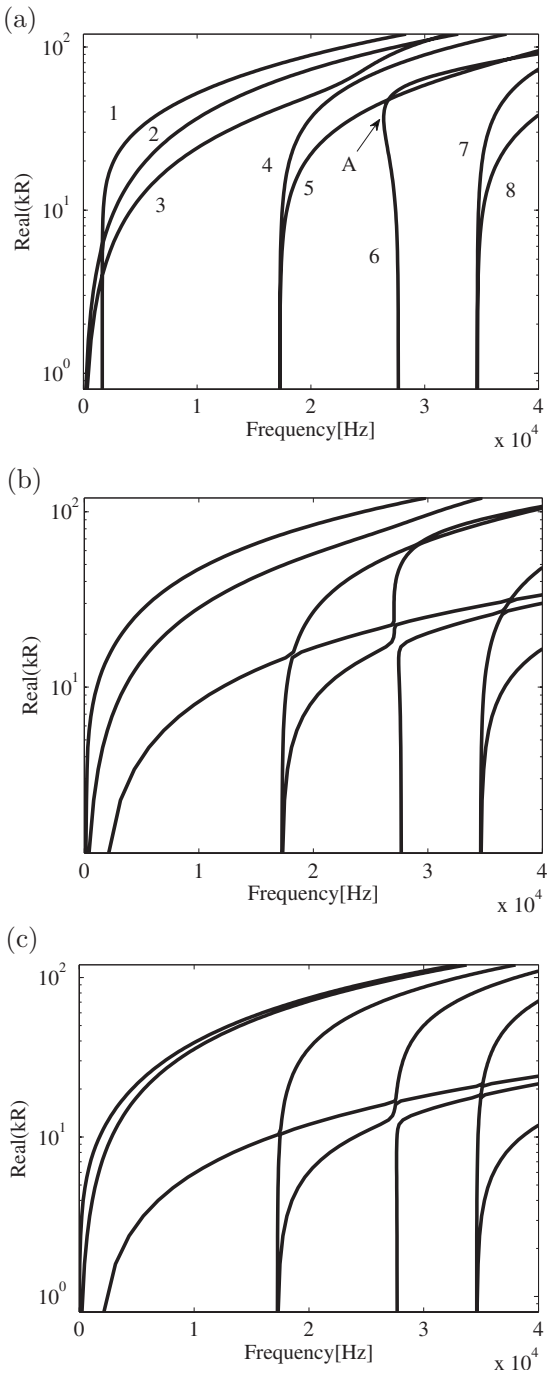


FIG. 5. Dispersion curves for the orthotropic curved panel. Heading direction: (a) θ=90°, (b) θ=45°, and (c) θ=0°.

branches at 20 kHz and in the sixth propagating branch at 30 kHz. The deformations are shown as functions of the position at one instant of time. These are indicative of the wave behavior, but they are not drawn to scale. Figure 6(a) shows that for branch 4 the deformations in the axial and radial directions (y and r directions, respectively) are negligible with respect to the antisymmetric deformation in the circumferential direction (α direction). The relative contributions of the kinetic energies in the circumferential, axial, and radial directions to the total kinetic energy are, respectively, $\Gamma_\alpha = 100\%\Gamma$, $\Gamma_y \approx 0\%\Gamma$, and $\Gamma_r \approx 0\%\Gamma$. Hence it can be concluded that branch 4 represents the first antisymmetric shear wave mode.

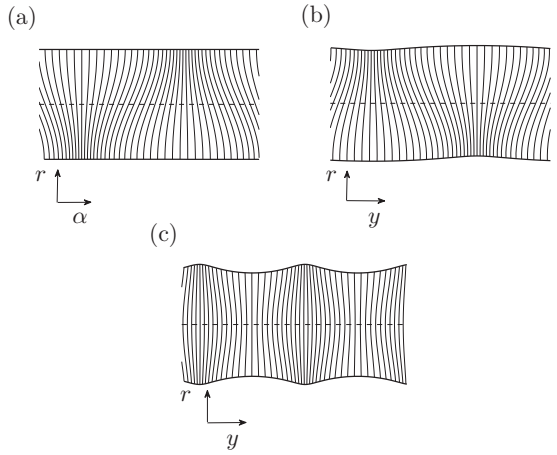


FIG. 6. Wave modes for the orthotropic curved panel, θ=90°: (a) branch 4 at 20 kHz, deformation in r and α directions; (b) branch 5 at 20 kHz, deformation in r and y directions; (c) branch 6 at 30 kHz, deformation in r and y directions (not to scale).

Branch 5 [Fig. 6(b)], is the first antisymmetric extensional wave mode, involving displacements primarily in the y direction although some Poisson contraction in the r direction can be observed. For this wave modes, $\Gamma_\alpha \approx 0\%\Gamma$, $\Gamma_y = 93.6\%\Gamma$, and $\Gamma_r = 6.4\%\Gamma$.

Branch 6 [Fig. 6(c)], involves primarily symmetric extensional motion in the axial direction. The out-of-plane displacement due to Poisson contraction is also significant. For this wave the contributions of the energies associated with the displacements in the α, y, and r directions to the kinetic energy are $\Gamma_\alpha \approx 0\%\Gamma$, $\Gamma_y = 73.6\%\Gamma$, and $\Gamma_r = 26.4\%\Gamma$. Branches 7 and 8 represent symmetric shear and extensional waves, respectively. Deformations are not shown here for the sake of brevity. These higher order wave modes are reminiscent of those for a flat panel. This is expected since the effects of the curvature are not significant at very high frequency. The behavior of branch 6 seems to not be significantly affected by changing the circumferential mode order. This branch starts propagating below the usual critical frequency, point A in Fig. 5. At this point a complex branch bifurcates into propagating waves with real wavenumbers, as shown better in Fig. 7. Above the bifurcation point A, the magnitude of one real wavenumber increases with frequency,

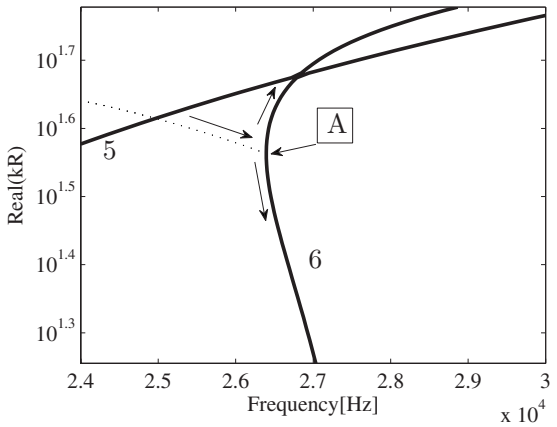


FIG. 7. Figure 5(a) in a region close to the bifurcation point A: (……) real part of complex valued wavenumbers; ( ) pure real wavenumbers.

while that of the second decreases with frequency. Close to this region, branch 6 exhibits wavenumbers and, consequently, phase velocities (the inverse of the slope of the chord between the point and the origin), which are double valued. The waves with smaller wavenumbers have phase and group velocities (the reciprocal of the slope of the dispersion curve at the point) of opposite signs. The behavior of this branch is similar to that of the S1 Lamb mode in a flat plate. The phenomenon has been theoretically predicted and then experimentally studied for plates by some authors.[27,28] At high frequencies, branch 3 and branch 2 cross. In particular the shape of the wave mode represented by branch 3 changes. This is perhaps due to the increasing Poisson contraction in the r direction. It can also be noticed that branch 6 crosses branch 7. At the crossover points, the frequency and the phase velocities are the same for both branches 6 and 7, but the group velocities are different.

C. Laminated sandwich cylinder and laminated sandwich curved panel

The sandwich panel comprises two laminated skins sandwiching a foam core. The two skins each comprise four orthotropic sheets of graphite epoxy with a lay up of $[+45/-45/-45/+45]$ and a total thickness of 4 mm. The material properties of the skins are the same as those given in Sec. III B for the orthotropic material. The core is a 10 mm polymethacrylamide ROHACELL foam whose material properties are Young's modulus $E=1.8\times10^8$ Pa, Poisson's ratio $\nu=0.286$, and density $\rho=110$ kg/m^3. The nondimensional thickness of the sandwich construction is $h/R=0.018$ with mean radius $R=1$ m. The WFE model was realized using 18 SOLID45 elements in ANSYS, four for each skin and ten for the core, resulting in 57DOFs after the reduction. Sandwich composite curved panels were studied by Heron[29] and later by Ghinet *et al.*[30] using a discrete layer theory. Heron assumed a classical theory for sandwich structures (a thick core that carries shear stress and thin skins that work in bending and extension), while Ghinet and Atalla used the Flügge theory to describe the strain-displacement relations. The two models lead to the 47th and 42nd order dispersion systems, respectively, which were subsequently solved numerically to determine the dispersion relations. Figure 8 shows the complex dispersion curves for waves propagating in the positive y direction for circumferential orders $n=0$, $n=1$, and $n=3$. The ring frequency for this sandwich cylinder is 622.7 Hz. The dispersive behavior is very complex and cannot be described simply in terms of torsional, extensional, and flexural waves alone. However some features can be observed, particularly concerning cut-off and cut-on with nonzero wavenumber and bifurcations between various wave modes. As an example, consider Fig. 8(b). With a behavior similar to that already shown for branch 6 in Sec. III B, it can be seen that between 0 and 600 Hz, as the imaginary part of the first complex branch becomes zero, its real part cuts off and bifurcates into two propagating waves, point A in Fig. 8(b). One of these propagating branches has a wavenumber that increases with frequency and phase and group velocities of the same sign, while the other has a wavenumber that

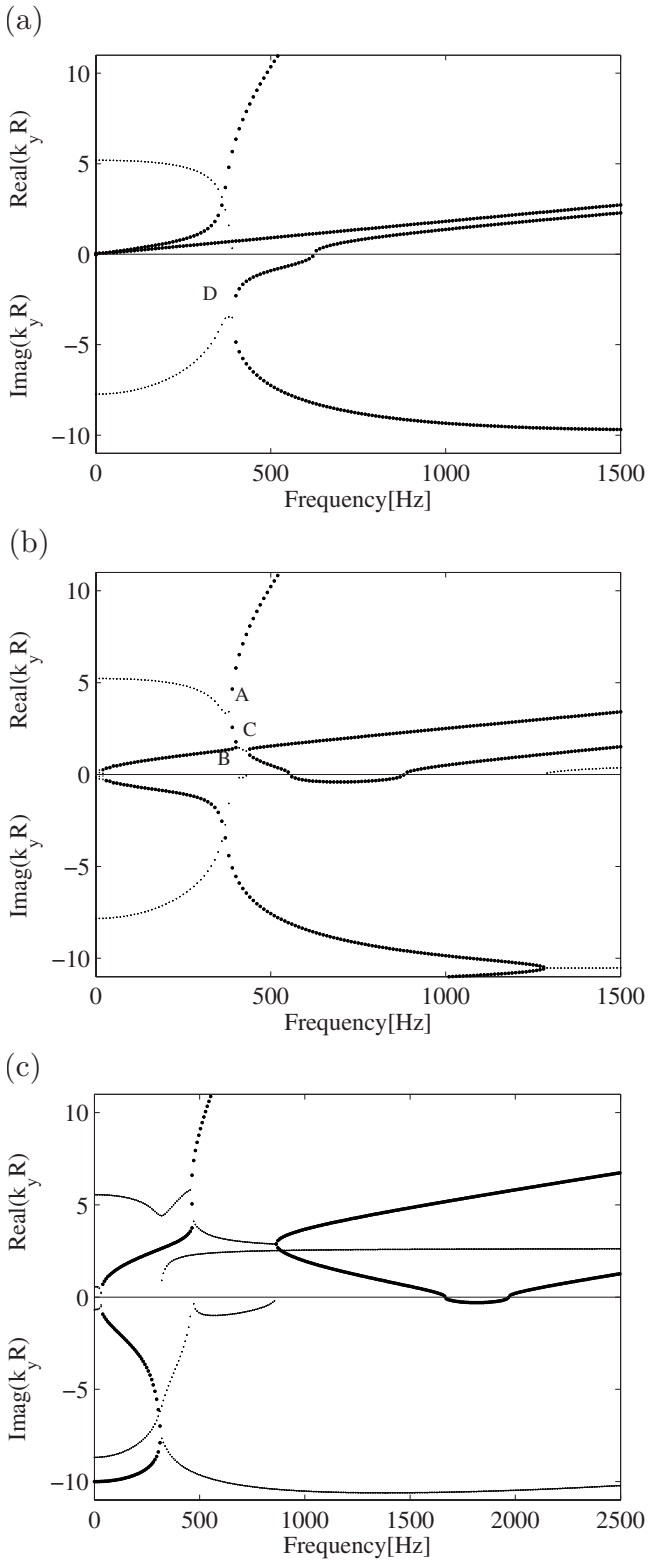


FIG. 8. Dispersion curves for the laminated sandwich cylindrical shell. Circumferential modes: (a) $n=0$, (b) $n=1$, and (c) $n=3$. (.....) Complex valued wavenumbers; (**.....**) pure real and pure imaginary wavenumbers.

decreases with frequency and has phase and group velocities of opposite signs. When this lower branch meets the first propagating branch (the first branch for which the critical frequency is zero), there is another bifurcation point, point B in Fig. 8(b). Here a complex branch cuts on. The real part of this second complex branch leaves the first propagating

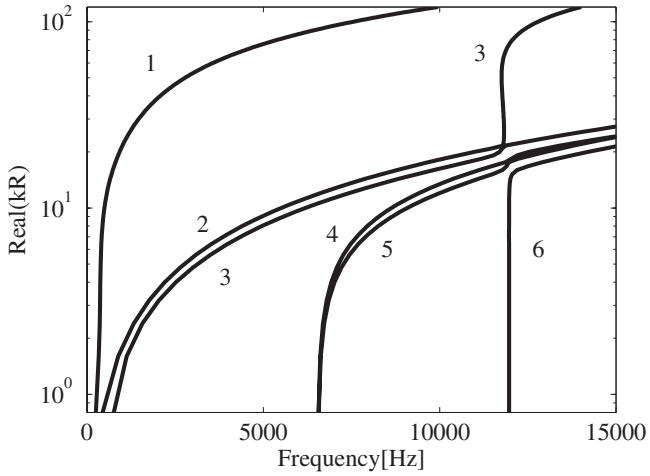


FIG. 9. Dispersion curves for the laminated sandwich curved panel. Heading direction $\theta=90°$ corresponding to the circumferential mode $n=0$.

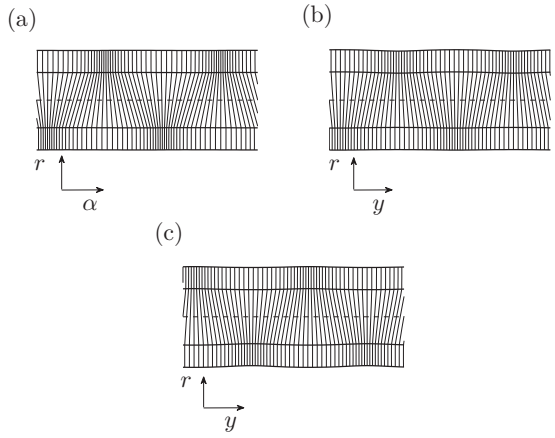


FIG. 11. Wave modes for the laminated sandwich cylindrical shell: (a) branch 4 at 10 kHz, deformation in α and r directions; (b) branch 5 at 10 kHz, deformation in y and r directions; (c) branch 6 at 14 kHz, deformation in y and r directions (not to scale).

branch at a stationary point, point B in Fig. 8(b), and reenters at a stationary point of the next pure real branch, point C in Fig. 8(b). At point C again two waves, one of these having phase and group velocities of opposite signs, start propagating. In this region pure real branches exhibit more than one possible value of k_yR for the same value of the frequency. Similar observations can be drawn for the dispersion curves in Fig. 8(c). The imaginary solutions also exhibit a similar behavior. Consider Fig. 8(a). When the real part of the first complex branch becomes zero, there is bifurcation from the corresponding imaginary part of the first complex branch to two different purely imaginary branches, point D in Fig. 8(a). For one of these branches, the imaginary wavenumber increases with frequency, and it represents rapidly decaying waves. The other one has an imaginary wavenumber that decreases with frequency until it becomes zero, where a third propagating branch cuts on at its critical frequency.

Higher order branches are shown in Fig. 9 for $\theta=90°$ or, equivalently, for the breathing mode shape. Figure 10 shows the deformations of the cross-section for the first three propagating wave branches in Fig. 9 at 2 kHz. In Fig. 10(a) it can be noticed that the wave of branch 1 at 2 kHz involves predominant bending of the skins. Displacements in the radial direction are almost constant, displacements in the circumferential direction are negligible, and displacements in the axial direction arise from bending and rotation of the laminas. The skins and the core are moving in antiphase. For this wave the contributions in the α, y, and r directions to the total kinetic energy are $\Gamma_\alpha \approx 0\%\Gamma$, $\Gamma_y=0.25\%\Gamma$, and $\Gamma_r=99.7\%\Gamma$, respectively. Figure 10(b) shows that the wave of branch 2 is a quasishear wave for which $\Gamma_\alpha=100\%\Gamma$, $\Gamma_y \approx 0\%\Gamma$, and $\Gamma_r \approx 0\%\Gamma$. The wave of branch 3 [Fig. 10(c)], resembles a quasiextensional wave mode involving primarily displacements in the y direction (some Poisson contraction can be noticed). The kinetic energies in the α, y, and r directions are $\Gamma_\alpha \approx 0\%\Gamma$, $\Gamma_y=93.4\%\Gamma$, and $\Gamma_r=6.6\%\Gamma$. As the frequency increases, branch 3 crosses branch 2, and its behavior changes: it approaches the first branch. The wave mode represented by branch 3 at a high frequency is evaluated at 14 kHz, and it is shown in Fig. 10(d). This mode involves axial motion of the core and a significant out-of-plane displacement. Both axial and out-of-plane displacements are symmetric. The kinetic energy is predominantly in the r direction, $\Gamma_r=97.3\%\Gamma$, while the contribution in the y direction is $\Gamma_y=2.7\%\Gamma$.

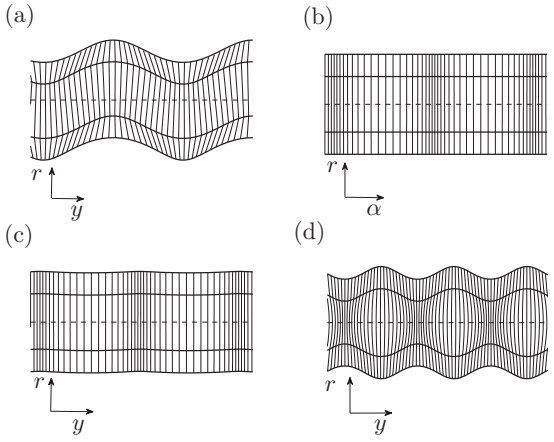


FIG. 10. Wave modes for the laminated sandwich curved panel: (a) branch 1 at 2 kHz, deformation in r and y directions; (b) branch 2 at 2 kHz, deformation in r and α directions; (c) branch 3 at 2 kHz, deformation in r and y directions; (d) branch 3 at 14 kHz, deformation in r and y directions (not to scale).

Deformations corresponding to higher order modes are shown in Fig. 11 for branches 4 and 5 at 10 kHz and for branch 6 at 14 kHz. Branch 4 [Fig. 11(a)], is the first antisymmetric quasishear wave mode. The kinetic energy for this wave is predominantly in the circumferential direction, while the contributions to the total kinetic energy in the y and r directions are negligible. Branch 5 [Fig. 11(b)], involves antisymmetric quasiextensional motion in the skins with some shear in the core. The distribution of the kinetic energy is $\Gamma_\alpha \approx 0\%\Gamma$, $\Gamma_y=99.09\%\Gamma$, and $\Gamma_r=0.91\%\Gamma$. Branch 6 [Fig. 11(b)] involves primarily axial motion of the core with Poisson contraction in the r direction. This wave resembles the wave of branch 5 at 10 kHz, as expected from Fig. 9.

The dispersion curves for the nonclosed curved sandwich panel can be represented in the form $\omega=f(k_\alpha,k_y)$, that

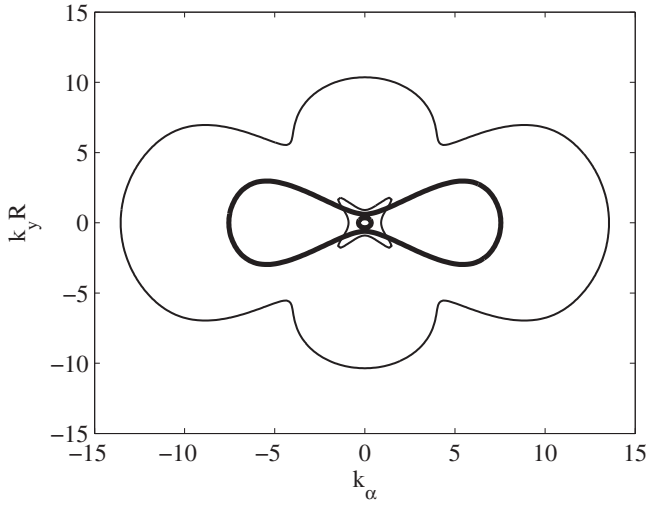


FIG. 12. Dispersion contour curves for the laminated sandwich curved panel: bold line, 200 Hz; thin line, 500 Hz.

is the constant frequency contour in the wavenumber plane. As an example, Figs. 12 and 13 show the dispersion contours at 200, 500, 800, and 1200 Hz. The curves with larger wavenumbers in Figs. 12 and 13 involve wave motion primarily in the r direction. The lower branches correspond to quasishear and quasiextensional wave modes.

Considerations about the direction of the group velocity and the direction of the wave propagation can be obtained by analyzing these isofrequency contour curves.[8] Since the group velocity is defined by $\mathbf{c}_g = d\omega/d\mathbf{k}$, the group velocity is in the direction of the normal to the contour curves in the (k_α, k_y) plane for constant ω. As an example, consider the direction of the group velocity for waves represented by the curve with larger wavenumbers at 200 Hz for $k_\alpha > 0$ and $k_y R > 0$ in Fig. 12. It can be noticed that waves having the same axial wavenumber exhibit normals to the contour curves, which have components with respect to the k_α direction of opposite sign. Thus they represent distinct waves having different directions of the group velocity in the circumferential direction.

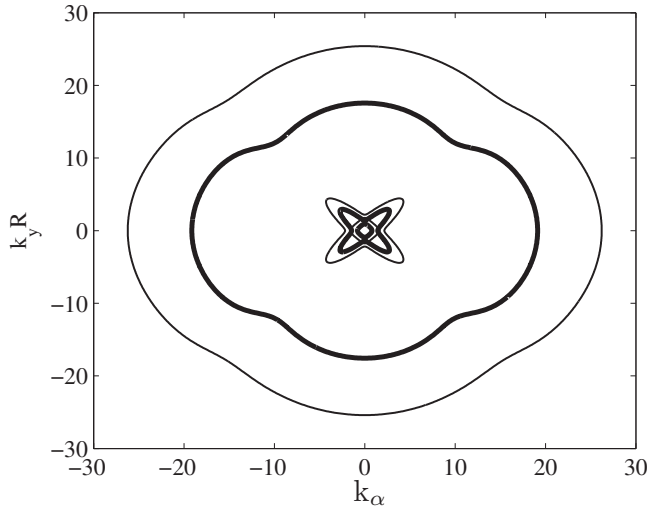


FIG. 13. Dispersion contour curves for the laminated sandwich curved panel: bold line, 800 Hz; thin line, 1200 Hz.

IV. CONCLUSION

In this paper the free wave characteristics of cylindrical and curved panels were predicted using a WFE method. The WFE method is a systematic and straightforward approach, which combines the theory of wave propagation in periodic structures with commercial FE tools. The technique involves the systematic postprocessing of element matrices of a small segment of the structure, which is modeled using either a single shell element or a stack of solid elements meshed through the cross-section. The full power of existing FE packages and their extensive element libraries is utilized. The method was described, and the WFE eigenproblem discussed. Several numerical examples were then presented. These include isotropic, orthotropic, and laminated sandwich shells. Comparison between the WFE results and analytical results for an isotropic cylindrical shell showed the accuracy of the method. Various wave propagation phenomena were observed, particularly concerning cut-off and bifurcations between various wave modes at high frequencies. The wave mode characteristics were identified, evaluating the cross-sectional deformations as functions of the distance and the kinetic energy associated with the individual DOFs. Dispersion contour curves for the laminated sandwich cylinder were also predicted. The application of the method was seen to be straightforward even in the complicated case of a laminated sandwich curved panel. The numerical results were obtained at negligible computational cost.

[1]K. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1991).
[2]D. C. Gazis, "Three dimensional investigation of the propagation of waves in hollow circular cylinders. I. Analytical foundation," J. Acoust. Soc. Am. **31**, 568–573 (1959).
[3]D. C. Gazis, "Three-dimensional investigation of the propagation of waves in hollow circular cylinders. II. Numerical result," J. Acoust. Soc. Am. **31**, 573–578 (1959).
[4]R. Kumar and R. W. B. Stephens, "Dispersion of flexural waves in circular cylindrical shells," Proc. R. Soc. London, Ser. A **329**, 283–297 (1972).
[5]A. W. Leissa, *Vibration of Shells* (U.S. GPO, Washington, D.C., 1973).
[6]P. M. Naghdi and R. M. Cooper,"Propagation of elastic waves in cylindrical shells, including the effects of transverse shear and rotatory inertia," J. Acoust. Soc. Am. **28**, 56–63 (1956).
[7]I. Mirsky and G. Herrmann, "Nonaxially symmetric motions of cylindrical shells," J. Acoust. Soc. Am. **29**, 1116–1123 (1957).
[8]R. S. Langley, "Wave motion and energy flow in cylindrical shells," J. Sound Vib. **169**, 29–42 (1994).
[9]V. V. Tyutekin, "Helical waves of an elastic cylindrical shell," Acoust. Phys. **50**, 273–277 (2004).
[10]V. V. Tyutekin, "Circumferential and helical normal waves of a cylindrical waveguide: Helical waves in a free space," Acoust. Phys. **52**, 471–476 (2006).
[11]R. B. Nelson, S. B. Dong, and R. D. Kalra, "Vibrations and waves in laminated orthotropic circular cylinders," J. Sound Vib. **18**, 429–444 (1971).
[12]K. H. Huang and S. B. Dong, "Propagating waves and edge vibrations in anisotropic composite cylinders," J. Sound Vib. **96**, 363–379 (1984).
[13]D. R. Mahapatra and S. Gopalakrishnan, "A spectral finite element for analysis of wave propagation in uniform composite tubes," J. Sound Vib. **268**, 429–463 (2003).
[14]D. J. Mead, "Wave propagation in continuous periodic structures: Research contribution from Southamtpon, 1964–1995," J. Sound Vib. **190**, 495–524 (1996).
[15]R. M. Orris and M. Petyt, "A finite element study of harmonic wave propagation in periodic structures," J. Sound Vib. **33**, 223–236 (1974).
[16]A. Abdel-Rahman, "Matrix analysis of wave propagation in periodic systems," Ph.D. thesis, University of Southampton (1979).
[17]C. Pany and S. Parthan, "Axial wave propagation in long periodic curved panels," J. Vibr. Acoust. **125**, 24–30 (2003).

[18]B. R. Mace and E. Manconi, "Modelling wave propagation in two-dimensional structures using finite element analysis," J. Sound Vib. **318**, 884–902 (2008).
[19]B. R. Mace, D. Duhamel, M. J. Brennan, and L. Hinke, "Finite element prediction of wave motion in structural waveguides," J. Acoust. Soc. Am. **117**, 2835–2843 (2005).
[20]D. Duhamel, B. R. Mace, and M. J. Brennan, "Finite element analysis of the vibration of waveguides and periodic structures," J. Sound Vib. **294**, 205–220 (2006).
[21]F. Tressède, "Numerical investigation of elastic modes of propagation in helical waveguides," J. Acoust. Soc. Am. **121**, 3398–3408 (2007).
[22]L. Houillon, M. N. Ichchou, and L. Jezequel, "Wave motion in thin-walled structures," J. Sound Vib. **281**, 483–507 (2005).
[23]M. Maess, N. Wagner, and L. Gaul, "Dispersion curves of fluid filled elastic pipes by standard FE models and eigenpath analysis," J. Sound Vib. **296**, 264–276 (2006).
[24]J. M. Mencik and M. N. Ichchou, "Wave finite elements in guided elastodynamics with internal fluid," Int. J. Solids Struct. **44**, 2148–2167 (2007).
[25]L. Brillouin, *Wave Propagation in Periodic Structures* (Dover, New York, 1953).
[26]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 1988).
[27]J. Wolf, T. D. K. Ngoc, R. Kille, and W. G. Mayer, "Investigation of Lamb waves having a negative group velocity," J. Acoust. Soc. Am. **83**, 122–126 (1988).
[28]M. F. Werby and H. Überall, "The analysis and interpretation of some special properties of higer order symmetric Lamb waves: The case for plate," J. Acoust. Soc. Am. **111**, 2686–2691 (2002).
[29]K. Heron, "Curved laminates and sandwich panels within predictive SEA," Proceedings of the Second International AutoSEA Users Conference, Troy, MI (2002).
[30]S. Ghinet, N. Atalla, and H. Osman, "The transmission loss of curved laminates and sandwich composite panels," J. Acoust. Soc. Am. **118**, 774–790 (2005).

Generalized eigenfunctions of layered elastic media and application to diffuse fields

Ludovic Margerin[a)]
Centre Européen de Recherche et d'Enseignement des Géosciences de l'Environnement, Université Aix Marseille, Centre National de la Recherche Scientifique, Europôle Méditerranéen de l'Arbois, BP 80, 13545 Aix en Provence, France

(Received 1 March 2008; revised 9 August 2008; accepted 15 October 2008)

The spectral decomposition of the elastic wave operator in a layered isotropic half-space is derived by means of standard functional analysis methods. Particular attention is paid to the coupled *P-SV* waves. The problem is formulated directly in terms of displacements which leads to a 2×2 Sturm–Liouville system. The resolvent kernel (Green's function) is expressed in terms of simple plane-wave solutions. Application of Stone's formula leads naturally to eigenfunction expansions in terms of generalized eigenvectors with oscillatory behavior at infinity. The generalized eigenfunction expansion is employed to define a diffuse field as a white noise process in modal space. By means of a Wigner transform, we calculate vertical to horizontal kinetic energy ratios in layered media, as a function of depth and frequency. Several illustrative examples are considered including energy ratios near a free surface, in the presence of a soft layer. Numerical comparisons between the generalized eigenfunction summation and a classical locked-mode approximation demonstrate the validity of the approach. The impact of the local velocity structure on the energy partitioning of a diffuse field is illustrated.
 [DOI: 10.1121/1.3021312]

PACS number(s): 43.40.Hb, 43.20.Bi, 43.40.Ph [RLW] Pages: 164–174

I. INTRODUCTION

Since its introduction in elastodynamics by Weaver,[1] the diffuse field concept and the principle of equipartition of elastic waves have been successfully applied to both field[2] and laboratory experiments.[3] It is an important ingredient of the spectacular reconstruction of Green's function from thermal noise[4] or seismic coda waves.[5] The principle of equipartition can also be used to predict the partition of energy in diffuse fields.[1,2,6,7] Yet, the practical implementation of the equipartition principle in a configuration as simple as a half-space poses some serious technical difficulties.[7] The simplest mathematical formulation of equipartition relies on the existence of a complete set of eigenfunctions. In the past, simple waveguides or a homogeneous half-space has been considered.[6–8] In this work, we construct a complete set of eigenfunctions of the elastic wave equation from the general spectral theory of self-adjoint operators in Hilbert space.[9,10] The present approach is based on applied functional analysis and differs from that usually followed in the acoustical or seismological literature which heavily relies on complex integration techniques.[11,12] A diffuse field is then represented as a white noise distributed over the complete set of eigenfunctions independent of the nature of the spectrum (discrete/continuous) or medium (open/closed). The paper is organized as follows: In Sec. II, the spectral problem is introduced and general properties of the elastodynamic operator are summarized. In Sec. III, the key mathematical results are presented and illustrated on a simple problem. Sections IV and V present the central results of the paper. The spectral theory of a layered elastic half-space is derived in detail and applied to energy partitioning of diffuse fields. Other applications of the theory are briefly discussed in the Conclusion.

II. PROBLEM STATEMENT

The elastodynamic equation governing wave propagation in elastic solids can be written as

$$\partial_t^2|u\rangle = -\mathbf{L}|u\rangle, \tag{1}$$

or more explicitly in the position representation:

$$\partial_t^2 u_i(\mathbf{x}) = \frac{1}{\rho(\mathbf{x})}\partial_j C_{ijkl}\partial_k u_l(\mathbf{x}). \tag{2}$$

In Eq. (1), the minus sign has been introduced in order to deal with a positive definite operator.

In an isotropic medium where properties depend only on the depth coordinate z, the elastodynamic equation decouples into independent scalar (*SH*) and vectorial (*P-SV*) equations. Our goal is to obtain a complete set of eigenfunctions for the latter problem. Following Ref. 13, we reduce the three-dimensional (3D) problem to a coupled system of second-order differential equations. Translational invariance suggests to look for solutions of the form $\mathbf{u}(z)e^{ikx}$. Additionally, cylindrical symmetry enables one to work in a single plane of propagation and ignore the third space dimension. Taking all the symmetries into account yields an eigenvalue problem for a second-order differential operator $\mathbf{L}_k$:[13]

[a)]Electronic mail: margerin@cerege.fr

$$\langle z|\mathbf{L}_k|u\rangle = -\frac{1}{\rho(z)}\partial_z\langle z|\tau|u\rangle + kB^t\partial_z\langle z|u\rangle + k^2C\langle z|u\rangle$$
$$= \lambda\langle z|u\rangle,$$
$$\langle z|\tau|u\rangle = A\partial_z\langle z|u\rangle + k\rho B\langle z|u\rangle. \tag{3}$$

The operator τ provides the tractions generated by the displacement field $|u\rangle$ and ρ denotes the density. The matrices A, B, and C are defined as

$$A = \rho(z)\begin{pmatrix} c^p(z)^2 & 0 \\ 0 & c^s(z)^2 \end{pmatrix},$$
$$B = \begin{pmatrix} 0 & 2c^s(z)^2 - c^p(z)^2 \\ c^s(z)^2 & 0 \end{pmatrix}, \tag{4}$$
$$C = \begin{pmatrix} c^s(z)^2 & 0 \\ 0 & c^p(z)^2 \end{pmatrix},$$

where c^p and c^s denote the longitudinal and the shear wavespeed, respectively. The vector $|u\rangle$ has components (iu_z, u_x), where u_x and u_z denote the original displacement field. This change of variables can be represented by a unitary transformation U and turns the original problem into an equivalent easier problem. Indeed, the multiplication of the vertical component by i makes $\mathbf{L}_k$ a real operator which simplifies the calculations in the layered case. Once the eigenvectors of the transformed operator have been obtained, those pertaining to the original operator are recovered by application of the inverse operation $U^\dagger$. As shown previously,[14,15] the elastodynamic operator is positive and self-adjoint in the linear vector space of functions equipped with the scalar product

$$\langle u|v\rangle = \int dz\rho(z)u_i(z)^*v_i(z). \tag{5}$$

The precise domain of definition of $\mathbf{L}_k$ is given in Ref. 14 but, loosely speaking, the operator acts on functions of finite elastic deformation energy. Depending on the problem at hand, the integral (5) is taken on the whole real line (section 3) or on a semi-infinite interval (section 4). In Eqs. (3) and (5), we have introduced the Dirac bra-ket notation for its compactness and convenience. Since it will be used throughout the paper, we have summarized the main properties of the Dirac formalism in the Appendix.

III. STONE FORMULA AND A SIMPLE APPLICATION

A. Mathematical results

The Stone formula[16] is a powerful functional analysis result for self-adjoint operators, and is particularly useful for the mathematical formulation of scattering theory.[9,10,17] In this paper, we will formally apply this formula to the elastic wave equation to obtain generalized eigenfunction expansions. The generalized eigenvectors $|e\rangle$ are solutions of the equation $\mathbf{L}_k|e\rangle = \lambda|e\rangle$ but do not belong to the domain of the operator, i.e., they are not solutions with finite energy. These eigenvectors are required to construct the complete modal solutions of the elastic wave equation.[18] Our work generalizes previous results obtained in the case of a homogeneous half-space.[14,19] Before stating the Stone formula, we introduce the resolvent $\mathbf{G}(\lambda)$ of the operator $\mathbf{L}_k$:

$$\mathbf{G} = (\mathbf{L}_k - \lambda\mathbf{I})^{-1}. \tag{6}$$

Because $\mathbf{L}_k$ is self-adjoint, its spectrum is a subset of the real axis and therefore the resolvent is defined for $\mathrm{Im}\,\lambda \neq 0$. In position and polarization space, the resolvent is represented by the Green's matrix

$$\langle z|(\mathbf{L}_k - \lambda\mathbf{I})^{-1}|z'\rangle = G_{ij}(\lambda, z, z'), \tag{7}$$

which possesses the Hermitian symmetry

$$G_{ij}(\lambda, z, z') = G_{ji}(\lambda^*, z', z)^*. \tag{8}$$

We first recall the fundamental spectral theorem and the Stone formula in an abstract setting. To every self-adjoint operator $\mathbf{L}_k$, one can associate a projection operator valued function (measure) $\mathbf{P}_\lambda$ having the following properties:[16]

$$f(\mathbf{L}_k) = \int_{-\infty}^{+\infty} f(\lambda)d\mathbf{P}_\lambda. \tag{9}$$

The spectral theorem guarantees that the spectral family $\mathbf{P}_\lambda$ is orthogonal, diagonalizes the operator $\mathbf{L}_k$, and provides the completeness relation. A practical method to construct the spectral projector $\mathbf{P}_\lambda$ is provided by the Stone formula, which relates functions of a self-adjoint operator to the discontinuity of its resolvent across the real axis:[16]

$$f(\mathbf{L}_k) = \frac{1}{2\pi i}\lim_{\epsilon\to 0^+}\int_{-\infty}^{+\infty}[\mathbf{G}(\lambda + i\epsilon) - \mathbf{G}(\lambda - i\epsilon)]f(\lambda)d\lambda. \tag{10}$$

For a positive operator, the integral can be taken from 0 to $+\infty$ only because the spectrum is a subset of the positive real axis. As a particular case of Eq. (10), one can formally obtain the completeness relation:

$$\mathbf{I} = \frac{1}{2\pi i}\lim_{\epsilon\to 0^+}\int_{-\infty}^{+\infty}[\mathbf{G}(\lambda + i\epsilon) - \mathbf{G}(\lambda - i\epsilon)]d\lambda. \tag{11}$$

For a positive operator with a mixed spectrum, with both discrete and continuous parts, Eq. (11) will take the following form:

$$\mathbf{I} = \int_{a\geq 0}^{+\infty} d\lambda\sum_m |e_m(\lambda)\rangle\langle e_m(\lambda)| + \sum_n |e_n\rangle\langle e_n|, \tag{12}$$

where $|e_m(\lambda)\rangle$ denotes generalized eigenfunctions of the continuum—we assume that no singular continuous spectrum exists—that are normalized in the following sense:

$$\langle e_m(\lambda)|e_{m'}(\lambda')\rangle = \delta_{mm'}\delta(\lambda - \lambda'). \tag{13}$$

In Eq. (12), we have introduced a discrete index in the continuous part of the spectrum, which corresponds to the possibility that the space of generalized eigenfunction be multidimensional. Such will be the case for the operator $\mathbf{L}_k$.

B. Generalized eigenvectors in homogeneous full space

We consider the system (3) with constant parameters c_∞^p, c_∞^s, and ρ_∞. First, we construct the resolvent $(\mathbf{L}_k-\lambda\mathbf{I})^{-1}$ outside of the real axis. In order to uniquely define the square root of a complex number z, we adopt the following convention: $\sqrt{z}=|z|^{1/2}e^{i\theta/2}$, with $\theta=\arg(z)\in[0,2\pi)$. This definition guarantees that $\mathrm{Im}\sqrt{z}>0$ when z lies outside the positive real axis. The 2×2 system of coupled second-order equation (3) has four linearly independent solutions $|u_{+\infty}^P(\lambda)\rangle$, $|u_{-\infty}^P(\lambda)\rangle$, $|u_{+\infty}^S(\lambda)\rangle$, and $|u_{-\infty}^S(\lambda)\rangle$, whose properties are summarized below:

$$\langle z|u_{\pm\infty}^P(\lambda)\rangle=\frac{1}{\epsilon_P(\lambda)}\begin{pmatrix}\pm iq_P(\lambda)\\ k\end{pmatrix}e^{\pm iq_P(\lambda)z},$$
$$\langle z|u_{\pm\infty}^S(\lambda)\rangle=\frac{1}{\epsilon_S(\lambda)}\begin{pmatrix}k\\ \mp iq_S(\lambda)\end{pmatrix}e^{\pm iq_S(\lambda)z}. \tag{14}$$

In Eq. (14), the prefactors $\epsilon_{P,S}$ ensure that the plane waves are energy normalized:[13]

$$\epsilon_{P,S}=\frac{1}{\sqrt{2\rho_\infty\sqrt{\lambda}q_{P,S}}}, \tag{15}$$

and the vertical wavenumbers $q_{P,S}$ are defined as

$$q_P=\sqrt{\frac{\lambda}{{c_\infty^p}^2}-k^2},\quad q_S=\sqrt{\frac{\lambda}{{c_\infty^s}^2}-k^2}. \tag{16}$$

The following symmetry relations will prove useful:

$$\langle z|u_{+\infty}^P(\lambda^*)\rangle=-\langle z|u_{+\infty}^P(\lambda)\rangle^*,$$
$$\langle z|u_{-\infty}^P(\lambda^*)\rangle=-\langle z|u_{-\infty}^P(\lambda)\rangle^*, \tag{17}$$

with completely analogous relations involving $|u_{+\infty}^S(\lambda)\rangle$, $|u_{-\infty}^S(\lambda)\rangle$. When $\mathrm{Im}\,\lambda=0$, $\lambda>(c_\infty^pk)^2$, assuming a time dependence of the form $e^{-i\sqrt{\lambda}t}$, $(u_{+\infty}^P,u_{-\infty}^P)$–$(u_{+\infty}^S,u_{-\infty}^S)$ behave like (outgoing, incoming) plane P-S waves at $+\infty$, respectively. Using this set of solutions, we can construct the Green's matrix for $\mathrm{Im}\,\lambda\neq0$, which obeys the following equation:

$$-\rho_\infty^{-1}\partial_z\tau[G]_{ij}(z,z')+K_{ik}(z)G_{kj}(\lambda,z,z')-\lambda G_{ij}(z,z')=\rho_\infty^{-1}\delta_{ij}\delta(z-z'), \tag{18}$$

where we have introduced the differential operator $K(z)=k^2C+kB^t\partial_z$. The Green's matrix should possess the following properties.

(1) For $z>z'$, each column of G must be a linear combination of solutions with finite energy at $+\infty$ $(|u_{+\infty}^P\rangle,|u_{+\infty}^S\rangle)$.
(2) For $z<z'$, each column of G must be a linear combination of solutions with finite energy at $-\infty$ $(|u_{-\infty}^P\rangle,|u_{-\infty}^S\rangle)$.
(3) G obeys the symmetry relation (8).
(4) G is continuous on the diagonal $z=z'$.
(5) The traction matrix of G has a jump discontinuity on the diagonal: $\tau[G]_{ij}|_{z'-}^{z'+}=-\delta_{ij}$.

Properties (1)–(5) are straightforward generalizations to systems of second-order differential equations of the standard Sturm–Liouville theory.[19] Properties (1)–(3) indicate that $\mathbf{G}(\lambda)$ should have the following form:

$$\mathbf{G}(\lambda)=\begin{cases}K_1(\lambda)|u_{-\infty}^P(\lambda)\rangle\langle u_{+\infty}^P(\lambda)^*|+K_2(\lambda)|u_{-\infty}^S(\lambda)\rangle\langle u_{+\infty}^S(\lambda)^*|, & z<z'\\ K_1(\lambda)|u_{+\infty}^P(\lambda)\rangle\langle u_{-\infty}^P(\lambda)^*|+K_2(\lambda)|u_{+\infty}^S(\lambda)\rangle\langle u_{-\infty}^S(\lambda)^*|, & z>z'.\end{cases} \tag{19}$$

In Eq. (19), we have separated the Green's matrix into a P and an S wave part. Application of the continuity condition of the Green's matrix on the diagonal (4) yields $K_1(\lambda)=K_2(\lambda)$. Next, we apply the condition (5) for the discontinuity of the traction on the diagonal to obtain

$$K_1(\lambda)=K_2(\lambda)=\frac{i}{\sqrt{\lambda}}. \tag{20}$$

The next step is to study the discontinuity of the resolvent across the real axis. We will denote by $|u_{+\infty}^{P+}\rangle$ and $|u_{+\infty}^{P-}\rangle$ the limiting values of the vectors $|u_{+\infty}^P(\lambda)\rangle$ as λ approaches the real axis from above and below, respectively, with analogous definitions for $|u_{-\infty}^{P+}\rangle$ and $|u_{+\infty}^{S+}\rangle,\ldots$. The jump of the resolvent $\mathbf{G}^+-\mathbf{G}^-$ across the real axis depends on the relations among these functions. For $z<z'$, one obtains

$$\mathbf{G}^+-\mathbf{G}^-=\frac{i}{\sqrt{\lambda^+}}(|u_{-\infty}^{P+}\rangle\langle(u_{+\infty}^{P+})^*|+|u_{-\infty}^{P-}\rangle\langle(u_{+\infty}^{P-})^*|+|u_{-\infty}^{S+}\rangle\langle(u_{+\infty}^{S+})^*|+|u_{-\infty}^{S-}\rangle\langle(u_{+\infty}^{S-})^*|). \tag{21}$$

Three cases have to be distinguished.

(1) $\lambda>(c_\infty^pk)^2$. In this case, $q_{P,S}$ and $\epsilon_{P,S}$ are real and all the waves are propagating. Applying the following symmetry relations:

$$|u_{\pm\infty}^{P+}\rangle=|(u_{\mp\infty}^{P+})^*\rangle,\quad |u_{\pm\infty}^{S+}\rangle=|(u_{\mp\infty}^{S+})^*\rangle,$$
$$|u_{\pm\infty}^{P-}\rangle=-|(u_{\pm\infty}^{P+})^*\rangle,\quad |u_{\pm\infty}^{S-}\rangle=-|(u_{\pm\infty}^{S+})^*\rangle, \tag{22}$$

one can rewrite the resolvent discontinuity (21) as a sum of projection operators on a generalized eigenfunction basis:

$$\mathbf{G}^+-\mathbf{G}^-=\frac{i}{\sqrt{\lambda^+}}(|u_{-\infty}^{P+}\rangle\langle u_{-\infty}^{P+}|+|u_{+\infty}^{P+}\rangle\langle u_{+\infty}^{P+}|+|u_{-\infty}^{S+}\rangle\langle u_{-\infty}^{S+}|+|u_{+\infty}^{S+}\rangle\langle u_{+\infty}^{S+}|). \tag{23}$$

From Eq. (23), one concludes that there are four linearly independent generalized eigenfunctions with eigenvalue

$\lambda > (c_\infty^p k)^2$. For instance, the first term on the right-hand side of Eq. (23) is an orthogonal projector on the space of upgoing P waves.

(2) $(c_\infty^p k)^2 > \lambda > (c_\infty^s k)^2$. q_P is pure imaginary and q_S is real which corresponds to evanescent P waves and propagating S waves. The symmetry relations for S waves are clearly the same as in Eq. (22). Upon application of the following symmetry relations for P waves:

$$|u_{\pm\infty}^{P+}\rangle = -i|(u_{\pm\infty}^{P+})^*\rangle \text{ and } |u_{\pm\infty}^{P-}\rangle = -i|u_{\pm\infty}^{P+}\rangle, \tag{24}$$

Eq. (21) simplifies to

$$\mathbf{G}^+ - \mathbf{G}^- = \frac{i}{\sqrt{\lambda^+}}(|u_{-\infty}^{S+}\rangle\langle u_{-\infty}^{S+}| + |u_{+\infty}^{S+}\rangle\langle u_{+\infty}^{S+}|). \tag{25}$$

There are two linearly independent S eigenfunctions.

(3) $\lambda < (c_\infty^s k)^2$. There are no singularities and the resolvent is continuous:

$$\mathbf{G}^+ - \mathbf{G}^- = \mathbf{0}. \tag{26}$$

Because of the symmetry of the Green's function (8), the same results are obtained for $z > z'$. The completeness relation now follows from the Stone formula (11). The continuous parameter λ is analogous to a squared eigenfrequency. Introducing $\lambda = \omega^2$ and generalized eigenvectors $|e_{+\infty}^P(\omega)\rangle, \ldots$, the resolution of the identity writes as a sum of integrals over frequencies:

$$\mathbf{I} = \int_{|c_\infty^p k|}^{+\infty} d\omega(|e_{+\infty}^P(\omega)\rangle\langle e_{+\infty}^P(\omega)| + |e_{-\infty}^P(\omega)\rangle\langle e_{-\infty}^P(\omega)|) + \int_{|c_\infty^s k|}^{+\infty} d\omega(|e_{+\infty}^S(\omega)\rangle\langle e_{+\infty}^S(\omega)| + |e_{-\infty}^S(\omega)\rangle\langle e_{-\infty}^S(\omega)|). \tag{27}$$

The eigenvectors $|e_{+\infty}^P(\omega)\rangle, \ldots$ can be deduced from the basis vectors $|u_{+\infty}^P(\lambda)\rangle$ and form an orthonormal set. Let us, for instance, consider the eigenvector $|e_{+\infty}^P(\omega)\rangle$:

$$\langle z|e_{+\infty}^P(\omega)\rangle = \frac{1}{\sqrt{2\pi\rho_\infty \omega q_P}}\begin{pmatrix} -iq_P \\ k \end{pmatrix} e^{+iq_P z},$$
$$\langle e_{+\infty}^P(\omega)|e_{+\infty}^P(\omega')\rangle = \delta(\omega - \omega'), \tag{28}$$

where $q_P = \sqrt{\omega^2/c_\infty^{p\,2} - k^2}$. Multiplying the wavefunctions $\langle z|e_{+\infty}^P(\omega)\rangle$ by the phase term $(1/\sqrt{2\pi})e^{ikx}$, we can form compound eigenvectors $|\psi^{P,S}\rangle$ that can be used to expand two-dimensional (2D) in-plane vector fields. The expansion takes a particularly simple form if instead of using a frequency integral, the new variables $p_z = \sqrt{\omega^2/c_\infty^{p\,2} - k^2}$ and $p_z = \sqrt{\omega^2/c_\infty^{s\,2} - k^2}$ are introduced in the P and S integrals of Eq. (27), respectively. In order to obtain eigenfunctions of the elastodynamic operator in its standard form, we apply the unitary transformation $U^\dagger$ to obtain a fairly simple completeness relation:

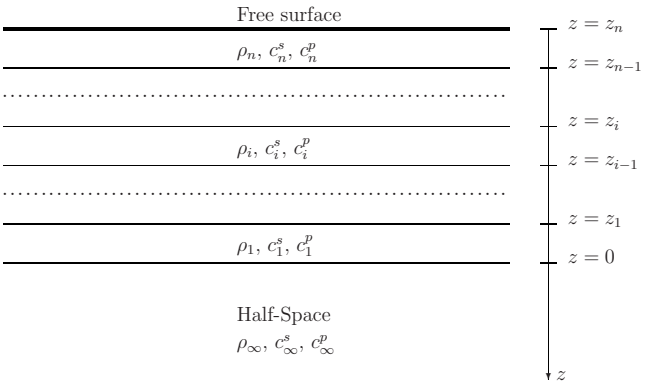


FIG. 1. Problem setting. We consider a stack of n elastic layers bounded by a free surface at $z=z_n$ and overlying a homogeneous half-space ($z>0$). The elastic properties of layer i are defined by the density ρ_i and the shear and compressional wavespeeds, denoted by c_i^s and c_i^p, respectively. We look for a complete set of eigenfunctions of the elastic wave equation satisfying the continuity of displacements and tractions across each interface $z=0,\ldots,z_{n-1}$ and the traction-free condition at the surface of the medium ($z=z_n<0$).

$$\mathbf{I} = \int\int_{\mathbb{R}^2} dp_x dp_z(|\psi^P(p_x,p_z)\rangle\langle\psi^P(p_x,p_z)| + |\psi^S(p_x,p_z)\rangle\langle\psi^S(p_x,p_z)|), \tag{29}$$

where $|\psi^P(p_x,p_z)\rangle$ and $|\psi^S(p_x,p_z)\rangle$ are simple plane P and SV waves:

$$\langle r|\psi^P(p_x,p_z)\rangle = \hat{p}\frac{e^{ip_x x + ip_z z}}{2\pi\sqrt{\rho_\infty}},$$
$$\langle r|\psi^S(p_x,p_z)\rangle = \hat{p}^\perp\frac{e^{ip_x x + ip_z z}}{2\pi\sqrt{\rho_\infty}}, \tag{30}$$

where $\hat{p}$ and $\hat{p}^\perp$ denote unit vectors parallel and perpendicular to the wavevector $\mathbf{p}$, respectively:

$$\hat{p} = \frac{1}{\sqrt{p_x^2 + p_z^2}}\begin{pmatrix} p_x \\ p_z \end{pmatrix}, \quad \hat{p}^\perp = \frac{1}{\sqrt{p_x^2 + p_z^2}}\begin{pmatrix} -p_z \\ p_x \end{pmatrix}. \tag{31}$$

$|\psi^P(p_x,p_z)\rangle$ and $|\psi^S(p_x,p_z)\rangle$ are eigenvectors of the elastic wave equation with eigenvalues $(p_x^2+p_z^2)c_\infty^{p\,2}$ and $(p_x^2+p_z^2)c_\infty^{s\,2}$, respectively. They form a complete set of generalized eigenfunctions which decompose 2D vector fields into longitudinal and transverse parts.

IV. GENERALIZED EIGENFUNCTIONS IN STRATIFIED MEDIA

A. Construction of the resolvent

We begin with the construction of the resolvent, which parallels the development of the previous section. To make comparison easier, we denote by ρ_∞, c_∞^s, and c_∞^p the density and seismic wavespeeds in the half-space located at depth $z>0$. The medium is assumed to be composed of a stack of n layers delimited by interfaces at $z=0,\ldots,z_i,\ldots,z_{n-1}$ and is bounded by a free surface at $z=z_n$ ($z_1,\ldots,z_n<0$), as depicted in Fig. 1. We wish to find the eigenfunction expansion associated with the eigenvalue problem $\mathbf{L}_k|u\rangle = \lambda|u\rangle$ introduced in

Eq. (3) and supplemented with the free surface traction condition $\langle z|\tau|u\rangle=0$ at $z=z_n$ and the continuity of stresses and displacements at each interface.

First, we introduce four linearly independent vector solutions of the elastic wave equation (3) denoted by $|u^P_{+\infty}(\lambda)\rangle$, $|u^P_{-\infty}(\lambda)\rangle$, $|u^S_{+\infty}(\lambda)\rangle$, and $|u^S_{-\infty}(\lambda)\rangle$. Although these solutions differ from those introduced in section 3, we employ the same notation because they share common properties. More precisely, in the half-space $z>0$, their analytical form is given by Eq. (14), in the previous section. The analytical dependence of the vectors $|u^{P,S}_{\pm\infty}(\lambda)\rangle$ in the stack of layers is obtained by integrating the equations of motion (3) from depth $z=0^+$ to $z=z_n^+$ by applying continuity of stresses and displacements at each interface. Note that, in general, these solutions do *not* verify the stress-free condition at $z=0$. For $\mathrm{Im}\,\lambda=0$ and $\mathrm{Re}\,\lambda>(c^p_\infty k)^2$, $|u^{P,S}_{+\infty}(\lambda)\rangle-(|u^{P,S}_{-\infty}(\lambda)\rangle)$ represent the outgoing (incoming) P and S waves at $z=+\infty$. For λ complex, only $|u^{P,S}_{+\infty}(\lambda)\rangle$ are bounded at $+\infty$.

Following Ref. 13, we introduce two linearly independent solutions $|u^P_0(\lambda)\rangle$ and $|u^S_0(\lambda)\rangle$ of the wave equation that verify the stress-free condition at $z=z_n$:

$$|u^P_0(\lambda)\rangle = |u^P_{-\infty}(\lambda)\rangle + r^{pp}(\lambda)|u^P_{+\infty}(\lambda)\rangle + r^{ps}(\lambda)|u^S_{+\infty}(\lambda)\rangle,$$

$$|u^S_0(\lambda)\rangle = |u^S_{-\infty}(\lambda)\rangle + r^{sp}(\lambda)|u^P_{+\infty}(\lambda)\rangle + r^{ss}(\lambda)|u^S_{+\infty}(\lambda)\rangle, \quad (32)$$

where r^{pp}, r^{ps}, r^{sp}, and r^{ss} are the reflection coefficients of the stack of layers including the free surface. More precisely, applying the free-surface condition at $z=z_n$ to Eq. (32) yields a system of two equations with two undetermined coefficients—(r^{pp}, r^{ps}) or (r^{ss}, r^{sp})—which can be solved since the vectors $|u^{P,S}_{+\infty}(\lambda)\rangle$ are linearly independent. For $\mathrm{Im}\,\lambda=0$, $\mathrm{Re}\,\lambda>(c^p_\infty k)^2$, the solutions (32) behave like propagating P and S waves incident from $+\infty$ together with their reflections. They are solutions of the wave equation with infinite energy that verify the free surface condition at $z=z_n$. For $\mathrm{Im}\,\lambda\neq 0$, we remark that because the matrix elements of $\mathbf{L}_k$ are real, the following relation holds: $\mathbf{L}_k|u^P_0(\lambda)^*\rangle = \lambda^*|u^P_0(\lambda)\rangle = \mathbf{L}_k|u^P_0(\lambda^*)\rangle$. Using Eq. (17), which is still valid in the stratified case, the following symmetry relations are established:

$$|u^{P,S}_0(\lambda^*)\rangle = -|u^{P,S}_0(\lambda)\rangle, \quad (33)$$

$$r^{ab}(\lambda^*) = r^{ab}(\lambda)^*, \quad (34)$$

where $a,b=p,s$. In addition, using the method of propagation invariants developed in Ref. 20, we can prove the following reciprocity relation:

$$r^{ps}(\lambda) = r^{sp}(\lambda). \quad (35)$$

To construct the Green's matrix, we again make use of the properties (2) and (3) of the previous section. Property (1) is replaced by the requirement that for $z<z'$ (z' is the source depth), the columns of G be linear combinations of the vectors $|u^{P,S}_0\rangle$ which are solutions of the wave equation satisfying the stress-free condition at $z=z_n$. For $z>z'$, we recall that the columns of the Green's matrix must be linear combinations of the vectors $|u^{P,S}_\infty\rangle$ which are bounded at $+\infty$ for λ complex. This suggests to try the following form for $\mathbf{G}(\lambda)$:

$$\mathbf{G}(\lambda) = \begin{cases} K_1(\lambda)|u^P_0(\lambda)\rangle\langle u^P_{+\infty}(\lambda)^*| + K_2(\lambda)|u^S_0(\lambda)\rangle\langle u^S_{+\infty}(\lambda)^*|, & z<z' \\ K_1(\lambda)|u^P_{+\infty}(\lambda)\rangle\langle u^P_0(\lambda)^*| + K_2(\lambda)|u^S_{+\infty}(\lambda)\rangle\langle u^S_0(\lambda)^*|, & z>z', \end{cases} \quad (36)$$

where use has been made of Eq. (34). In essence, Eq. (36) is completely similar to Eq. (19): the Green's function is separated into a P and an S wave part, although the vectors $|u^P_0\rangle$ and $|u^S_0\rangle$ are neither purely longitudinal nor purely transverse, due to the mode conversions at the medium boundaries. We must now determine the values of K_1 and K_2 in order to match the source conditions (4) and (5). When the source is located in the half-space $z>0$, we can make use of the analytical form of the vectors $|u^{P,S}_{+\infty}\rangle$ and $|u^{P,S}_0\rangle$, given in Eqs. (14) and (32). Following the same steps as in the homogeneous case and employing the reciprocity relation (35), one obtains

$$K_1(\lambda) = K_2(\lambda) = \frac{i}{\sqrt{\lambda}}. \quad (37)$$

Let us now demonstrate the validity of this expression for an arbitrary source depth. The main difficulty lies in the fact that the analytical form of the solutions $u^{P,S}_\infty$ and $u^{P,S}_0$ is not known in the stack of layers ($z_n<z<0$). Let us first introduce the following four-dimensional displacement-stress vectors:

$$w^{P,S}_\infty(\lambda,z) = \begin{pmatrix} u^{P,S}_{+\infty}(\lambda,z) \\ \tau[u^{P,S}_{+\infty}(\lambda,z)] \end{pmatrix}, \quad (38)$$

and

$$w^{P,S}_0(\lambda,z) = \begin{pmatrix} u^{P,S}_0(\lambda,z) \\ \tau[u^{P,S}_0(\lambda,z)] \end{pmatrix}. \quad (39)$$

For better readability, we employ standard vector and matrix notations in Eqs. (38) and (39) and the following. In order for the Green's matrix proposed in Eqs. (36) and (37) to verify the discontinuity of the displacement-stress vector for an *arbitrary* source depth, the following relation among the vectors $w^{P,S}_\infty$ and $w^{P,S}_0$ must hold at any $z>z_n$:

$$\frac{i}{\sqrt{\lambda}}(w^P_\infty(\lambda,z)w^P_0(\lambda,z)^t - w^P_0(\lambda,z)w^P_\infty(\lambda,z)^t + w^S_\infty(\lambda,z)w^S_0(\lambda,z)^t - w^S_0(\lambda,z)w^S_\infty(\lambda,z)^t) = N, \quad (40)$$

where we have introduced the notations:

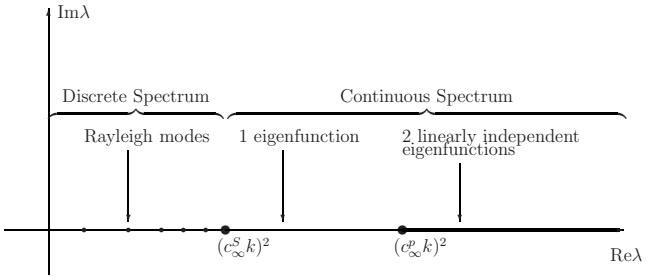


FIG. 2. Spectral properties of the elastodynamic operator in a layered half-space. λ is the eigenparameter.

$$N = \begin{pmatrix} 0 & -E \\ E & 0 \end{pmatrix} \text{ and } E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{41}$$

Note that we have already shown that Eq. (40) is valid for $z>0$. Indeed, by using the analytical expressions of the vectors $|u_{\pm\infty}^{P,S}\rangle$ in the half-space, we have been able to guess the form of the Green's matrix given by Eqs. (36) and (37). Now, the displacement-stress vector satisfies a first-order linear system $dw(\lambda,z)/dz = A(\lambda,z)w(\lambda,z)$, equivalent to the second order system (3), which can be integrated with the aid of a propagator matrix:

$$w(\lambda,z) = P(\lambda,z,z')w(\lambda,z'). \tag{42}$$

Since the propagator matrix has the symplectic symmetry:[13]

$$P(\lambda,z,z')NP(\lambda,z,z')^t = N, \tag{43}$$

relation (40) can be propagated at any depth in the stack of layers by left and right multiplications of Eq. (40) by P and its transpose, respectively. Therefore, relation (40) is established which in turn implies the validity of Eq. (37) for arbitrary source depth.

B. Singularities of the resolvent

The next step is to re-examine the discontinuity of the resolvent (36) across the real axis. This task is significantly facilitated by the unitary and symmetry properties of the reflection coefficients. Note that the unitary relations are valid in nondissipative media only. As shown in Ref. 20, the energy normalization of the eigenvectors is of fundamental importance in this respect. Other normalizations are, of course, possible but they lead to much more awkward calculations. We will denote by $|u(\lambda^+)\rangle - (|u(\lambda^-)\rangle)$ the limiting value of the ket $|u(\lambda)\rangle$ as λ approaches the real axis from above (below). The interrelations among the kets $|u_{+\infty}^{P,S}\rangle$ and $|u_{-\infty}^{P,S}\rangle$ across the real axis determine the form of the resolvent discontinuity. These relations, in turn, depend essentially on the analytical forms of these kets, i.e., on whether the P and S waves are propagating or evanescent in the half-space $z>0$. As in the homogeneous case, this leads to a separation of the real axis into three distinct domains, as shown in Fig. 2.

(1) For propagating P and S waves: $\lambda > (c_\infty^p k)^2$, the following relations apply:

$$|u_{\pm\infty}^P(\lambda^+)\rangle = |u_{\mp\infty}^P(\lambda^+)^*\rangle, \quad |u_{\pm\infty}^S(\lambda^+)\rangle = |u_{\mp\infty}^S(\lambda^+)^*\rangle,$$
$$|u_{\pm\infty}^P(\lambda^-)\rangle = -|u_{\pm\infty}^P(\lambda^+)^*\rangle, \quad |u_{\pm\infty}^S(\lambda^-)\rangle = -|u_{\pm\infty}^S(\lambda^+)^*\rangle. \tag{44}$$

These relations, together with the analytical properties of the reflection coefficients, enable us to write the kets $|u_0^{P,S}(\lambda^-)\rangle$ in terms of the kets $|u_{\pm\infty}^{P,S}(\lambda^+)\rangle$ only:

$$|u_0^P(\lambda^-)\rangle = -|u_{+\infty}^P(\lambda^+)\rangle - \overline{r^{pp}}|u_{-\infty}^P(\lambda^+)\rangle - \overline{r^{ps}}|u_{-\infty}^S(\lambda^+)\rangle,$$
$$|u_0^S(\lambda^-)\rangle = -|u_{+\infty}^S(\lambda^+)\rangle - \overline{r^{ss}}|u_{-\infty}^S(\lambda^+)\rangle - \overline{r^{sp}}|u_{-\infty}^P(\lambda^+)\rangle, \tag{45}$$

where the reflection coefficients are defined by their limiting value as λ approaches the real axis from above. For notational convenience, we have denoted the complex conjugation with an overbar. When both P and S waves are propagating in the half-space, the matrix of reflection coefficients is unitary:[20]

$$R = \begin{pmatrix} r^{pp} & r^{ps} \\ r^{sp} & r^{ss} \end{pmatrix}, \quad RR^\dagger = E. \tag{46}$$

For $z<z'$, the discontinuity of the resolvent writes

$$\mathbf{G}^+ - \mathbf{G}^- = \frac{i}{\sqrt{\lambda^+}}(|u_0^P(\lambda^+)\rangle\langle u_{-\infty}^P(\lambda^+)| + |u_0^S(\lambda^+)\rangle\langle u_{-\infty}^S(\lambda^+)| - |u_0^P(\lambda^-)\rangle\langle u_{+\infty}^P(\lambda^+)| - |u_0^S(\lambda^-)\rangle\langle u_{+\infty}^S(\lambda^+)|). \tag{47}$$

Using Eqs. (45) and (46) and after some algebra, one obtains

$$\mathbf{G}^+ - \mathbf{G}^- = \frac{i}{\sqrt{\lambda^+}}(|u_0^P(\lambda^+)\rangle\langle u_0^P(\lambda^+)| + |u_0^S(\lambda^+)\rangle\langle u_0^S(\lambda^+)|). \tag{48}$$

The same result applies for $z>z'$, and we have therefore put the discontinuity of the resolvent in suitable dyadic form. It appears that for $\lambda > (c_\infty^p k)^2$ there are two linearly independent eigenvectors. They are energy normalized incident P and S waves incident from $+\infty$ together with their reflections.

(2) For $(c_\infty^p k)^2 > \lambda > (c_\infty^s k)^2$, there are propagating S and evanescent P waves in the half-space $z>0$ and the following relations apply:

$$|u_{\pm\infty}^P(\lambda^+)\rangle = -i|u_{\pm\infty}^P(\lambda^+)^*\rangle \text{ and } |u_{\pm\infty}^S(\lambda^+)\rangle = |u_{\mp\infty}^S(\lambda^+)^*\rangle, \tag{49}$$

which yields

$$|u_0^P(\lambda^-)\rangle = -i|u_{-\infty}^P(\lambda^+)\rangle - i\overline{r^{pp}}|u_{+\infty}^P(\lambda^+)\rangle - \overline{r^{ps}}|u_{-\infty}^S(\lambda^+)\rangle,$$
$$|u_0^S(\lambda^-)\rangle = -|u_{+\infty}^S(\lambda^+)\rangle - \overline{r^{ss}}|u_{-\infty}^S(\lambda^+)\rangle - i\overline{r^{sp}}|u_{+\infty}^P(\lambda^+)\rangle. \tag{50}$$

For propagating S and evanescent P waves in the half-space, the unitary relations take a more complicated form:[20] $\overline{r^{ss}}r^{sp} = i\overline{r^{sp}}$. With the aid of Eq. (49), this implies $\overline{r^{ss}}|u_0^S(\lambda^+)\rangle = -|u_0^S(\lambda^-)\rangle$. For $z<z'$, using the unitary relations, the resolvent discontinuity can be put into the following form:

$$\mathbf{G}^+ - \mathbf{G}^- = -[i|u_0^P(\lambda^+)\rangle + |u_0^P(\lambda^-)\rangle]\langle u_{+\infty}^P(\lambda^+)| + |u_0^S(\lambda^+)\rangle[\langle u_{-\infty}^S(\lambda^+)| + \overline{r^{ss}}\langle u_{+\infty}^S(\lambda^+)|]. \quad (51)$$

Using the symmetry relation $r^{pp} - \overline{r^{pp}} = i\overline{r^{sp}}$, the following relation is established:

$$i|u_0^P(\lambda^+)\rangle + |u_0^P(\lambda^-)\rangle = -\overline{r^{sp}}|u_0^S(\lambda^+)\rangle, \quad (52)$$

which in turn implies

$$\mathbf{G}^+ - \mathbf{G}^- = \frac{i}{\sqrt{\lambda^+}}|u_0^S(\lambda^+)\rangle\langle u_0^S(\lambda^+)|. \quad (53)$$

This last equation has the desired form and shows that in the range of λ considered there is only one generalized eigenfunction. In the half-space, it consists of the sum of an incident and a reflected propagating S wave together with a reflected evanescent P wave. Note that for some λ in the range $](c_\infty^s k)^2, (c_\infty^p k)^2[$, it may happen that the vector $|u_{+\infty}^P\rangle$ satisfies the zero traction condition at the free surface. Such a situation can be considered as accidental, since a slight modification of the thickness or velocity in one of the layers is likely to make the mode disappear (Y. Colin de Verdière, personal communication). Such possible complications will be neglected.

(3) For $(c_\infty^s k)^2 > \lambda$, both P and S waves are evanescent in the half-space $z>0$. In this case, the reflection matrix R is simultaneously symmetric and Hermitian, so that all its coefficients are real. Further, using the relations

$$|u_{\pm\infty}^P(\lambda^-)\rangle = -i|u_{\pm\infty}^P(\lambda^+)\rangle \text{ and } |u_{\pm\infty}^S(\lambda^-)\rangle = -i|u_{\pm\infty}^S(\lambda^+)\rangle, \quad (54)$$

one establishes

$$|u_0^{P,S}(\lambda^-)\rangle = -i|u_0^{P,S}(\lambda^+)\rangle, \quad (55)$$

which seems to imply that the resolvent discontinuity vanishes identically. This is not true as there can exist values of λ for which the reflection coefficients have poles. For these particular points of the spectrum, the kets $|u_0^{P,S}(\lambda^+)\rangle$ can be written solely as linear combinations of the kets $|u_{+\infty}^{P,S}(\lambda^+)\rangle$.[13] This means that there exist eigenvectors with finite energy that satisfy the traction-free condition, which are the well-known Rayleigh surface waves. To each pole, we can associate a one-dimensional eigenspace spanned by a normalized Rayleigh wave eigenvector. For a given value of the horizontal wavenumber k, there exists m distinct solutions denoted by $0<\lambda_0<\lambda_1<\cdots<\lambda_{m-1}<(c_\infty^s k)^2$. The index serves to label the different surface wave mode branches. Note that the number of modes m is a function of the wavenumber k and that the index 0 refers to the fundamental Rayleigh mode.

C. Eigenfunction expansions

We have now identified all the singularities of the resolvent on the real axis and are in position to write down the completeness relation for 2D in-plane problems using formula (10). In the frequency domain, introducing the new variable $\omega^2=\lambda^+$, one obtains

$$\mathbf{I} = \int_{-\infty}^{+\infty} dk \sum_n |e^R(\omega_n)\rangle\langle e^R(\omega_n)| + \int_{-\infty}^{+\infty} dk \int_{|c_\infty^s k|}^{|c_\infty^p k|} d\omega |e^S(\omega)\rangle\langle e^S(\omega)| + \int_{-\infty}^{+\infty} dk \int_{|c_\infty^p k|}^{+\infty} d\omega (|e^S(\omega)\rangle\langle e^S(\omega)| + |e^P(\omega)\rangle\langle e^P(\omega)|), \quad (56)$$

where the compound generalized eigenvectors $|e^{S,P,R}\rangle$ are defined as

$$\langle r|e^{S,P}(\omega)\rangle = \sqrt{2}\frac{e^{ikx}}{\sqrt{2\pi}}\langle z|u_0^{S,P}(\omega^2)\rangle, \qquad \langle r|e^R(\omega_n)\rangle = \frac{e^{ikx}}{\sqrt{2\pi}}\langle z|u^R(\omega_n)\rangle. \quad (57)$$

In the first of Eq. (57), the $\sqrt{2}$ factor is the result of introducing of the new variable ω, whereas the last equation defines surface waves with unit-energy normalized z-eigenfunction. Although this is implicit in Eq. (57), the reader should keep in mind that all eigenvectors are functions of the horizontal wavenumber k. This applies to the eigenfrequencies ω_n as well as the number of branches. As in the case of the homogeneous space, it is possible to replace the frequency integral by a vertical wavenumber integral in Eq. (56) to obtain

$$\mathbf{I} = \int_{-\infty}^{+\infty} dp_x \sum_n |\psi^R(\omega_n)\rangle\langle\psi^R(\omega_n)| + \int_{-\infty}^{+\infty} dp_x \int_{-\infty}^{0} dp_z |\psi^S(p_x,p_z)\rangle\langle\psi^S(p_x,p_z)| + \int_{-\infty}^{+\infty} dp_x \int_{-\infty}^{0} dp_z |\psi^P(p_x,p_z)\rangle\langle\psi^P(p_x,p_z)|. \quad (58)$$

The eigenvectors $|\psi^{P,S,R}\rangle$ are obtained by application of the transformation $U^\dagger$ to the kets $|e^{S,P,R}\rangle$. The eigenvectors $|\psi^{P,S}\rangle$ correspond to properly normalized incoming P and S waves exactly as given in Eqs. (30) and (31) together with their reflections from the stack of layers, including the free surface. This is a general result in scattering theory: the eigenvectors of the medium+scatterer can be obtained by calculating the complete response of the scatterer to the unperturbed eigenvectors.[9,21] This is basically what the double integrals of Eq. (58) say. Note that by calculating the scattering of plane waves, one does not obtain the surface waves directly. They show up as poles of the reflection coefficient located on the real axis.

V. APPLICATION TO DIFFUSE FIELDS

A diffuse field is defined as a random field where all the modes are equally represented. More precisely, according to Ref. 1, it is a narrow band signal which is a sum of eigenfunctions of the system excited randomly at equal energy:

$$u(\mathbf{x},t) = \sum_{n,R,L} \int\int_{\omega_n \in B} \frac{dp_x dp_y}{2\pi} (A_n^{R,L}(p_x,p_y)\psi_n^{R,L}(z)e^{i(p_x x+p_y y)} \times e^{-i\omega_n^{R,L}(p_x,p_y)t}) + \int_{p_z<0} dp_z \sum_{P,S} \int\int_{\omega \in B} \frac{dp_x dp_y}{2\pi} \times (A^{P,S}(p_x,p_y,p_z)\psi_{p_z}^{P,S}(z)e^{i(p_x x+p_y y)}e^{-i\omega^{P,S}(p_x,p_y,p_z)t}). \quad (59)$$

In Eq. (59), we consider a semi-infinite 3D layered elastic body and write the displacement field as a sum over Love and Rayleigh surface waves (indicated by L and R superscripts), together with the generalized P and S (SH-SV) generalized eigenfunctions. B denotes the frequency band of the signal and u refers to the complex analytic signal whose real part is the measured displacement vector. In a diffuse field, the amplitudes $A_n^{R,L}$ and $A^{P,S}$ are assumed to be white-noise random processes:

$$\langle A_n^{R,L} A^{P,S}\rangle = 0,$$

$$\langle A_n^{R,L}(p_x,p_y)A_m^{R,L}(p_x',p_y')^*\rangle = \sigma^2 \delta^{R,L}\delta_{nm}\delta(p_x-p_x')\delta(p_y-p_y'), \quad (60)$$

$$\langle A^{P,S}(p_x,p_y,p_z)A^{P,S}(p_x',p_y',p_z')^*\rangle = \sigma^2 \delta^{P,S}\delta(p_z-p_z')\delta(p_x-p_x')\delta(p_y-p_y').$$

In Eq. (60), the symbol $\delta^{L,R}$ ($\delta^{P,S}$) equals 1 for surface waves (body waves) of the same type and 0 otherwise. In the case of a semi-infinite medium, the sum over all modes involves a continuous index that labels the eigenfunctions of the continuous spectrum and a discrete index that refers to the surface wave mode branch.

Formulas (59) and (60) are now applied to the calculation of the vertical to horizontal kinetic energy ratio of a diffuse field in a stratified half-space. The vertical kinetic energy density at central frequency ω_0 is obtained by means of the Wigner distribution of the complex analytic wavefield:

$$E_z(t,\tau,\mathbf{x}) = \tfrac{1}{2}\rho(z)\langle \partial_t u_z(t+\tau/2,\mathbf{x})\partial_t u_z(t-\tau/2,\mathbf{x})^*\rangle, \quad (61)$$

with similar expressions for the horizontal kinetic energies along axes x and y. In Eq. (61), the brackets denote an ensemble average. Inserting the spectral decomposition (59), applying the equipartition principle (60), and taking a Fourier transform with respect to the time variable τ, the local vertical kinetic energy density of a diffuse field at circular frequency ω_0 is obtained:

$$E_z(\omega_0,z) = \frac{\rho(z)\omega_0^2\sigma^2}{8\pi^2}\sum_{n,R,L}\int\int_{\mathbb{R}^2} dp_x dp_y |\psi_{n,z}^{R,L}(z)|^2 \times \delta(\omega_0-\omega_n^{R,L}(p_x,p_y)) + \frac{\rho(z)\omega_0^2\sigma^2}{8\pi^2}\int_{p_z<0} dp_z \sum_{P,S}\int\int_{\mathbb{R}^2} dp_x dp_y |\psi_{p_z,z}^{P,S}(z)|^2 \times \delta(\omega_0-\omega^{P,S}(p_x,p_y,p_z)). \quad (62)$$

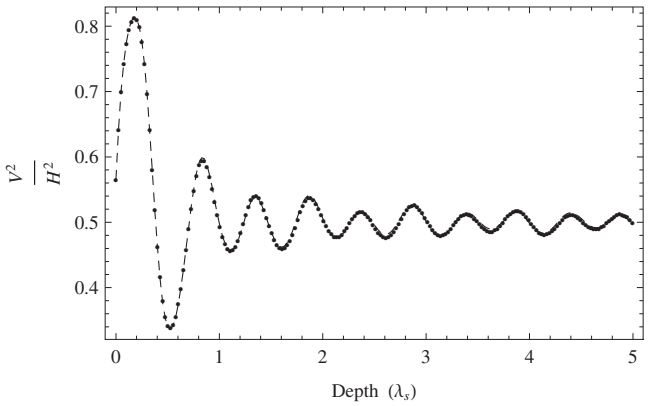


FIG. 3. Depth dependence of the vertical to horizontal kinetic energy ratio near the free surface of a Poisson half-space. Dots: locked-mode approximation. Dashed line: generalized eigenfunctions summation. The depth unit is the shear wavelength λ_S.

In Eq. (62), the subscript z refers to the vertical component of the eigenfunctions. The delta functions represent the density of states of the surface and body waves, which are closely related to the imaginary part of the Green's function.[22] The link between the Wigner distribution of the wavefield and the Green's function is the theoretical basis of the Green's function reconstruction in diffuse fields.[21,23,24] Introducing the cylindrical and spherical coordinates in the double and triple integrals of Eq. (62), respectively, one finds

$$E_z(\omega_0,z) = \sum_{n,R}\frac{\rho(z)\omega_0^3\sigma^2}{4\pi c_n^R u_n^R}|\psi_{n,z}^R(z)|^2 + \sum_{P,SV}\frac{\rho(z)\omega_0^4\sigma^2}{4\pi v_\infty^{p,s3}}\int_{\pi/2}^{\pi} d\theta \sin\theta \times |\psi_{p_z,z}^{P,SV}(z)|^2\Big|_{p_z=(\omega_0/v_\infty^{p,s})\cos\theta}. \quad (63)$$

In Eq. (63), we have introduced the following notations: c_n and u_n are the phase and group velocities of the nth surface wave mode, respectively. The calculation of the eigenfunctions and the remaining sums and integrals have to be performed numerically. Following similar reasoning, the total horizontal kinetic energy can be expressed as

$$E_h(\omega_0,z) = \sum_{n,R,L}\frac{\rho(z)\omega_0^3\sigma^2}{4\pi c_n^{R,L} u_n^{R,L}}|\psi_{n,h}^{R,L}(z)|^2 + \sum_{P,SV,SH}\frac{\rho(z)\omega_0^4\sigma^2}{4\pi v_\infty^{p,s3}}\int_{\pi/2}^{\pi} d\theta \sin\theta \times |\psi_{p_z,h}^{P,SV,SH}(z)|^2\Big|_{p_z=(\omega_0/v_\infty^{p,s})\cos\theta}. \quad (64)$$

In Eq. (64), the subscript h denotes the horizontal component of the wavefunction measured in the plane of polarization. Note that Love waves as well as generalized SH eigenfunctions contribute to the horizontal kinetic energy only.

In Fig. 3, we consider the vicinity of a free surface which has been previously investigated by several authors.[2,7,8] The calculations were performed both with for-

mula (63) and with a locked-mode approximation where the medium is bounded at great depth by a rigid boundary where the displacements vanish exactly. In the latter case, the eigenvalue problem in the depth coordinate only has a discrete spectrum, which is standard. For more details on the locked-mode technique and geophysical applications, the reader is referred to Ref. 25. The outcome of the two calculations for a 3D medium are superposed in Fig. 3. In the locked-mode technique, the lower boundary is at a depth of 16 shear wavelengths. A total of 32 Love and 50 Rayleigh modes were found. The agreement is very satisfactory and confirms the validity of our approach. Our work shows that the generalized eigenfunctions of the continuum can be treated like standard normal modes, albeit with a continuous index. The results presented in Fig. 3 were first obtained in Refs. 2 and 8, using the eigenfunctions of a thick plate, the so-called Lamb modes. Our calculations illustrate the fact that the elastodynamic operator is in the limit point case at $+\infty$ according to the classification of Ref. 26. This simply means that independent of the self-adjoint boundary condition imposed at the lower boundary—Neumann, Dirichlet, or mixed—the eigenfunctions converge to a common limit as the depth of that boundary tends to ∞. A few wavelengths away from the free surface, the kinetic energy ratio oscillates around 0.5, which is the expected value in a homogeneous half-space.

We now investigate the case of a soft layer overlying a homogeneous space. In the layer, the P and S wave velocity is one-third smaller than in the half-space and the density is reduced by a factor of 2. The vertical to horizontal kinetic energy ratio is plotted as a function of depth in Fig. 4. The unit depth is the shear wavelength inside the layer and the interface between the layer and the half-space lies at about 0.2 unit depth. In the locked-mode approximation, two situations were considered where the rigid boundary is located at 5.25 and 42 wavelengths below the bottom of the layer. In the former case, 7 Love and 11 Rayleigh modes were found, while in the latter case we identified a total 56 Love and 88 Rayleigh modes. In the generalized eigenfunction expansion method, only the fundamental Love and Rayleigh modes are present. Thus, in the locked-mode method, the higher Love and Rayleigh modes serve as an approximation for the propagating P and S waves incident from below the layer. As illustrated in Fig. 4, the agreement between the two methods is extremely good and, as expected, improves as the depth of the rigid boundary increases. In the vicinity of the layer, the vertical to horizontal kinetic energy ratio shows rapid variations and, at greater depth, converges toward the ratio of a homogeneous space.

In Fig. 5, we investigate the frequency dependence of the V^2/H^2 kinetic energy ratio at the surface of a solid with a superficial soft layer, where the shear velocity and density are reduced by a factor of 2 with respect to the underlying Poisson half-space. The ratio of longitudinal to shear wavespeeds in the layer is taken to be 2.5. To facilitate the interpretation of the V^2/H^2 calculations, we show in the top panel the normalized density of states of body, Rayleigh and Love waves as a function of frequency. The frequency unit is the fundamental resonance frequency of the layer for vertically propagating shear waves, $f_0=v^s/4h$, where v^s is the shear wavespeed and h is the layer thickness. In Fig. 5, we have also indicated high- and low-frequency asymptotics of the V^2/H^2 energy ratio. At high frequency, we expect the waves to be largely insensitive to the properties of the underlying half-space. We therefore calculate an approximate high-frequency V^2/H^2 ratio by replacing the true model with a simple half-space where the P to S velocity ratio equals 2.5. The expected value is 0.513, in good agreement with the full calculations. At low frequency, we can neglect the presence of the shallow layer to recover the classical 0.56 ratio at the surface of a Poisson solid. The V^2/H^2 ratio is slightly smaller at high frequencies mainly because of the increased amount of horizontally polarized shear waves. To the contrary, the V^2/H^2 energy ratio of the Rayleigh wave tends to increase with the P to S velocity ratio. Close to the resonance frequency f_0 of the layer, the vertical to horizontal kinetic energy ratio drops dramatically. Careful analysis reveals that this is caused by the increasing contribution of the fundamental Love mode to the density of states at the surface. Around twice the resonance frequency, the V^2/H^2 ratio presents a marked overshoot which is due to an abrupt inversion of the Rayleigh wave ellipticity. At high frequencies, the energy density is largely dominated by the Rayleigh and Love waves trapped in the low-velocity layer. The fluctuations of

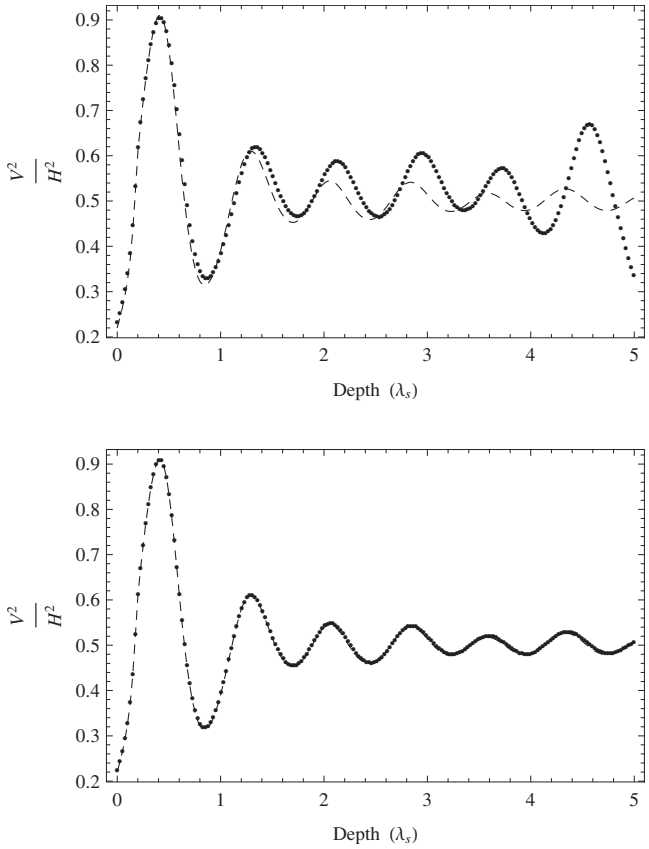


FIG. 4. Depth dependence of the vertical to horizontal kinetic energy ratio in the presence of a soft layer. Dots: locked-mode approximation. Dashed line: generalized eigenfunctions summation. The depth unit is the shear wavelength λ_S. In the locked-mode approximation, the lower boundary is located at a depth of $5.25\lambda_s$ (top) and $42\lambda_s$ (bottom), respectively.

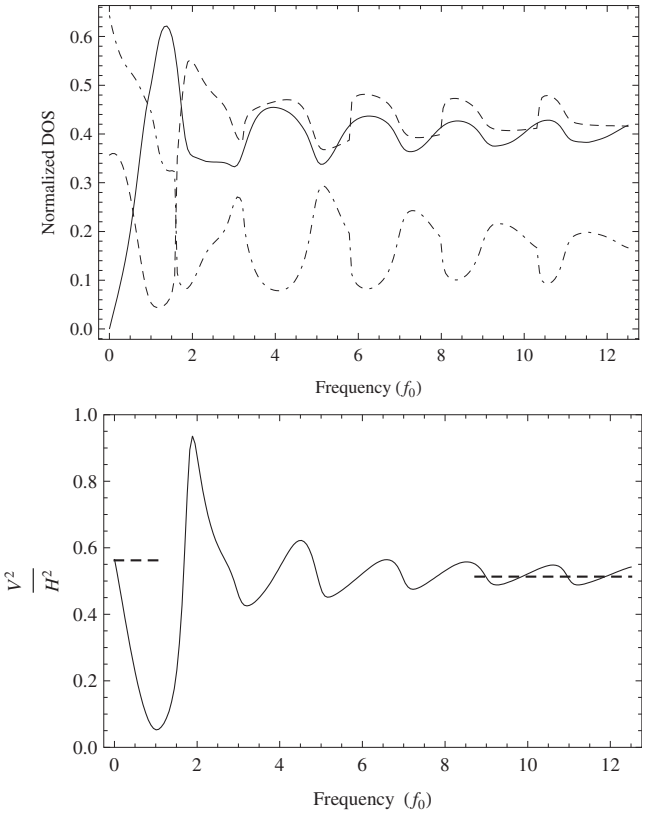


FIG. 5. Properties of a diffuse field at the surface of a half-space with a superficial low-velocity low-density layer. Top: normalized density of states of Rayleigh (dashed line), Love (solid line), and bulk waves (dash-dotted line) as a function of normalized frequency. The resonance frequency f_0 of vertically propagating S waves in the layer is the unit frequency. Bottom: frequency dependence of the vertical to horizontal kinetic energy ratio. Dashed lines: low- and high-frequency asymptotics. Solid line: generalized eigenfunctions summation.

the local density of states slowly decrease with increasing frequency.

VI. CONCLUSION

In this work, we have shown that the definition of a diffuse field as a white noise in modal space can be applied equally well to closed[1] and open systems.[27] In the framework of the spectral theory presented in this paper, discrete, continuous, or mixed spectra can be treated on the same footing. Other applications of the generalized eigenfunction expansion could be envisaged, such as the calculation of synthetic seismograms for geophysical applications.[18] The theory could also be used to give more rigorous foundations to empirical civil engineering techniques. In particular, the large drop of the vertical to horizontal kinetic energy ratio in diffuse fields close to the resonance frequency of a low-velocity layer sounds reminiscent of the so-called Nakamura's method used for site effect evaluation with ambient noise.[28] Although extremely popular, the limitations of the technique are still to be understood. The diffuse field concept offers a potentially useful tool for this purpose.

ACKNOWLEDGMENTS

I would like to express my gratitude to Professor Y. Colin de Verdière for his lectures on mathematical topics closely related to the present article. This paper has also been influenced by discussions with Dr. H. Nakahara, who suggested to examine the wave content of diffuse fields. The careful comments of an anonymous reviewer helped to improve the clarity of the manuscript. The author acknowledges the support of the Pôle Grenoblois d'Etudes et de Recherche pour la Prévention des Risques Naturels and of the French Agence Nationale de la Recherche (ANR), under grant SISDIF (JCJC08_313906) "Sismologie différentielle: caractérisation et imagerie dynamique avec les ondes diffuses."

APPENDIX: SUMMARY OF DIRAC FORMALISM

In this Appendix, we summarize the bra-ket notations used in this paper. A vector or ket of an abstract vector space will be denoted by $|e\rangle$. The representation of the vector in position space is written as $\langle x|e\rangle = e(x)$. To each ket of our space of function, we associate a bra denoted by $\langle e| = |e\rangle^\dagger$. In position space, a bra has representation $\langle e|x\rangle = e(x)^*$, where * denotes complex conjugation. Mathematically speaking, a bra would be more appropriately defined as a linear functional acting on a space of test functions. However, we will not insist on these technicalities. We shall also make use of the conjugated versions of kets such that $\langle x|e^*\rangle = e(x)^*$. The scalar product between two vectors $|e\rangle$ and $|f\rangle$ is denoted by

$$\langle e|f\rangle = \int dx \rho(x) e_i(x)^* f_i(x), \tag{A1}$$

where a summation over the repeated index i is implied. The completeness relation or resolution of the identity for a set of eigenvectors $|e_p\rangle$ is written as

$$\sum_p |e_p\rangle\langle e_p| = \mathbf{I}, \tag{A2}$$

where $\mathbf{I}$ denotes the identity in the abstract vector space, and p denotes a label running over the whole set of eigenfunctions. Equation (A2) introduces the outer product between a bra and a ket. The outer product of two (properly normalized) eigenvectors is an orthogonal projector on the subspace generated by $|e\rangle$. In the position representation, Eq. (A2) reads

$$\sum_p \langle x|e_p\rangle\langle e_p|x'\rangle = \delta_{ij}\frac{\delta(x-x')}{\rho(x')}. \tag{A3}$$

The appearance of the weighting function ρ in the denominator is consistent with the definition of the scalar product (A1). The matrix elements in position and polarization space of a general outer product between a ket $|e\rangle$ and a bra $\langle f|$ are given by

$$\langle x|e\rangle\langle f|x'\rangle = e_i(x) f_j(x')^*. \tag{A4}$$

The matrix elements of an abstract operator $\mathbf{L}$ in position space are given by the kernel of an integral operator. If $\mathbf{L}$ represents a differential operator, it will be assumed to be diagonal in position space.

[1]R. L. Weaver, "On diffuse waves in solid media," J. Acoust. Soc. Am. **71**, 1608–1609 (1982).
[2]R. Hennino, N. Trégourès, N. M. Shapiro, L. Margerin, M. Campillo, B. A. van Tiggelen, and R. L. Weaver, "Observation of equipartition of seismic waves," Phys. Rev. Lett. **86**, 3447–3450 (2001).
[3]A. E. Malcolm, J. A. Scales, and B. A. van Tiggelen, "Extracting the Green function from diffuse, equipartitioned waves," Phys. Rev. E **70**, 015601 (2004).
[4]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[5]M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[6]R. L. Weaver, "Diffuse waves in finite plates," J. Sound Vib. **94**, 319–335 (1984).
[7]R. L. Weaver, "Diffuse elastic waves at a free surface," J. Acoust. Soc. Am. **78**, 131–136 (1985).
[8]N. P. Trégourès and B. A. van Tiggelen, "Quasi-two-dimensional transfer of elastic waves," Phys. Rev. E **66**, 036601 (2002).
[9]M. Reed and B. Simon, *Scattering Theory*, Methods of Mathematical Physics Vol. **1** (Academic, New York, 1979).
[10]D. B. Pearson, *Quantum Scattering and Spectral Theory* (Academic, London, 1988).
[11]K. Aki and P. G. Richards, *Quantitative Seismology* (University Science Books, Sausalito, CA, 2002).
[12]L. M. Brekovskikh, *Waves in Layered Media* (Academic, New York, 1960).
[13]B. L. N. Kennett, *Seismic Wave Propagation in Stratified Media* (Cambridge University Press, Cambridge, 1983).
[14]Y. Dermenjian and J.-C. Guillot, "Les ondes élastiques dans un demi-espace isotrope. Développement en fonctions propres généralisées. Principe d'absorption limite" ("Elastic waves in an isotropic half-space. A generalized eigenfunction expansion and a limiting absorption principle"), C. R. Acad. Sci., Ser. I: Math. **301**, 617–619 (1985).
[15]Y. Dermenjian and J.-C. Guillot, "Scattering of elastic waves in a perturbed isotropic half space with a free boundary. The limiting absorption principle," Math. Methods Appl. Sci. **10**, 87–124 (1988).
[16]M. Reed and B. Simon, *Functional Analysis*, Methods of Mathematical Physics Vol. **1** (Academic, New York, 1980).
[17]C. Wilcox, *Sound Propagation in Stratified Fluids*, Applied Mathematical Sciences Vol. **50** (Springer-Verlag, Berlin, 1984).
[18]V. Maupin, "The radiation modes of a vertically varying half-space—A new representation of the complete Green's function in terms of modes," Geophys. J. Int. **126**, 762–780 (1996).
[19]P. Sécher, "Étude spectrale du système différentiel 2×2 associé à un problème d'élasticité linéaire," Ann. Fac. Sci. Toulouse Math. **7**, 699–726 (1998).
[20]B. L. N. Kennett, N. J. Kerry, and J. H. Woodhouse, "Symmetries in the reflection and transmission of elastic waves," Geophys. J. R. Astron. Soc. **52**, 215–229 (1978).
[21]Y. C. de Verdière, "Mathematical models for passive imaging I: General background," URL: http://www.citebase.org/abstract?id=oai:arXiv.org:math-ph/0610043 (Last viewed February, 2008).
[22]P. Sheng, Introduction to Wave Scattering, Localization and Mesoscopic Phenomena (Academic, San Diego, 1995).
[23]B. A. van Tiggelen, "Green function retrieval and time reversal in a disordered world," Phys. Rev. Lett. **91**, 243904 (2003).
[24]R. L. Weaver and O. I. Lobkis, "On the emergence of the Green's function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**, 3011–3017 (2001).
[25]G. Nolet, R. Sleeman, V. Nijhof, and B. L. N. Kennett, "Synthetic reflection seismograms in three dimensions by a locked-mode approximation," Geophysics **54**, 350–358 (1989).
[26]A. M. Krall, "$m(\lambda)$ theory for singular Hamiltonian systems with one singular point," SIAM J. Math. Anal. **20**, 664–700 (1989).
[27]R. Weaver and O. Lobkis, "Diffuse fields in open systems and the emergence of the Green's function (L)," J. Acoust. Soc. Am. **116**, 2731–2734 (2004).
[28]P.-Y. Bard, "Microtremor measurement: A tool for site effect estimation?," in *The Effects of Surface Geology on Seismic Motion*, edited by K. Irikura, K. Kudo, H. Okada, and T. Sasatami (Balkema, Rotterdam, 1999), pp. 1251–1279.

Minimizing sleep disturbance from blast noise producing training activities for residents living near a military installation

Edward T. Nykaza,[a] Larry L. Pater, and Robert H. Melton
U.S. Army Engineer Research and Development Center, Construction Engineering Research Laboratory, 2902 Newmark Drive, Champaign, Illinois 61822-1078

George A. Luz
Luz Social and Environmental Associates, 4910 Crowson Avenue, Baltimore, Maryland 21212

(Received 20 March 2008; revised 22 September 2008; accepted 22 October 2008)

Field research was conducted during 2004 in the vicinity of a United States military installation to determine if awakening of residents due to blast noise from large military weapons might vary during the night. Analysis of the data indicates that awakening from blast noise is significantly less likely during the time period between midnight and 0200 h compared to time periods before midnight and approaching dawn. These findings suggest that postponing noisy evening training until after midnight could effectively reduce the negative impact of nighttime training on local residents and thus help to preserve nighttime training capabilities.
 [DOI: 10.1121/1.3026325]

PACS number(s): 43.50.Qp, 43.28.Mw, 43.50.Pn, 43.50.Sr [BSF] Pages: 175–184

I. INTRODUCTION

The United States military conducts training with heavy weapons (e.g., artillery, tank guns, and other explosive weapons) at installations that had been located in unpopulated areas when they were originally established in the late 19th and early 20th centuries. Although some are still in sparsely populated areas, others are in areas where residential populations are increasing.[1] In an effort to limit incompatible land use on private lands near heavy weapon ranges, the U.S. Department of Defense publishes maps of the *C*-weighted day-night average sound level (*C*-weighted DNL) for use by those local governments having regulatory authority to control residential land use. The calculation method for generating these maps follows along the lines established by the U.S. Environmental Protection Agency in 1974.[2] The calculation method of DNL was applied to blast noise by the Committee on Hearing, Bioacoustics and Biomechanics (CHABA) in 1981,[3] and became official army policy as described in Chap. 7 of Army Regulation 200–1.

In the years since the U.S. Army first adopted the CHABA methodology, the methods for predicting acoustical signatures and variability in propagation of weapon noise have undergone continuous improvement, but the psychophysical rules for translating the physical stimuli into estimates of community annoyance are still in need of improvement.

Particularly elusive are the rules for analyzing the statistical distribution of impulsive noise levels received at specific locations in the community. There is considerable evidence that the highest intensity blasts in the distribution contribute more to the cumulative annoyance than the less intense blasts.[4–6] Consequently, maps of the long term (e.g., annual) *C*-weighted DNL can obscure the relationship between the physical stimuli and the psychological response, community annoyance.[7] This conclusion in regard to annoyance echoes an earlier observation about complaints—that complaints about weapon noise are generated by short term increases above the long term average.[8]

Statistical variability in the propagation of heavy weapon noise is particularly important in the evening and at night. Not only are more people at home during these hours[9] but also the prevailing meteorological conditions at these times often cause low-frequency weapon blasts to propagate farther.[10] As reported by Rylander and Lundquist,[11] when Swedish citizens living near heavy weapon ranges were asked about the time of day when they were most annoyed, about half chose the evening and about a third chose the night. About 10% stated that the shooting made it difficult for them to fall asleep and about the same percentage stated that the shooting resulted in awakening.

An increasing number of installation range control offices have access to blast noise monitors and/or sound level prediction software, which provide detailed information regarding sound levels in residential communities. Literature suggests that levels below a *C*-weighted sound exposure level (SEL) of 90 dB or a flat-weighted peak level of 115 dB will contribute little to annoyance[11] and are unlikely to generate complaints,[12] but these guidelines are of limited use for assessing impacts on sleep. The threshold at which people are likely to be awakened by weapon noise, and how this threshold varies across the night, would provide enhanced decision and planning guidance.

Some information is available in literature. Successive artillery, tank gun, and mortar blasts are generally separated by several minutes, and, in this regard, sleep disturbance from heavy weapons would be expected to be more similar to disturbance from sonic booms than from aircraft flyovers.

[a]Electronic mail: edward.t.nykaza@usace.army.mil

Differences between awakening from a high intensity impulse and from a subsonic aircraft flyover were noted by Lukas and Dobbs 35 years ago.[13] Working with female subjects, they found clear differences in brain response between subsonic jet noise and simulated sonic booms. Sonic booms were more likely to leave the electroencephalogram (EEG) unchanged and aircraft noises were more likely to cause awakenings. The awakening threshold for sonic booms was approximately a flat-weighted peak of 124 dB outdoors. 71% of the sleeping subjects showed no EEG changes at this level and only 1.5% awoke enough to push a button. Working with tape recordings of 120 mm tank gun noise, Griefahn[14] found sleep motility falling back to preshot baseline within a minute. In contrast, for aircraft, Fidell *et al.*[15] found that "autocorrelation analyses demonstrated that episodes of sleep motility persist for periods of minutes, rather than seconds." These researchers found that the correlation between probability of awakening and individual noise event *A*-weighted maximum levels increased when the postevent window was stretched from 1 to 2–5 min.

An observed similarity between sleep disturbance to blast and aircraft noise has been the change in awakening as a function of sleep duration. For aircraft, Fidell *et al.*[16] found time since retiring to be positively related to awakening: The longer the participant had slept, the greater the likelihood of awakening to the aircraft event. Griefahn[14] also found the probability of awakening to tank shots increasing the longer the subject had been asleep, and she warned against firing in the early morning. In contrast, Vos and Houben,[17] working with digitally reproduced gun shots, volleys of gun shots, and airplane flyovers, found no difference in awakening between the first and second halves of sleep time. Vos and Houben[17] also found that a single shot was less likely to awaken subjects than a volley of shots. For aircraft and volleys of shots, Vos and Houben[17] observed some habituation over sessions, but for single shots they found none.

Although these laboratory studies are important for understanding the processes underlying awakening to heavy weapon noise, they all have one shortcoming for determining a realistic threshold for sleep disturbance in the field. None of the stimuli match the acoustic power in the lower 1/3 octave bands within the first second of the firing of a large gun. During an equipment shakedown test conducted in preparation for the field study reported here, the research team used a 50 000 W speaker system to attempt to simulate the level and spectrum of a 120 mm tank gun blast recorded at a distance of 2 km.[18] Comparison of the original and reproduced spectra showed that there were no significant differences between the midrange 1/3 octave bands, namely, the frequencies that are unaltered by the *A*-weighting filter, but the lowest bands in the original acoustic signature were under-represented. The low-frequency bands in the range 12–35 Hz are unique to large guns, and the only way to truly define their potential for disturbing sleep is to conduct a field study with people who have lived long enough in the vicinity of heavy weapon ranges to be habituated to the noise. Such people are the subject of the following study. Of particular interest is whether their thresholds of awakening are unchanged during the night, systematically decreasing during the night, or exhibit a maximum or minimum value at some time during the night.

II. METHODS

A. Subject selection

Residents living near a military installation were recruited for participation in the study by mail.[19] Approximately 600 letters were sent out and 148 residents responded. Potential subjects were screened according to the following criteria: (1) Must not be deaf (although no audiograms were given), (2) must not have a sleep disorder, (3) must not be pregnant, (4) must be between the ages of 18 and 75, (5) must sleep at home for the duration of the study, (6) typically sleep between the hours of 2200 and 0800, (7) must be a resident of the area for at least one year, and (8) must be willing to have equipment set up in their bedroom and outside their home.

Due to the limited number of responses, limited equipment available for the study, and relatively small blast noise impact area, subjects that met the screening criteria were chosen on the basis of proximity to the nearest firing point. Although this precluded achievement of a strictly random sample, the sample was representative of the census tract from which it was drawn.

B. Sound measurement

Ambient and blast event sound levels were documented indoors and outdoors at each subject's home. Each noise monitor recorded 1 s SELs, 1/3 octave band SEL, and peak pressure levels between 2000 and 0800 h. Each measurement was also *A*-, *C*-, and flat-weighted.

All outdoor noise measurements were made with Norsonic-121 (Ref. 20) noise monitors. The outdoor microphones were located at least 3 m from the subjects' windows and any other reflective surfaces. There were a few instances where neighbors with adjacent homes participated in the study. In these situations one outdoor microphone was used as the outdoor measurements for both subjects. Indoor noise measurements were made in each subject's bedroom with either the second channel of a Norsonic-121 noise monitor or with Larson Davis-870 (Ref. 20) noise monitors. The frequency range of the Norsonic units was 6.3 Hz–5 kHz and that of the Larson Davis units was 5 Hz–16 kHz. Each microphone was located 2 m from each wall (if possible). These bandwidths are adequate; SEL spectrum for a typical blast event is shown in Fig. 1.

C. Sleep monitoring

Each subject wore a commercially available sleep actimeter, the Mini Mitter Actiwatch® (Ref. 20) each night of the study. This actimeter is an integrating actimeter; the sensor integrates the degree and speed of motion and produces an electrical current that varies in magnitude. The increase in degree of speed and motion produces an increase in voltage, which is stored as an activity count (AC). The maximum sampling rate of these actimeters was 32 Hz. The accuracy

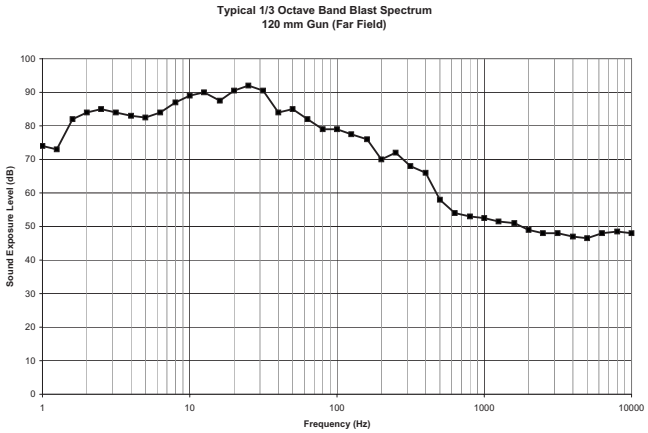


FIG. 1. Typical 1/3 octave band blast spectrum from a 120 mm gun as measured in the far field.

and reliability of the actimeter data have been documented in Refs. 21 and 22, and in a pilot study[23] that preceded this effort. The actimeter was also equipped with a marker switch that was inset on the device's front panel. The marker switch provided tactile and audible feedback, a "click" sound, when pressed that informed the subject of a successful marking. Subjects were instructed to press the inset button on the actimeter when they first got into bed for the night, when they got out of bed in the morning, and when they woke up during the night for any reason. At the time of the equipment installation, instructions and demonstrations were given to each subject on how to use the sleep actimeter. Subjects were also given a questionnaire to complete each morning that contained questions regarding their previous day and night of sleep, and were left a paper copy of the instructions that included a 24 h toll-free number they could call if they had any problems or questions.

D. Time drift corrections

The internal clocks of both the sleep and noise monitors used during the study had a time drift that was corrected before the data were analyzed.

1. Noise monitors

The noise monitors were manually calibrated to the "correct" time obtained from the WWVB broadcast out of Fort Collins, CO via an atomic watch prior to their use. After data were collected the time drift was calculated by comparing the correct time on an atomic watch to the time on the monitor at the end of their use. The time drift correction was done under the assumption that the time drift was linear. This assumption was reasonable given that the time drift measured in the field was relatively small and the time drift measured in the laboratory after the study was linear. The monitors drifted no more than 1 min over the course of 11 days.

2. Sleep monitors

No drift in the actimeters had been observed during the pilot study because the instruments had been read and reset within 12 h of beginning data collection. For longer periods, drift had to be taken into consideration. The internal clocks of the actimeters were also calibrated to the correct time via atomic watch prior to their use; however, these actimeters did not allow for direct comparison of the correct time to the internal clock time at the end of a usage period. A series of poststudy experiments was conducted to determine the amount of time drift for each actimeter. The actimeters were tested at temperatures reflective of the temperatures observed during the study. The results of the poststudy experiment found that each actimeter had a unique time drift and that the dependence on temperature was minimal. The measured time drift for each actimeter was used to correct all data.

E. Operational definitions

1. Noise events

The outdoor monitors were set up to record noise events. A noise event was defined as a noise level that exceeded a *C*-weighted peak threshold level of 100 dB (with regard to 20 μPa). For each noise event, noise stimulus metric levels were recorded during a time period from 1 s before the event to 5 s after the event, and an audio recording of the event was made for later reference and source identification. The audio recording or wave file was recorded only on the outdoor microphone to protect the subject's privacy.

2. Response evaluation period

For each blast event that a subject experienced, it was determined whether the subject was already awake when the blast occurred, if the subject woke up from the blast, or if the subject slept through the blast event. In order to determine the subject response to each blast event, an appropriate evaluation period had to be chosen. The evaluation period was defined as the length of the time window, before and after the blast event, within which the subjects' sleep state was determined.

The vegetative arousal to a blast is complete within a cycle of 30 s. It begins with a sympathetic outflow to the peripheral muscles at about 2 s followed by heart acceleration at 3–3.5 s and an increase in blood pressure at 5–6 s. The magnitude of the pulse measured at the finger reaches a minimum around 8 s and awakening begins around 10 s.[14] Thus, any instrument for measuring awakening to blasts must be sensitive to movement within the first 30 s. In this case, the minimum period available in the actimeter chosen in the pilot study was 15 s and the evaluation period chosen was 60 s.

3. Button press and activity count awakenings

Sleep awakenings were defined in two ways: Awakenings determined from button press (BP) and awakenings calculated from the AC. These two definitions of awakenings are defined as BP awakenings and AC awakenings throughout the text. A BP awakening refers to a subject pressing the button on the actimeter after waking from sleep and an AC awakening is based on the amount of motion calculated by an algorithm provided by the actimeter company.

As previously mentioned the actimeter used in this study was an integrating actimeter, i.e., the degree of motion reported for each time bin was calculated by integrating the

amount of motion from the start of the time bin to the end of the time bin. The algorithm used to calculate whether the subject was asleep or awake for each time bin is taken from a table in the user's manual[24] and is written as

$$\mathrm{AC}(\beta_0) = 4\beta_0 + 0.2(\beta_{-4} + \beta_{-3} + \beta_{-2} + \beta_{-1} + \beta_1 + \beta_2 + \beta_3 + \beta_4) + 0.04(\beta_{-8} + \beta_{-7} + \beta_{-6} + \beta_{-5} + \beta_5 + \beta_6 + \beta_7 + \beta_8), \quad (1)$$

where $\mathrm{AC}(\beta_0)$ is the activity count for the time bin of interest, β_0 is the time bin of interest, and β_n is the time bin relative to the time bin of interest.

The minimum available time bin length of 15 s was used for this study. Therefore, the algorithm used to determine sleep state was based on the weighted sum of AC that occurred within 2 min before and after a stimulus event. The sleep state is calculated according to the following criteria as suggested by the manufacturer and as used in the pilot study:[23]

$$\text{Sleep} = \mathrm{AC}(\beta_0) \leq 40$$

$$\text{Wake} = \mathrm{AC}(\beta_0) > 40$$

It should be noted that manufacturer's recommendation of the $\mathrm{AC}(\beta_0) > 40$ as threshold of awakening was based on a comparison of different criteria published by Kushida *et al.* in 2001.[22] In their comparison paper they tested low, medium, and high AC thresholds: 20, 40, and 80, respectively. They found that "the number of awakenings as calculated by the medium-threshold and high-threshold actigraphic algorithms was in close agreement and not significantly different from that obtained by polysomnography."

F. Data analysis

The data were analyzed in two ways to look at the effect of blast noise on overall sleep quality and to determine if there are preferred time periods at night to conduct nighttime training. Using terms from earlier aircraft disturbance studies,[25] the analyses are referred to as "whole night" and "behavioral awakening response." All regression analyses were performed using SAS® (Ref. 20) version 8.02. Significances for the regression analyses in this report are defined at the alpha <0.05 level.

1. Whole night analyses

To analyze the effects of blast noise on overall sleep quality over the entire night, a windowed subset of data was generated from the blast data set by restricting the data to the times between the time when a subject went to bed and the time when the subject arose out of bed in the morning. The dependent variables analyzed were mean sleep bout time, mean wake bout time, total time in bed, proportion of the night awake, count of button pushes, bedtime, and the response to the morning question regarding how well the subject slept the previous night. Only nights that included at least one blast were included in the analyses. Each analysis also included a baseline covariate that was calculated as the median of the mean values of the dependent variable on nights when no blasts were recorded. Other covariates used were the recorded bedtime for that night measured as hours after 2000, subject age, sex, and the number of resident years. Separate analyses were run for each dependent variable using each independent variable. The following independent variables were used: total of number of blast events, responses to several of the morning questions, and the mean value for the specified acoustic metric. Acoustic metrics included outdoor *A*-, *C*-, and flat-weighted peak and SELs.

General linear model analyses were performed using the SAS® GLM procedure. No transformations were applied since plots of residual versus predicted values were reasonably homoscedastic in all cases. Logistic regressions were performed using the SAS® LOGISTIC procedure, with input in the "trials/responses" format. Correction for overdispersion was achieved by including the "SCALE=WILLIAMS" option in the model statement. Negative binomial regressions were performed using the SAS® GENMOD procedure with options "DIST=NEGBIN" and "OFFSET=LNSLPDUR," where LNSLPDUR was the natural logarithm of the sleep duration, which was calculated as the difference between final awake time and get in bed time. Use of total time in bed as an offset variable served to correct the count of BPs for subject-night variability of the time in bed.

2. Behavioral awakening response analyses

Behavioral awakening response logistic regression analyses were conducted to understand the functional relationship between stimulus noise metrics and awakenings. In each analysis, data points in which the subject was determined to be awake within the designated time window before the blast event were excluded from the analysis. In essence, these analyses looked at the effect of each blast event experienced by subjects when they were determined to be asleep.

To determine if there are preferred time periods to conduct nighttime training, the data set for the entire study was divided into five time categories (2100–2300, 2300–0000, 0000–0100, 0100–0200, and 0200–0400). Divisions of the time categories were made with the goal of having an equal number of data points in each time period.

Each analysis was run separately for each combination of the following: AC and BP awakenings, outdoor and indoor *A*-, *C*-, and flat-weighted peak sound pressure levels and SELs. Two dependent variables, each a binary variable (0 =asleep after blast and 1=awake after blast), were examined.

As a control for individual variation in propensity to wake up, two variables used as covariates in the whole night analyses were used again as covariates in these analyses: the median percent of the time awake on nights with no blast events determined by AC awakenings, and the median percent of time awake on nights with no blast events determined by BP awakenings. Logistic regressions were performed using the SAS® LOGISTIC procedure. The option "SCALE =PEARSON" was included in the model statement to implement Pearson's correction for overdispersion. Williams' method was not used to model overdispersion, as it was in

TABLE I. Demographics of 33 study subjects.

Parameter	Mean	Standard deviation	Min	Max
Age (years)	44.6	15.4	19	72
Distance from firing point (km)	5.4	2.9	1.8	8.9
Years in the neighborhood (years)	10.9	8.8	1	29
Go to bed time (hh:mm)	22:31	01:13	19:12	02:45

the whole night analyses, since this option could be used only with the events/trials syntax.

III. RESULTS

A. Data collection

1. Subject demographics

The final subject pool consisted of 33 subjects and the average length each subject participated was 21 nights. The subject population was diverse (Table I) and the average age and percentage of male/female participants were representative of census data for the zip code that the study population was drawn from (Table II). The census data are from year 2000.

2. Stimulus

The noise events that served as stimuli for this research were from 120 mm tank gun fire from a predeployment training exercise that lasted approximately 6 weeks spanning 21 September–31 October, 2004. The actual number of night firing days over this time period was 15, which did include firing over the weekend. On the average, 40 shots were fired per night. Night firing typically started around 2200 and ended just before 0400.

There were 3101 received events that exceeded the threshold of 100 dB, *C*-weighted peak. A wave file was recorded for each of these events so that the event could be audited to confirm it was a blast. Out of the total sample, there were 2845 blasts and 256 nonblast events. The most common nonblast events included wind, rain, and thunder. A detailed analysis of the effects of nonblast events on sleep was not conducted given the relatively small number of nonblast events. Figure 2 shows the distribution of outdoor flat-weighted peak sound pressure levels for blasts that were recorded while subjects were sleeping in bed. The distribution is approximately normal and illustrates the large variance in received blast noise level that can occur even over a short time span. Variations in received blast noise levels due to meteorological conditions such as wind and temperature structure have been shown to range by as much as 50 dB.[26]

TABLE II. Comparison of subject demographics with census data.

Age/gender parameter	Subjects	Census data for zip code of study population (population 18–75 years of age)
Average age (years)	44.6	43.4
Percent male	36%	50%
Percent female	64%	50%

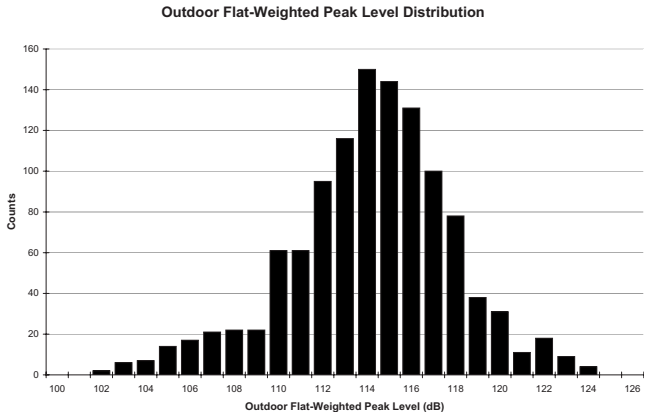


FIG. 2. Distribution of outdoor flat-weighted peak blast event levels.

3. Temporal structure of button press awakenings

Table III shows the percentages of all BP awakenings occurring within 1 min before and 1 min after a blast summed over all subject/nights. Presumably, the slight increase above the approximate baseline of 1.44% in the 15 s period before the events reflects small errors in synchronizing between the sound measurement equipment and the actimeters. As previously noted the total cycle for the vegetative arousal to blasts is complete within 30 s. Table III reflects this phenomenon and confirms the appropriateness of the preblast and postblast evaluation periods used in this study. Determining the subjects' sleep state during the preblast evaluation period showed that subjects were already awake for approximately 3% of the blasts.

4. Morning questionnaires

As part of the morning questionnaire, subjects were asked to write down the reasons they awoke the previous night along with whether they pressed the actimeter button, the amount of time it took them to fall back asleep after each awakening, and their annoyance on a five-point scale. Subjects reported being more annoyed when they woke up from a blast event in comparison to other reasons, as shown in Table IV.

B. Whole night analyses

The results of the whole night analyses are shown in Table V for analyses in which the effect of the explanatory

TABLE III. Subjects' BP responses relative to blast events.

Time relative to the blast event [τ(s)]	Percentage of button presses
$-60 \leq \tau < -45$	1.44
$-45 \leq \tau < -30$	0.86
$-30 \leq \tau < -15$	1.44
$-15 \leq \tau < 0$	3.46
$0 \leq \tau < 15$	8.36
$15 \leq \tau < 30$	14.70
$30 \leq \tau < 45$	5.76
$45 \leq \tau < 60$	3.17
$60 \leq \tau < 75$	1.44

TABLE IV. Comparison of reason for waking and annoyance.

Reason for waking	Percentage of responses	Percent not very annoyed	Percent slightly annoyed	Percent moderately annoyed	Percent very annoyed	Percent extremely annoyed
Bathroom	51	69	22	6	2.5	0.3
Do not know	16	68	21	8.5	2.5	0
Blast noise	10	10	43	25	12.5	10

variable was significant, controlling for the covariate effects. Salient features of these results were as follows.

(1) The covariates' bedtime, age, sex, and resident years had little or no effect on sleep bout duration or probability of awakening.
(2) As the number of blast events increased during a night, the number of BP awakenings also increased ($p=0.0075$). A covariate was used during this analysis to account for individual propensities of awakening, which was based on the median number of awakenings on nights that no blast events occurred. That is, the increase in awakenings was statistically adjusted for the number of awakenings that typically occurred for that individual on nights when there was no blast activity.
(3) As the number of blast events increased during a night, the mean sleep bout time decreased ($p=0.0365$). A covariate was used during this analysis to account for typical sleep bout lengths, which was based on the median mean sleep bout time on nights that no blast events occurred. This finding is not profound since the more times people are awakened during their normal sleeping hours, the shorter their sleeping bouts. The statistic is reported here to confirm the orderliness of the data.
(4) As the reported openness of subjects' windows increased, the number of AC awakenings increased ($p=0.0337$). The median percent of time awake on nights with no blast events was used as a covariate to account for individual propensities of awakening.

TABLE V. Whole night analyses with significant effects of the predictor variable of primary interest.

Sleep bout duration[a]					
Effect	DF	Estimate	Standard error	Student's t	p-value
Intercept	1	97.3	81.9	1.2	0.238
Median mean sleep bout time on nights with no blasts	1	0.8	0.1	6.5	<0.001
Bedtime	1	−2.1	16.1	−0.1	0.895
Age	1	0.2	1.1	0.2	0.857
Sex	1	29.5	48.0	0.6	0.540
No. of resident years	1	0.6	1.7	0.3	0.742
Total No. of blast events	1	−1.9	0.9	−2.1	0.037

Button press (BP) awakenings[b]					
Effect	DF	Estimate	Standard error	X^2	p-value
Intercept	1	−10.99	0.432	646.1	<0.001
Median BP hourly rate on nights with no blasts	1	2.89	0.488	35.0	<0.001
Bedtime	1	−0.03	0.090	0.1	0.777
Age	1	0.003	0.006	0.2	0.664
Sex	1	0.374	0.183	4.2	0.041
No. of resident years	1	0.012	0.009	1.6	0.205
Total No. of blast events	1	0.013	0.005	7.2	0.008
Dispersion	1	0.224	0.066	⋯	⋯

Activity count (AC) awakenings[c]					
Effect	DF	Estimate	Standard error	Wald X^2	p-value
Intercept	1	−2.543	0.310	67.1	<0.001
Median percent of time awake on nights with no blasts	1	0.079	0.010	65.3	<0.001
Bedtime	1	0.030	0.048	0.4	0.540
Age	1	0.001	0.003	0.2	0.651
Sex	1	−0.105	0.090	1.4	0.243
No. of resident years	1	−0.001	0.005	0.1	0.827
Reported openness of window	1	−0.141	0.066	4.5	0.034

[a]Analyzed using general linear model.
[b]Analyzed using negative binomial regression, with ln(sleep duration) used as an offset variable.
[c]Analyzed using logistic regression analysis. "Dispersion" refers to William's correction for overdispersion.

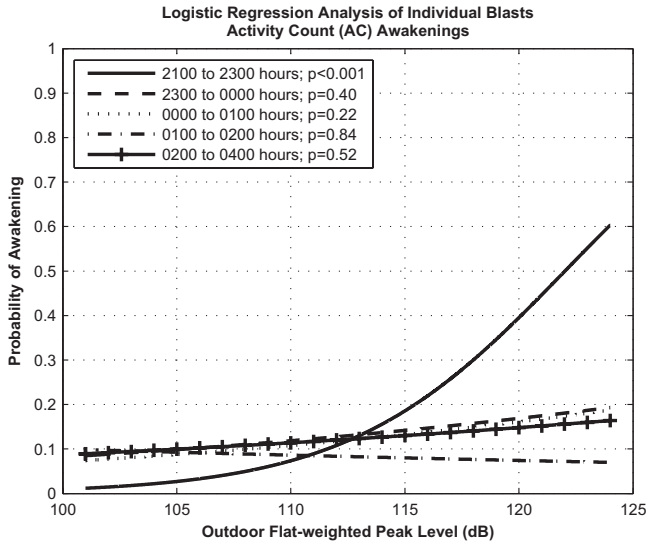


FIG. 3. Logistic regression analysis of the probability of AC awakening from outdoor flat-weighted peak levels for various time periods.

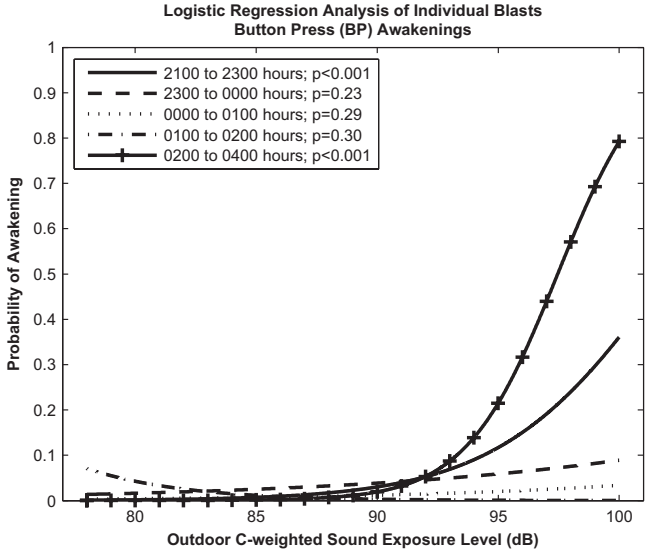


FIG. 5. Logistic regression analysis of the probability of BP awakening from outdoor *C*-weighted SELs for various time periods.

C. Behavioral awakening response analyses

The BP and AC awakening analyses were run for all combinations of the three weighting scales (*A*, *C*, and flat), the two ways of quantifying a blast (peak and SEL), the two locations (indoor and outdoor), and the five time periods. Given the large number of dose-response curves from these analyses only the outdoor blast noise metrics are presented in this paper. The entire set of dose-response curves may be found in an ERDC technical report.[19]

The behavioral analyses show that awakenings strongly depended on blast noise event level in the beginning of the night or more specifically in the time period 2100–2300; as the blast noise level increased during this time period, the probability of awakening increased. These results were significant for both definitions of awakenings for the following stimulus metrics: outdoor flat-weighted peak and SEL (Figs. 3 and 4), outdoor *C*-weighted SEL (Fig. 5), and indoor flat-weighted peak and SEL. For outdoor *C*-weighted peak (Fig. 6), BP awakenings reached statistical significance but the AC awakenings did not. Only the outdoor *A*-weighting failed as a level-dependent predictor of behavioral awakening during these critical hours (Figs. 7 and 8).

Conversely, awakening during the middle of the night (0000–0200) was unrelated to sound level. Even more significant is the fact that a given blast level produced a smaller percentage of awakenings after 2300 h. It is also significant that higher noise levels, 100–104 dB flat-weighted SEL and flat-weighted peak levels greater than 115 dB, which are judged to carry a medium to high risk of noise complaints during the day,[12] caused only a very small percentage of the subjects to register a BP (Figs. 4 and 5) during this time period. No noise complaints were received during this study and the public was not informed that a sleep disturbance study was in progress.

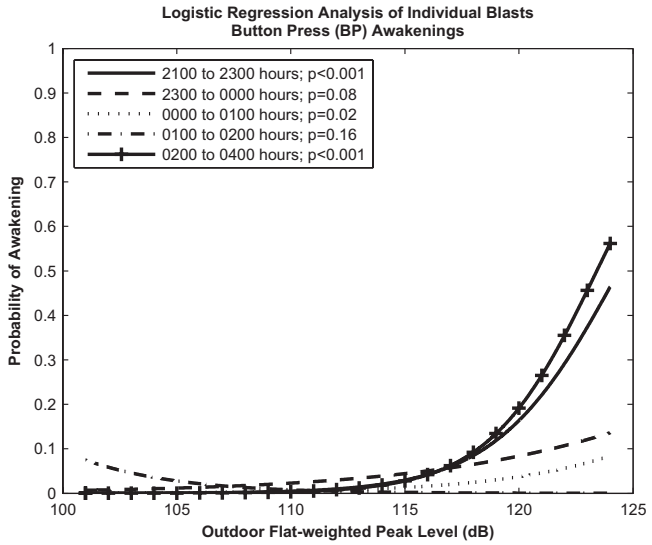


FIG. 4. Logistic regression analysis of the probability of BP awakening from outdoor flat-weighted peak levels for various time periods.

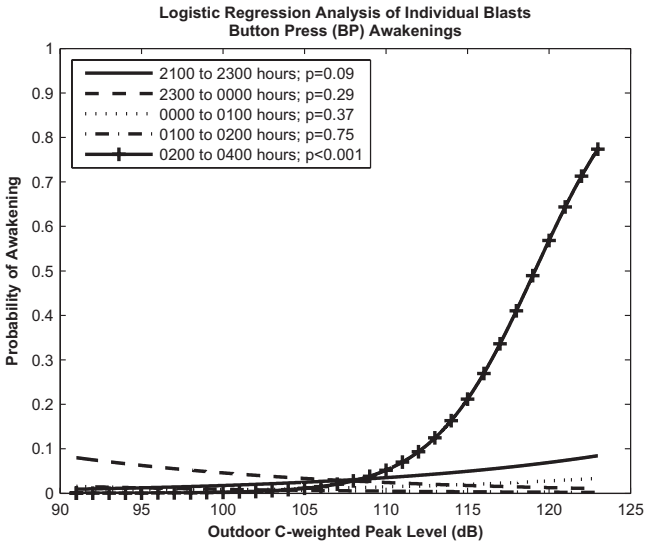


FIG. 6. Logistic regression analysis of the probability of BP awakening from outdoor *C*-weighted peak levels for various time periods.

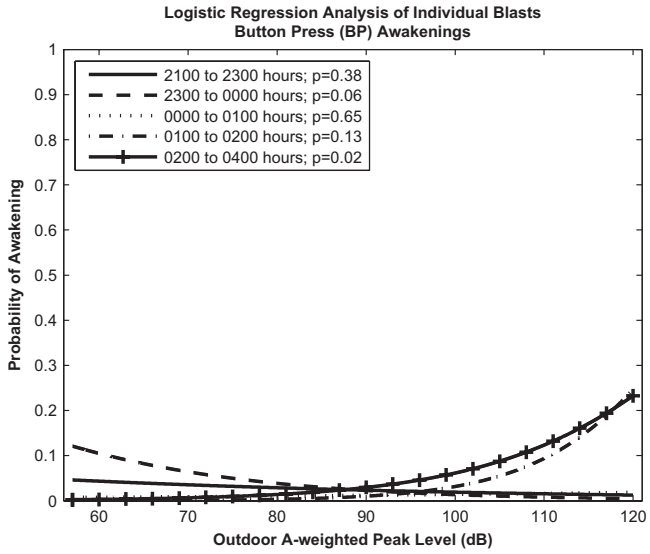


FIG. 7. Logistic regression analysis of the probability of BP awakening from outdoor *A*-weighted peak levels for various time periods.

The hours between 2300 and 0000 were a transition zone between the earlier level-dependent and the later level-independent periods. BP awakenings during the hours 2300–0000 were slightly dependent on noise event level, but to a lesser extent than 2100–2300. This correlation was significant only for the outdoor and indoor flat-weighted SEL metrics.

BP awakenings during the hours 0200–0400 were also level-dependent. However, the metrics yielding the best predictions were different from those found significant for the beginning of the night. The outdoor *C*-weighted peak and SEL yielded the best predictions during the time period 0200–0400, while the flat-weighted metrics were the best during the beginning of the night (2100–2300). In order of statistical significance, the best predictors of BP awakenings during the time period 0200–0400 were outdoor *C*-weighted peak (Fig. 6) and SEL (Fig. 5), outdoor flat-weighted peak level (Fig. 3), and outdoor *A*-weighted peak (Fig. 7) and SEL (Fig. 8).

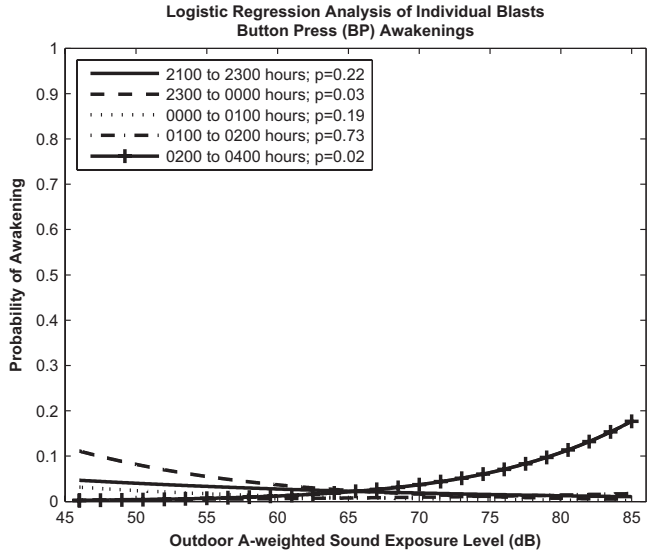


FIG. 8. Logistic regression analysis of the probability of BP awakening from outdoor *A*-weighted SELs for various time periods.

IV. DISCUSSION

A. Limitations due to sample size

The sample sizes available for analysis did not permit inclusion of more than one independent variable and one covariate. Separate analyses were run for each combination of independent variable and associated covariate, and only one-quarter of the 2845 recorded blasts events during the six-week study exceeded the putative flat-weighted peak "awakening threshold" of 115 dB. Nevertheless, the bulk of the analyses supports the conclusion that sleepers were least sensitive to blast noise during the middle part of the night (0000–0200).

B. Outdoor and indoor stimulus metrics

Military noise policy and guidance is primarily based on outdoor measurements because it is not feasible to accurately predict indoor levels. It is difficult to reliably predict the amount of attenuation each individual home provides to its residents. Even if it were possible, structural vibrations might be a better predictor of annoyance and sleep disturbance than indoor sound pressure level. In general, the findings of this study did not uncover a good correlation between indoor stimulus metrics and awakenings. In order to accurately assess the relationship between indoor stimulus metrics and awakenings, ambient indoor noise levels, the level above the ambient noise level, and secondary noises caused by structural vibrations or bric-a-brac noise must be accurately measured. However, outdoor blast noise levels seem to reliably predict sleep disturbance.

C. Stimulus metric weighting filters

During the beginning part of the night (2100–2300), the stimulus metrics that best predict the probability of awakening are the flat-weighted stimulus metrics; whereas the outdoor *C*-weighted metrics best predict the probability of awakenings during the later part of the night (0200–0400). These results could mean that different spectral frequencies are responsible for awakenings at different times of the night, or they could be an artifact of the relatively sparse BP awakening data set. For example, the findings for the beginning part of the night were significant for both AC and BP awakenings but the findings for that later part of the night (0200–0400) were only significant for the analyses that used BP awakenings.

D. Comparison of button press and activity count awakenings

When comparing the results from the two definitions of awakening, it can be observed that AC awakenings are a more sensitive measure of awakening than BP awakenings. That is, there were more AC awakenings than BP awaken-

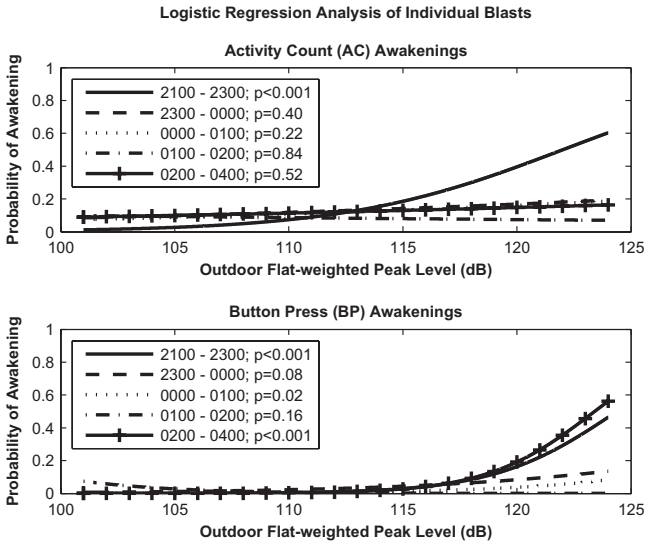


FIG. 9. Logistic regression analysis of the probability of AC and BP awakenings from outdoor flat-weighted peak level for various time periods.

ings, as shown in Fig. 9, the curves for both AC and BP awakenings are similar and suggest that both methods were meaningfully measuring the same effect.

Choice of criteria for awakening is an important consideration in sleep studies and both methods used in this study have their weaknesses. A BP awakening reflects the conscious waking state of the subject. However, the method is dependent on whether the sleeping person remembers to press the button when awakened, which in turn is dependent on whether the person is consciously awake. Another confounding factor is whether the task of pressing the button upon waking influences the subject's normal sleeping habits. Pressing the button could cause the subject to reach a state of consciousness that would not normally be reached, and may make it harder for the person to fall back asleep.

The AC method of awakening is a less precisely interpretable method because it is never fully known whether the person was actually awake or just happened to move during sleep. On the other hand, the actimeter AC method of awakening is less invasive and has no dependence on human consciousness, measuring the degree to which sleep was restful. Sleep research conducted solely using the AC awakening method reduces the dependence of the results on a subjects' conscious self-assessment.

E. Findings in comparison to the pilot study

The field study was preceded by a pilot laboratory study[23,18] designed to test the reliability of the instrumentation and confirmed that the probability of awakening to blast noise varies during the course of the night. The pilot study also provided the justification for limiting the field study to the relatively nonintrusive wrist-worn actimeter and the choice of actimeter. The pilot study, which involved 15 subjects, found that blast noise presented at a peak level of 120 dB was approximately 1.5 times more likely to wake someone than blast noise at 110 dB, regardless of the way the response to the blast noise was measured. It was also found that time had a slight effect on awakening. Subjects were more likely to respond to the stimulus between late evening (2300–0000) than the middle of the night (0000–0400), which agrees with the findings of this study.

TABLE VI. Complaint risk criteria.

Risk of noise complaints	Large caliber weapon noise [unweighted peak SPL (dB)]
Low	<115
Medium	115–130
High	130–140
Risk of physiological damage to unprotected human ears and structural damage claims	>140

F. Findings in comparison to complaint risk criteria

The two most common criteria used to measure human response to blast noise are complaints and annoyance. At night, another useful way to measure community response to blast noise is through sleep awakenings and disturbances. During the day, Pater's[12] complaint risk criteria (Table VI) are often used at the U.S. Department of Defense installations to judge the risk of receiving noise complaints. During the beginning and later part of the night (2100–2300) and (0200–0400), the threshold values of the probability of awakening and complaint risk criteria are very similar as shown in Fig. 9. When the level of an outdoor blast event is less than a flat-weighted peak level of 115 dB, there is a low probability of awakening and a low risk of complaints. As the decibel level increases above 115 dB, the probability of awakening and complaint risk increases.

V. CONCLUSIONS

A. Blast noise negatively affects local residents

Nighttime blast noise can affect residents while sleeping or trying to fall asleep. The whole night logistic regression analyses found that as the number of blast events increased during a night, the number of BP awakenings increased. The behavioral awakening response analyses also found that the time of the blast event significantly affected subject awakenings. In the beginning of the night (2100–2300) blast noise level had an effect on the probability of awakening, whereas the dependence on level was not apparent in the middle part of the night (0000–0200). For all stimulus metric levels measured, the probability of sleep disturbance was smaller in the middle of the night compared to the evening hours before midnight.

B. Current nighttime training restrictions and recommendations

Military installations clearly recognize that both day and night trainings are required to enable realistic rehearsal with all weapons to ensure combat proficiency and to minimize loss of life. However, to preserve training capability and to be a good neighbor, installations also endeavor to minimize community noise disturbance and resulting negative public reaction. To this end, installations self-impose firing curfews

that typically stipulate that noise training should be completed before midnight. The results of this research project clearly and strongly indicate that community disturbance is more effectively reduced by conducting training between 0000 and 0200 h, and avoiding noisy training during the evening hours before midnight.

C. Recommendations for future work

The results of this study suggest that postponing nighttime firing until after midnight, or at least after 2300, when the majority of residents have already fallen asleep, could potentially minimize the negative impact of blast noise on local residents. These findings are statistically significant for the present study population, but should be investigated further before accepting their validity as general. This study demonstrated the practicality and importance of conducting this type of field research with actual military blast stimuli and with subjects sleeping in their own beds. One of the issues with this study was the sparsity of data. It is therefore recommended that a future larger scale effort should include a random sample of several communities exposed to a large number of nighttime blast noise events over a longer period of time to determine to what extent the findings from this study can be applied to all communities surrounding installations. Future work should explore the effect of ambient noise level, the level of noise events above the ambient noise level, and 1/3 octave band level on awakening for various time periods. Time periods should also include the end of the night hours (0400–0600); no blast events occurred during this time period in this study.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their comments and the acoustics teams at ERDC and USACHPPM for their support and guidance throughout this project. Thanks are extended to the command and staff of the installation at which the study was conducted for their cooperation and support. The authors would like to add a special thanks to Bonnie Storey for all the time and effort she put into this project.

[1]Government Accounting Office (GAO), "Military training: DoD lacks a comprehensive plan to manage encroachment on training ranges," GAO-02–614, Washington, DC, 2002.
[2]Environmental Protection Agency, "Information on levels of environmental noise requisite to protect public health and welfare with an adequate margin of safety," Report No. 550/9–74–004, Office of Noise Abatement and Control, Arlington, VA, 1974.
[3]Committee on Hearing, Bioacoustics and Biomechanics (CHABA), "Assessment of community response to high-energy impulsive sounds," Report No. WG-84, National Academy of Science, National Research Council, Washington, D.C., 1981.
[4]P. D. Schomer, L. R. Wagner, L. J. Benson, K. W. Buchta, and D. Krahé, "Human and community response to military sounds: Results from field-laboratory tests of small arms, tracked-vehicle, and blast sounds," Noise Control Eng. J. **42**, 71–84 (1994).
[5]J. Vos, "A re-analysis of the relationship between the results obtained in laboratory and field studies on the annoyance caused by high energy impulsive sounds," Noise Control Eng. J. **45**, 123–131 (1997).
[6]E. Buchta and J. Vos, "A field survey on the annoyance caused by sounds from large firearms and road traffic," J. Acoust. Soc. Am. **104**, 2890–2902 (1998).
[7]CHABA, *Community Response to High-Energy Impulsive Sounds: An Assessment of the Field Since 1981* (National Academy of Sciences, National Research Council, Washington, D.C., 1996).
[8]G. A. Luz, R. Raspet, and P. D. Schomer, "Analysis of community complaints to noise," J. Acoust. Soc. Am. **73**, 1229–1235 (1983).
[9]R. Hoeger, "Aircraft noise and time of day: Possibilities of redistributing and influencing noise exposure," Noise Health **6**, 55–58 (2004).
[10]D. Larom, M. Garstang, M. Lindeque, R. Raspet, M. Zunckel, Y. Hong, K. Brassel, S. O'Beirne, and F. Sokolic, "Meteorology and elephant infrasound at Etosha National Park," J. Acoust. Soc. Am. **101**, 1710–1717 (1997).
[11]R. Rylander and B. Lundquist, "Annoyance caused by noise from heavy weapon shooting ranges," J. Sound Vib. **192**, 199–202 (1996).
[12]L., Pater, "Noise abatement program for explosive operations at NSWC/DL," Paper presented at the 17th Explosives Safety Seminar of the Department of Defense Safety Board, Denver, CO (1976).
[13]J. S. Lukas and M. E. Dobbs, "Effects of aircraft noises on the sleep of women," NASA Contractor Report No. NASA-CR-2041, Stanford Research Institute, Menlo Park, CA, 1972.
[14]B. Griefahn, "Effects of military noise during sleep. Relations to sex and time of night," in Proceedings of the Fifth International Congress on Noise as a Public Health Problem, edited by B. Berglund, U. Berglund, J. Karlsson, and T. Lindvall, Swedish Council for Building Research, Stockholm, Sweden, Vol. **II**, pp. 39–45 (1988).
[15]S. Fidell, R. Howe, B. Tabachnick, K. Pearsons, L. Silvati, M. Sneddon, and E. Fletcher, "Field studies of habituation to change in nighttime aircraft noise and of sleep motility measurement methods," BBN Report No. 8195, BBN Technologies, Canoga Park, CA, 1988.
[16]S. Fidell, K. Pearson, B. Tabachnick, R. Howe, L. Silvati, and D. S. Barber, "Field study of noise-induced sleep disturbance," J. Acoust. Soc. Am. **98**, 1025–1033 (1995).
[17]J. Vos and M. M. J. Houben, "Sleep disturbance caused by impulse sounds," TNO Report No. TNO-DV 2007 A331, TNO Defence, Security and Safety, The Netherlands, 2007.
[18]G. A. Luz, E. T. Nykaza, C. Stewart, and L. Pater, "Use of actimeters to determine awakenings by sounds of large guns," Noise Control Eng. J. **56**, 211–217 (2008).
[19]E. T. Nykaza, L. Pater, and R. Melton, "Sleep impacts of nighttime training noise from large weapons on residents living near a military installation," ERDC/CERL Report No. TR-06–21, U.S. Army Engineering Research and Development Center, Champaign, IL, 2006.
[20]The use of commercial names in this paper does not constitute advertisement or endorsement of the any products or commercial devices.
[21]N. Oakley, "Validation with polysomnography of the sleepwatch sleep/wake scoring algorithm used by the actiwatch activity monitoring system," Cambridge Neurotechnology Ltd., Cambridge.
[22]C. Kushida, A. Chang, C. Gadkary, C. Guilleminault, O. Carillo, and W. Dement, "Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients," Sleep Med. **2**, 389–396 (2001).
[23]G. A. Luz, E. T. Nykaza, C. Stewart, and L. Pater, "The role of sleep disturbance in predicting community response to the noise of heavy weapons," ERCD/CERL Report No. TR-04–26, U.S. Army Engineering Research and Development Center, Champaign, IL, 2004.
[24]Actimeter instruction manual software version 3.3 and earlier, Mini Mitter Part No. 910–0007–08, 2001, www.minimitter.com.
[25]S. Fidell, R. R. Howe, B. G. Tabachnick, K. S. Pearsons, and M. D. Sneddon, "Noise-induced sleep disturbance in residences near two civil airports," NASA Report Nos. BBN-8096, NAS 1.26.198252, and NAS-CR-1982–52, National Aeronautics and Space Administration, Langley Research Center, Hampton, VA, 1995.
[26]P. D. Schomer, R. J. Goff, and L. M. Little, "Statistics of amplitude and spectrum of blasts propagated in the atmosphere," J. Acoust. Soc. Am. **63**, 1431–1443 (1978).

Measuring hybrid passive-active sound absorption of a microperforated liner at oblique incidence

Pedro Cobo[a] and María Cuesta
Instituto de Acústica, CSIC, Serrano 144, 28006 Madrid, Spain

(Received 26 November 2007; revised 29 August 2008; accepted 23 October 2008)

The performance of hybrid passive-active absorbers at normal incidence has been previously demonstrated. However, their feasibility in realistic environments remains to be confirmed. This paper proposes a hybrid passive-active liner absorbing at oblique incidence. The passive absorber is a $2.44 \times 2.44 \times 1$ mm^3 microperforated panel backed by a 5 cm deep air layer and a rigid ending. An active multichannel cell, composed of five single-input single-output controllers, releases the sound pressure at the entrance of the air cavity in a reduced area of the microperforated panel absorber. Hybrid passive/active absorption of the prototype is measured in an anechoic room up to a maximum incidence angle of 21°. For instance, at 20° such a system is able to provide an average absorption of 94% between 266 and 1500 Hz. In addition, the performance that would provide such a hybrid absorber in a reverberant room has been quantified theoretically by an index describing the active-to-passive reverberation time ratio in terms of the active-to-passive absorption area and absorption coefficient ratios. The proposed prototype, implementing 7% of the active proportion area, would afford a reverberation time reduction between 77.5% and 12% from 100 to 500 Hz.
 [DOI: 10.1121/1.3026328]

PACS number(s): 43.55.Ev, 43.50.Ki [KAC] Pages: 185–190

I. INTRODUCTION

Passive absorbers are typically used to improve the acoustics in buildings, concert halls, and other environments. Traditionally, such absorbers are accomplished with a porous material backed by an air cavity and a rigid wall. Depending on the acoustic properties of the material as well as the thickness of each layer, its absorption spectrum can be tuned in a specific frequency band, with maxima at the frequencies where the air layer thickness is an odd integer of the quarter wavelength. In practice, such absorbers are unable to control low frequency fields due to the prohibitive size and weight requirements. However, this lack of low frequency absorption can be solved by using active techniques.

Hybrid solutions, combining passive absorbers and active techniques, are promising strategies for many broadband noise control applications (vehicles, aircraft, buildings, etc.) where the traditional strategy remains too bulky. Such hybrid strategies are based on the notion of active absorption, first introduced by Olson and May (1953) and afterward validated experimentally by Guicking and Lorentz (1984). These authors proposed the equivalent active $\lambda/4$ absorber by substituting the rigid termination of the passive absorber with a loudspeaker driven so as to minimize the pressure in a microphone just behind the material [pressure-release (PR) condition]. This first active absorber provided 60%–70% of absorption in the 200–300 Hz frequency band using an analog controller. Later on, Furtoss *et al.* (1997) designed a low frequency absorber actively controlling the input impedance of the liner. This methodology was recently applied to flow duct applications (Sellen *et al.*, 2006). Another strategy to implement active absorbers is the *impedance matching condition* (Beyene and Burdisso, 1997), which attempts to cancel the reflected sound field inside the air gap at the rear face of the material. Cobo *et al.* (2003) demonstrated that the performance of both approaches (*PR and matching impedance*) depends on the properties of the material, the PR condition affording higher absorption when the flow resistance of the porous layer matches the acoustic impedance of the air. Since for practical implementations the final prototype must remain lightweight and thin, traditional porous layers can be replaced by microperforated panels (MPPs). Cobo *et al.* (2004) and Cobo and Cuesta (2007) demonstrated recently the feasibility of designing thinner hybrid passive-active absorbers using properly designed MPPs.

Most of the results concerning the performance of such absorbers have been obtained at normal incidence in an impedance tube, and so, their performance in more realistic conditions remains to be analyzed. Some difficulties are encountered when trying to extrapolate the results obtained in the impedance tube to the free field. One relates to the measurement technique. Reflection techniques are more suitable to afford the absorption coefficient of extended samples (Garai, 1993). The main difficulty of such a technique is to separate by time windowing the reflected from the diffracted events at the measured reflection trace. Cobo *et al.* (2007) demonstrated that this procedure can be improved by shaping the frequency response of the loudspeaker-microphone in order to radiate shorter pulses. Another limitation concerns the complexity of the sound field in an extended absorber with respect to that in an impedance tube where only plane waves are propagated at low frequencies. In this case, when the sound pressure just behind the material is released, the input impedance of the system matches that of the air at low frequencies, and high active absorption is guaranteed. Re-

[a]Electronic mail: iacpc24@ia.cetef.csic.es

cently, Cobo and Cuesta (2007) confirmed that releasing the sound pressure at the rear face of the material also provides active absorption in an extended panel at normal incidence.

This paper represents a further step in this research and proposes a new prototype able to provide both passive and active absorptions at oblique incidence. First, such a hybrid MPP absorber is analyzed in Sec. II, where special attention is paid to describe the experimental setup as well as the measurement methodology. Then, some results regarding hybrid absorption coefficient measurement at oblique incidence are presented in Sec. III. Comparisons with predictions are included. Finally, Sec. IV investigates theoretically the potential benefit of using such an active absorber in more real conditions, for instance, a reverberant room. This benefit has been quantified by an index describing the active-to-passive reverberation time ratio of the room.

II. HYBRID PASSIVE/ACTIVE MPP ABSORBER AT OBLIQUE INCIDENCE

Cobo *et al.* (2004) already validated experimentally the hybrid passive-active absorption performance of a MPP in an impedance tube. Also, Cobo and Cuesta (2007) measured the hybrid passive-active absorption of an extended MPP in an anechoic room at normal incidence. Thus, the next step might be the validation of the active absorptive properties of the MPP at oblique incidence. A new prototype has been proposed for this purpose. Theoretically, the absorption coefficient of such a liner at any angle of incidence, θ, should be

$$\alpha(\theta) = 1 - |R(\theta)|^2 = 1 - \left| \frac{z_i(\theta)\cos(\theta) - 1}{z_i(\theta)\cos(\theta) + 1} \right|^2, \quad (1)$$

where $R(\theta)$ is the reflection coefficient of the liner at oblique incidence and $z_i(\theta)$ is the input impedance to the absorber normalized by the characteristic acoustic impedance of the air. In the following, all the impedances are assumed to be normalized by that of the air. For the case of an MPP in front of an air cavity, this input impedance becomes (Cobo *et al.*, 2004)

$$z_i(\theta) = z_m + z_c(\theta), \quad (2)$$

where z_m is the acoustic impedance of the MPP (Maa, 1998) and $z_c(\theta)$ is the acoustic impedance of the air cavity. Depending on the control condition at the back of the cavity, either a rigid ending (passive absorber) or a moving wall releasing the pressure (PR) behind the material (active absorber), $z_c(\theta)$ is

$$z_c(\theta) = \begin{cases} -j\dfrac{\cot(kD\cos\theta)}{\cos\theta} & \text{passive condition} \\ 0 & \text{active condition (PR)}. \end{cases} \quad (3)$$

A. Experimental setup

The proposed hybrid MPP absorber combines the passive material (MPP backed by a 5 cm deep air cavity and a rigid termination) with an active cell in the back panel (Fig. 1). The MPP is designed according to Pfretzschner *et al.* (2006) to provide maximum absorption around 1.1 kHz. It consists of a 2.44×2.44 m^2 steel panel, 1 mm thick, with holes of diameter of 6 mm and perforation ratio of 0.23 (perforation ratio is the quotient open surface/total surface) [Fig. 1(a)]. This sheet is back covered with a mesh 39 μm thick, with perforation ratio of 0.14 and holes with diameter of 35 μm. The passive condition in the air cavity (rigid termination) is implemented using a 1.5 cm thick wood panel. To impose the PR condition, an active cell containing five loudspeakers (secondary sources) and five microphones placed just behind the material, each one in front of the corresponding loudspeaker, has been implemented at the back of the prototype [Fig. 1(b)]. Note that the active cell has a cross configuration, covering about 81 cm in each direction. The active-to-passive absorption area ratio of this cell is 0.07. This active cell is completely sealed to the remaining wooden panel so that the air gap between the MPP and the active cell is still 5 cm deep. In order to create local units, the air cavity contained in the area cell is separated into five equal cavities, each one containing a loudspeaker and a microphone (Fig. 2). Each loudspeaker is driven to reduce the sound pressure field in the related microphone, according to the FXLMS algorithm (Elliott, 2001), implemented in a commercial controller based on a TI C40 DSP. The reference

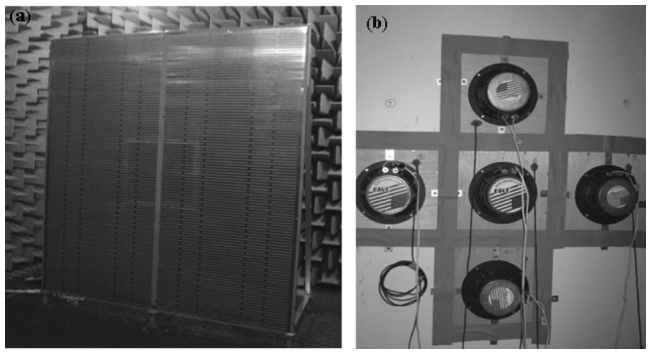


FIG. 1. (a) Front view of the MPP absorber and (b) details of the active cell in the rear wall within an anechoic room.

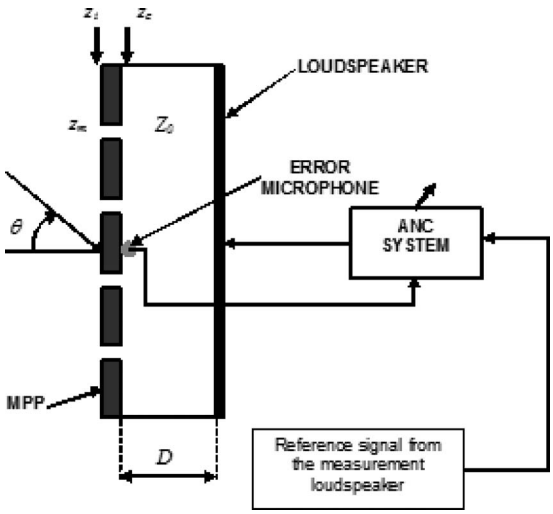


FIG. 2. Sketch of one of the hybrid passive-active MPP absorbers within the multichannel 5I5O cell.

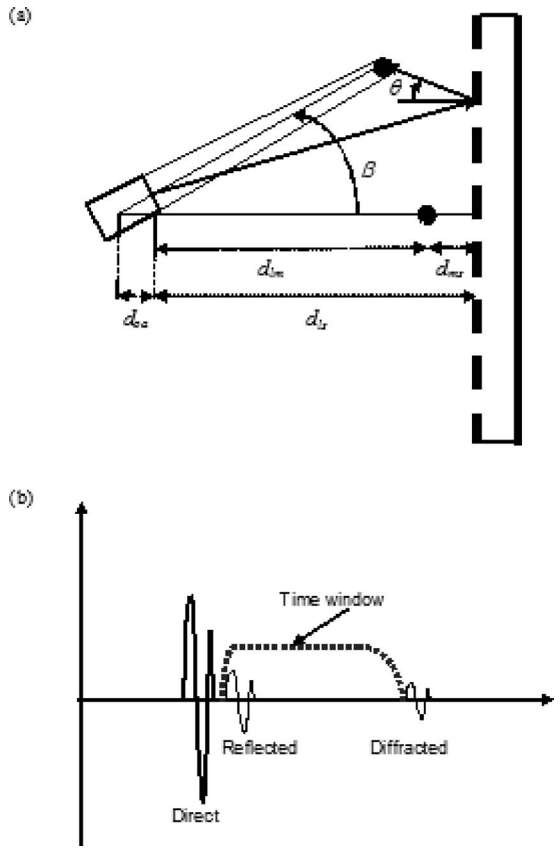


FIG. 3. (a) Measurement setup of absorption at oblique incidence. (b) Sketch of the measured trace with the time window superimposed on the reflected event.

signal for each of these units is taken from the maximum-length sequence (MLS) signal driving the measurement loudspeaker (see Sec. II B). Therefore, the active cell is configured as five local single-input single-output (5I5O) units, thus allowing us to decrease the computational burden of this active cell. According to the model described in Cobo *et al.* (2004), such an absorber provides between 90% and 100% of hybrid passive-active absorption at normal incidence in the range of 100–1500 Hz.

B. Measurement method

Both the passive and active absorption coefficients of such a prototype are measured according to the procedure outlined by Cobo *et al.* (2006). This MLS technique requires a loudspeaker-microphone system optimally positioned according to their relative distances to the sample. The microphone is kept at a fixed distance from the loudspeaker by thin rigid rods. The distance from the loudspeaker to the sample also remains constant.

First, the electroacoustic system should record the direct event (without an absorber). Then, it is addressed toward the absorber, and the reflection trace is measured. Figure 3(a) shows a sketch of the measurement system at oblique incidence. The trace received at the microphone at each incidence angle is drawn in Fig. 3(b). Since this trace also contains diffractions, the direct and reflected events must be picked up by windowing. Direct and reflected events, both filtered with the same window, are used to obtain the absorption coefficient of the prototype. Note that the length of the window, which depends on the geometry, limits the lowest reliable frequency of this procedure. The position of the reflected and diffracted events within the trace also varies with the turn angle of the electroacoustic system. As a consequence, the lowest reliable frequency of this measurement method depends on the turn angle. The turn angle of the system, β, and the incidence angle on the sample, θ [Fig. 3(a)], are related by (Cobo *et al.*, 2006)

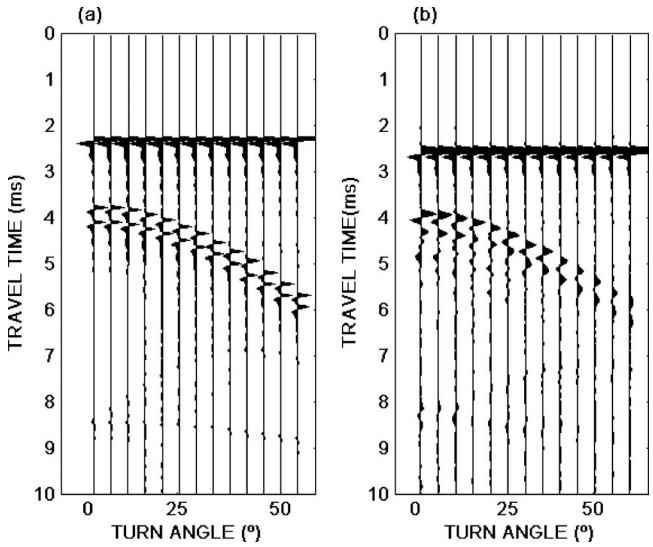


FIG. 4. (a) Synthetic and (b) measured echograms of the travel time traces at the measuring microphone, as a function of the turn angle of the electroacoustic system, for an extended absorber of 2.44×2.44 m^2, loudspeaker-microphone distance of 79 cm, and loudspeaker-sample distance of 104 cm.

$$\tan\theta = \frac{d_{\mathrm{lm}}\sin\beta}{2d_{\mathrm{ls}} + 2d_{oa}(1-\cos\beta) - d_{\mathrm{lm}}\cos\beta}, \tag{4}$$

where d_{lm} is the loudspeaker-microphone distance, d_{ls} is the loudspeaker-sample distance, and d_{oa} is the distance between the turn center and the loudspeaker cone. Another constraint of this method is the maximum incidence angle which can be measured, since the finite size of the sample limits the maximum turn angle of the electroacoustic system. For the absorber prototype of Fig. 1(a), with $d_{\mathrm{lm}}=79$ cm and $d_{\mathrm{ls}}=104$ cm, the maximum turn angle is 60°. Therefore, the maximum incidence angle on the sample is about 21°.

Figure 4 shows the synthetic [Fig. 4(a)] and measured [Fig. 4(b)] echograms of the reflection traces as a function of the turn angle. The synthetic echogram is calculated taking into account the theoretical travel times of the direct, sample reflected, edge diffracted, and loudspeaker diffracted events for the above described prototype. As can be seen, both echograms are well matched. Since the loudspeaker-microphone distance is kept invariable, the travel time of the direct event is also constant. When the turn angle increases, both the travel time of the sample reflected event and that of the loudspeaker diffracted event increase, but the travel time of the edge diffracted event decreases. Thus, the time window to retrieve the sample reflected signal, which must fit between the sample reflected and first diffracted events [Fig. 3(b)], shortens with the turn angle.

The procedure to measure the passive and active absorp-

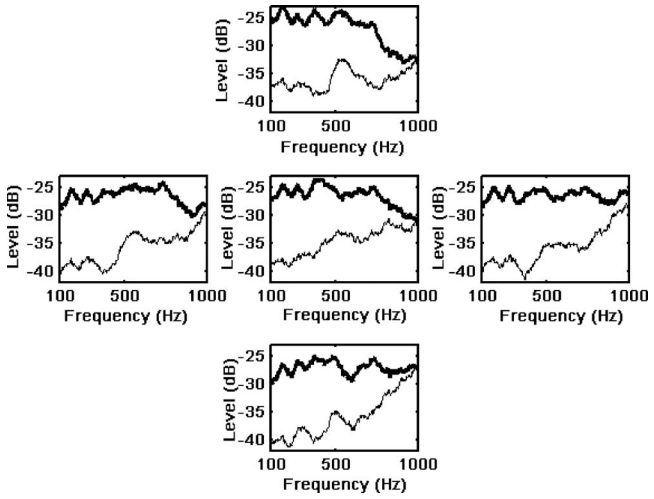


FIG. 5. Spectral levels of the five error microphones inside the active cell with the active controller switched off (thick line) and on (thin line).

tions of the MPP absorber proceeds is as follows:

- For each incidence angle, the passive absorption of the prototype (active cells off) is measured first.
- Then, the multichannel controller is configured to reduce optimally the incident sound field at the error sensor locations. Once this cancellation is optimal, the active absorption coefficient is measured.

III. RESULTS

The hybrid passive-active absorption of the MPP absorber has been measured up to a turn angle of 60°, which corresponds to a maximum incidence angle of 21° [see Eq. (4)]. Figures 5 and 6 illustrate the performance of such an absorber for a turn angle of 50° (incidence angle of 20°). Figure 5 shows the error signal spectra of each microphone inside the active cell, with the active control off (solid line) and on (thin line). These error signals are low-pass filtered at 900 Hz, which is the cutoff frequency between the theoretical passive and active absorption coefficient curves of the

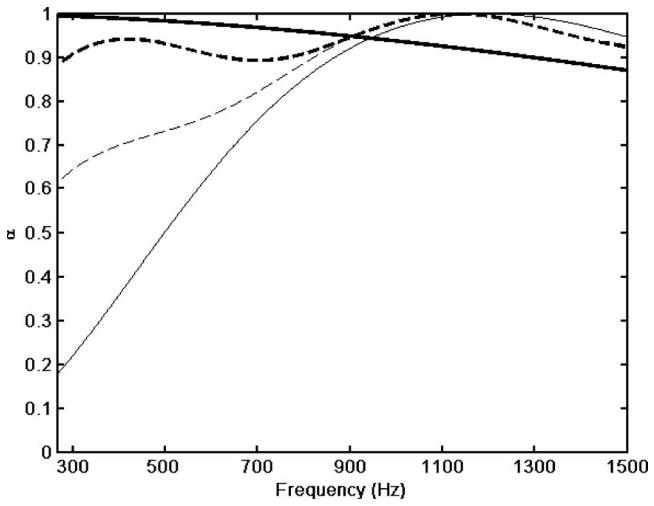


FIG. 6. Theoretical (solid lines) and experimental (dashed lines) passive (thin lines) and hybrid passive-active (thick lines) absorption coefficients in the frequency band of 266–1500 Hz at 20° of incidence angle.

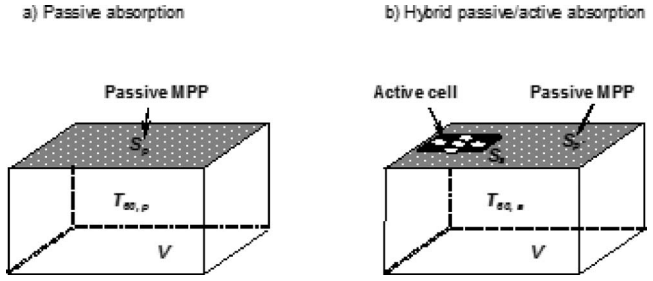


FIG. 7. Scheme of the (a) passive and (b) hybrid passive/active performances of a microperforated absorber standing at the roof of a reverberant room.

prototype, at this angle of incidence. This permits us to design the hybrid absorber, functioning as a passive (active) system above (below) the cutoff frequency. The relative level of the primary MLS field picked up by the microphones is reduced between 5 and 15 dB below this frequency. The controller has been configured with 70 taps in both the control and error path filters and with a convergence coefficient of 0.1.

Figure 6 compares the hybrid absorption of the prototype, measured with this optimized active cell, to the passive one, between 266 Hz (the lowest reliable frequency of the measurement) and 1500 Hz. Predictions for both the passive and active absorbers at this angle of incidence are also included for comparisons. Note that the passive measured absorption coefficient is higher than the expected one below 700 Hz. This is thought to be due to the effect of the active cell in the passive measurement. This passive curve was measured after incorporating the active cell in the back panel. Although this was switched off during the passive measurement, this back panel could no longer be considered rigid. Thus, some sound is transmitted through the active cell, this in turn providing less reflection and more absorption. On the other hand, the measured hybrid passive-active absorption coefficient is slightly lesser than the active predicted one below 900 Hz, the cutoff frequency between the active and passive curves. Above this frequency, the hybrid curve matches the theoretical passive one, as expected since the active cell is low-pass filtered at this frequency. These results confirm the benefit of designing hybrid strategies. The active system is able to provide absorption in the low frequency region, where the passive materials are intrinsically limited for such performance. Similar results have been obtained at other angles.

IV. POTENTIAL BENEFIT OF HYBRID ACTIVE ABSORPTION IN A DIFFUSE FIELD

The final purpose of an active absorber is to be lined in a real environment, for example, a room or an aircraft or vehicle cabin, where a diffuse field is likely set up. Therefore, its performance should be described in terms of the statistical absorption coefficient. It would be very useful to quantify the benefit of using such an active solution in a real application, for example, a room. Let us consider a room of volume V whose roof, of area S_p, is initially treated with the passive MPP absorber previously described [Fig. 7(a)]. This

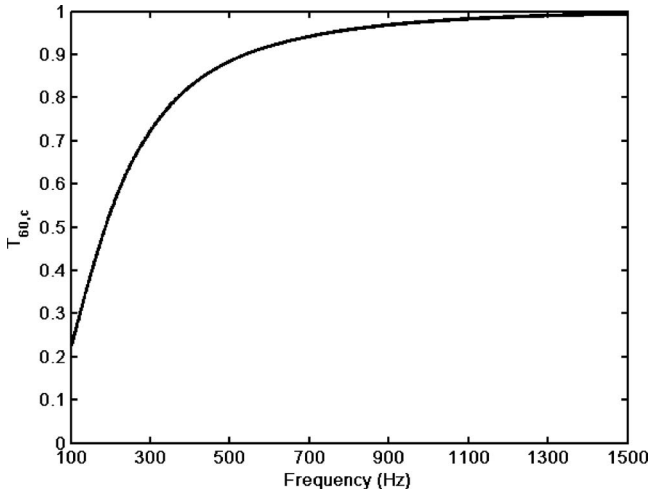


FIG. 8. Time reverberation ratio as a function of frequency for an active cell with an absorption surface ratio of 0.07 and the passive and active absorption coefficients shown in Fig. 6.

passive MPP is characterized by its statistical absorption coefficient α_p. Therefore, if the air absorption of the empty room is ignored, the passive reverberation time of this room, $T_{60,p}$, is (Pierce, 1989)

$$T_{60,p}(f) = \frac{0.161V}{S_p \alpha_p(f)}. \tag{5}$$

In the second stage, the proposed active cell is implemented in section S_a of the roof [Fig. 7(b)]. This reduces the area of the purely passive absorber to $S_{p'} = (S_p - S_a)$. Taking into account that the active cell provides an absorption coefficient α_a, the new reverberation time of the room, $T_{60,a}$, becomes

$$T_{60,a}(f) = \frac{0.161V}{(S_p - S_a)\alpha_p(f) + S_a \alpha_a(f)}. \tag{6}$$

Thus, the performance of such an active absorber can be quantified by an index, $T_{60,c}$, describing the active-to-passive reverberation time ratio of the room,

$$T_{60,c} = \frac{T_{60,a}(f)}{T_{60,p}(f)} = \frac{S_p \alpha_p(f)}{(S_p - S_a)\alpha_p(f) + S_a \alpha_a(f)} = \frac{1}{1 + S_c[\alpha_c(f) - 1]}, \tag{7}$$

where $S_c = S_a / S_p$ and $\alpha_c(f) = \alpha_a(f) / \alpha_p(f)$ are the active-to-passive absorption area and absorption coefficient ratios, respectively. Therefore, the reverberation time ratio shows a hyperbolic dependence with both the absorption coefficient and the absorption area ratios. The reverberation time ratio decreases when both the absorption coefficient and the absorption area ratios increase. On the other hand, the reverberation time ratio approaches 1 when the absorption coefficient ratio tends to 1 and the absorption area ratio tends to zero.

In the above proposed hybrid prototype, the area of the active cell is much lesser than the area of the passive absorption surface. In particular the active-to-passive proportion area is 7% ($S_c = 0.07$). Figure 8 shows the predicted reverberation time reduction that could be expected when lining a reverberant room with such a hybrid passive/active solution at its roof. The main reverberation time reduction is obtained below 500 Hz, as expected. This reverberation time reduction amounts to 77.5% at 100 Hz and to 12% at 500 Hz. Therefore, this study reveals that even a rather small active cell could be able to provide a significant reduction of the reverberation time at low frequencies.

V. CONCLUSIONS

It has already been proved in many previous papers that several embodiments of active absorbers can successfully absorb low frequency sound in an impedance tube. However, the response of such active absorbers to oblique incidence waves remained to be addressed. The main contribution of this paper is the demonstration of the capabilities of hybrid passive-active absorbers at oblique incidence. The proposed hybrid absorber combines a passive MPP with an active 5I5O cell in the back panel.

The measurement of the absorption coefficient of such a hybrid passive-active MPP is carried out by the reflection method. This method measures the impulse response of a loudspeaker-microphone system in front of the panel, at a prescribed incidence angle. Since this impulse response contains also diffractions from the edges of the panel and from the loudspeaker, the direct and sample reflected events must be picked up by time windowing. The time window length, which depends on the size of the sample, limits the lower reliable frequency of the method. For a MPP of 5.95 m^2, the lowest reliable frequency, at an incidence angle of 20°, is 266 Hz. To measure lower frequencies using this method, a larger sample should be used. The measurements of the passive and active absorption coefficients using this method, at 20° of incidence angle, in an anechoic room, are in agreement with the theoretical predictions. A multichannel active cell composed of 5I5O controllers, covering 7% of the panel area, is able to provide an average absorption of 94% in the frequency band of 266–1500 Hz.

Finally, the benefit of using such a hybrid passive-active liner on the reverberation time of an enclosed sound field has been theoretically explored in terms of the size of the absorption area and the increased absorption coefficient, both relative to their passive counterparts. The passive-to-active reverberation time ratio shows a hyperbolic dependence with both the passive-to-active absorption area and absorption coefficient ratios. The reverberation time decreases when both the absorption area and absorption coefficient ratios increase. It is shown that even a rather small active cell, with 7% of the passive absorption area, could provide a 77.5% of reverberation time reduction at 100 Hz.

ACKNOWLEDGMENTS

We are grateful to the Spanish Ministry of Education and Science for funding this research through Grant No. DPI2004-05504-C02-01.

Beyene, S., and Burdisso, R. A. (**1997**). "A new hybrid passive/active noise absorption system," J. Acoust. Soc. Am. **101**, 1512–1515.
Cobo, P., and Cuesta, M. (**2007**). "Hybrid passive-active absorption of a

microperforated panel in free field conditions," J. Acoust. Soc. Am. **121**, EL251–EL255.
Cobo, P., Cuesta, M., Pfretzschner, J., Fernández, A., and Siguero, M. (**2006**). "Measuring the absorption coefficient of panels at oblique incidence by using inverse filtered MLS signals," Noise Control Eng. J. **54**, 414–419.
Cobo, P., Fernández, A., and Cuesta, M. (**2007**). "Measuring short impulse responses with inverse filtered maximum length sequences," Appl. Acoust. **68**, 820–830.
Cobo, P., Fernández, A., and Doutres, O. (**2003**). "Low frequency absorption using a two-layer system with active control of input impedance," J. Acoust. Soc. Am. **114**, 3211–3216.
Cobo, P., Pfretzschner, J., Cuesta, M., and Anthony, D. K. (**2004**). "Hybrid passive-active absorption using microperforated panels," J. Acoust. Soc. Am. **116**, 2118–2125.
Elliott, S. (**2001**). *Signal Processing for Active Control* (Academic, San Diego), Chap. 3.
Furtoss, M., Thenail, D., and Galland, M. A. (**1997**). "Surface impedance control for sound absorption: Direct and hybrid passive/active strategies," J. Sound Vib. **203**, 219–236.
Garai, M. (**1993**). "Measurement of the sound absorption coefficient in situ: The reflection method using periodic pseudo-random sequences of maximum length," Appl. Acoust. **39**, 119–139.
Guicking, D., and Lorentz, E. (**1984**). "An active sound absorber with porous plate," ASME J. Vib., Acoust., Stress, Reliab. Des. **106**, 389–392.
Maa, D. Y. (**1998**). "Potential of microperforated panel absorber," J. Acoust. Soc. Am. **104**, 2861–2866.
Olson, H. F., and May, E. G. (**1953**). "Electronic sound absorber," J. Acoust. Soc. Am. **25**, 1130–1136.
Pfretzschner, J., Cobo, P., Simón, F., Cuesta, M., and Fernández, A. (**2006**). "Microperforated insertion units: An alternative strategy to design microperforated panels," Appl. Acoust. **67**, 62–73.
Pierce, A. D. (**1989**). *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, Melville), Chap. 6.
Sellen, N., Cuesta, M., and Galland, M. A. (**2006**). "Noise reduction in a flow duct: Implementation of a hybrid passive/active solution," J. Sound Vib. **297**, 492–511.

A broadband fast multipole accelerated boundary element method for the three dimensional Helmholtz equation

Nail A. Gumerov[a),b)] and Ramani Duraiswami[b),c)]
Perceptual Interfaces and Reality Laboratory, Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland 20742

(Received 2 January 2008; revised 3 October 2008; accepted 10 October 2008)

The development of a fast multipole method (FMM) accelerated iterative solution of the boundary element method (BEM) for the Helmholtz equations in three dimensions is described. The FMM for the Helmholtz equation is significantly different for problems with low and high kD (where k is the wavenumber and D the domain size), and for large problems the method must be switched between levels of the hierarchy. The BEM requires several approximate computations (numerical quadrature, approximations of the boundary shapes using elements), and these errors must be balanced against approximations introduced by the FMM and the convergence criterion for iterative solution. These different errors must all be chosen in a way that, on the one hand, excess work is not done and, on the other, that the error achieved by the overall computation is acceptable. Details of translation operators for low and high kD, choice of representations, and BEM quadrature schemes, all consistent with these approximations, are described. A novel preconditioner using a low accuracy FMM accelerated solver as a right preconditioner is also described. Results of the developed solvers for large boundary value problems with $0.0001 \lesssim kD \lesssim 500$ are presented and shown to perform close to theoretical expectations. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3021297]

PACS number(s): 43.55.Ka, 43.20.Fn, 43.28.Js [NX] Pages: 191–205

I. INTRODUCTION

Boundary element methods (BEMs) have long been considered as a very promising technique for the solution of many problems in computational acoustics governed by the Helmholtz equation. They can handle complex shapes, lead to problems in boundary variables alone, and lead to simpler meshes where the boundary alone must be discretized rather than the entire domain. Despite these advantages, one issue that has impeded their widespread adoption is that they lead to linear systems with dense and possibly nonsymmetric matrices. As the domain size increases to many wavelengths, the number of variables in the discretized problem, N, should increase correspondingly to satisfy the Nyquist sampling criterion. For a problem with N unknowns, a direct solution requires $O(N^3)$ solution cost and storage of $O(N^2)$. Use of iterative methods does not reduce the memory but can reduce the cost to $O(N_{\text{iter}}N^2)$ operations, where N_{iter} is the number of iterations required, and the $O(N^2)$ per iteration cost arises from the dense matrix-vector product. In practice this is still quite large. An iteration strategy that minimizes N_{iter} is also needed. Other steps in the BEM are also expensive, such as the computation of the individual matrix elements, which require quadrature of nonsingular, weakly singular, or hypersingular functions. To reduce the singularity order and achieve symmetric matrices, many investigators employ Galerkin techniques, which lead to further $O(N^2)$ integral computations. Because of these reasons, the BEM was not used for very large problems. In contrast finite-difference and finite element methods, despite requiring larger volumetric discretizations, have well established iterative solvers and are more widely used.

The combination of the fast multipole method[1] (FMM) and the preconditioned Krylov iterative methods presents a promising approach to improving the scalability of BEMs and is an active area of research. The FMM for potential problems allows the matrix-vector product to be performed to a given precision ϵ in $O(N)$ operations and further does not require the computation or storage of all N^2 elements of the matrices, reducing the storage costs to $O(N)$ as well. Incorporating the fast matrix-vector product in a quickly convergent iterative scheme allows the system of equations to be rapidly solved with $O(N_{\text{iter}}N)$ cost. The FMM was initially developed for gravity or electrostatic potential problems. Later this method was intensively studied and extended to the solution of problems arising from the Helmholtz, Maxwell, biharmonic, elasticity, and other equations. While the literature and previous work on the FMM is extensive, reasons of space do not permit a complete discussion of the literature. The reader is referred to Ref. 2 for a comprehensive review.

II. FMM AND FMM ACCELERATED BEM FOR THE HELMHOLTZ EQUATION

The FMM and FMM accelerated BEM for the Helmholtz equation have seen significant work, and several authors have recently published on various aspects of the problem,[3–9] and new articles have appeared when this article was in review.[10] This article presents a FMM accelerated boundary element solver for the Helmholtz equation that has

[a)]Electronic mail: gumerov@umiacs.umd.edu
[b)]Also at Fantalgo, LLC, Elkridge, MD 20751.
[c)]Electronic mail: ramani@umiacs.umd.edu

the following contributions that distinguish it from previous work:

- achieves good performance at both low and high frequencies by changing the representation used,
- has a quickly convergent iterative scheme via a novel preconditioner, which is described here,
- achieves further efficiency by considering the error of the quadrature in the BEM in the overall error analysis of the FMM, and
- uses the Burton–Miller (combined) boundary integral formulation for external problems and avoids the problems with spurious resonances.

A. Error and fast multipole accelerated boundary elements

During the approximate solution of the Helmholtz equation via the FMM accelerated BEM, the following different sources of error are encountered:

- geometric error due to the discretization of the surface with meshes,
- quadrature error in computation of boundary integrals,
- matrix-vector product error due to the FMM, and
- residue error used as a termination criterion during the iterative solution process.

One of the contributions of this paper is to consider all the errors together and design an algorithm that provides the required accuracy and avoids wasteful computations. It should be noted that the error achieved in practice in many reported simulation error tolerances is quite high, from a high of a few percent to at most 10^{-4} (see, e.g., Refs. 4, 5, and 11). Many previous FMM/BEM simulations contain computations that are wasteful of CPU time and memory when considering the final error desired or achieved.

B. Resolving calculations over a large range of wavenumbers

To be accurate, any calculation must resolve the smallest wavelengths of interest, and to satisfy the Nyquist criterion, the discretization must involve at least two points per wavelength. The restriction imposed by this requirement manifests itself at large frequencies since at lower frequencies the discretization is controlled by the necessity to accurately represent the boundary. Thus, two basic regimes for the FMM BEM are usually recognized in acoustic simulations: the low-frequency regime and the high-frequency regime. These regimes can be characterized by some threshold value $(kD)_*$ of the parameter kD, where k is the wavenumber and D is the computational domain size. For each of these regimes the computational complexity of the FMM exhibits a different behavior.[12]

1. Low-frequency regime

In the low-frequency regime, $kD<(kD)_*$, the per iteration step cost of the FMM is proportional to N and not very much affected by the value of kD. Here, the most efficient representation is in terms of spherical multipole wavefunctions, and the translation schemes are based on the rotation-coaxial translation-backrotation (RCR) decompositions,[12,13] which have $O(p^3)$ complexity. Here p^2 is the number of terms in the multipole expansion (p in this regime can be constant). Alternately one may use the low-frequency exponential forms,[14,15] which have the same complexity, but with a different asymptotic constant. The method of function representation based on sampling of the far-field signature function[16] is not stable in this region due to exponential growth of terms in the multipole-to-local translation kernel.

2. High-frequency regime

In the high frequency regime, $kD>(kD)_*$, and the value of kD heavily affects the cost. Since the wavenumber k is inversely proportional to wavelength and, in practice, five to ten points per wavelength are required for accuracy, for a surface-based numerical method (such as the BEM) the problem size N scales as $O(kD)^2$, while for volumetric problems (e.g., for many scatterers distributed in a volume) N scales as $O(kD)^3$. For this regime, the size of the wavefunction representation, which is $O(p^2)$, must increase as the levels go up in the hierarchical space subdivision, with p proportional to the size of the boxes at a given level.[12] Because of this, the complexity of the FMM is heavily affected by the complexity of a single translation. It was shown[12] that $O(p^3)$ schemes (such as the RCR scheme) result in an overall complexity of the FMM $O(kD)^3$ for simple shapes and $O((kD)^3 \log(kD))$ for space-filling surfaces. The use of translation schemes of $O(p^4)$ and $O(p^5)$ complexities in this case provides the overall complexity of the FMM that is slower than the direct matrix-vector product. Where FMM with translation schemes of such complexity have been used with the BEM (e.g., see Ref. 7), one must recognize that the software is only usable in the low-frequency regime.

To reach the best scaling algorithm in the high-frequency regime, translation methods based on representations that sample the far-field signature function[16] are necessary. The translation cost in this case scales as $O(p^2)$, while at least $O(p^2 \log p)$ additional operations are needed for the spherical filtering necessary for numerical stabilization of the procedure. In this case the overall FMM complexity will be $O((kD)^2 \log^\alpha (kD))$ $(\alpha \geqslant 1)$ for simple shapes and $O((kD)^3)$ for space-filling shapes.

3. Switch in function representations

As discussed above, different representations are appropriate for low and high kD. However, even for high kD problems, since the FMM employs a hierarchical decomposition, at the fine levels, the problems behave as a low kD problem. Indeed at the fine levels, parameter ka, where a is a representative box size, is smaller than $(kD)_*$, and translations appropriate to the low-frequency regime should be used. For coarser levels, ka is large and the high-frequency regime should be used, and a combined scheme in which the spherical wavefunction representation can be converted to signature function sample representation is needed. Such a switch was also suggested and tested recently in Ref. 17. The

present scheme, however, is different from that of Ref. 17 and does not require interpolation/anterpolation.

4. Comparison with volumetric methods

Similarly volumetric methods, such as the finite-difference time domain (FDTD) method for the wave equation, have the requirement that to resolve a simulation they need several points per wavelength. Thus for a problem on a domain of size kD, in three dimensions we have $N \sim (kD)^3$, with this restriction being controlling at higher frequencies. Fast iterative methods, e.g., based on multigrid solve these systems in constant number of iterations and for a cost of $O(N^{4/3})$ per time step. So for a problem with M time steps, we have $O(MN^{4/3}) = O(M(kD)^4)$ complexity for the volumetric methods (this is based on the discussion in Ref. 18). In contrast a fast multipole accelerated BEM, which is preconditioned with a preconditioner requiring a constant number of steps and is solved at M frequencies, can achieve solution in $O(M(kD)^2 \log(kD))$ or $O(M(kD)^3)$ steps. However, in terms of programming ease the FDTD and finite element TD methods are much easier to implement, and the preconditioners developed for them work better. Further they are easier to generalize to nonisotropic media. Their difficulty in handling infinite domains has been largely solved via perfectly matched layer methods. Accordingly, despite the advantages provided by integral equation approaches for the Helmholtz equation, volumetric approaches are popular.

C. Iterative methods and preconditioning

Preconditioning can be very beneficial for fast convergence of Krylov subspace iterative methods. Preconditioning for boundary element matrices is, in general, a lesser studied issue than for finite-element- and finite-difference-based discretizations. From that theory, it is known that for high wavenumbers preconditioning is difficult and an area of active research. Many conventional preconditioning strategies rely on sparsity in the matrix, and applying them to dense BEM matrices requires computations that have a formal time or memory complexity of $O(N^2)$, which negates the advantage of the FMM.

One strategy that has been applied with the FMBEM is the construction of approximate inverses for each row based on a local neighborhood of the row. If K neighboring elements are considered, then constructing this matrix has a cost of $O(NK^3)$, and there is a similar cost to applying the preconditioner at each step.[3,4] However such local preconditioning strategies appear to work well only for low wavenumbers. Instead in this paper the use of a low accuracy FMM itself as a preconditioner by using a flexible generalized minimal residual (fGMRES) procedure[19] is considered. This novel preconditioner appears to work reasonably at all wavenumbers considered and stays within the required cost.

III. FORMULATION AND PRELIMINARIES

A. Boundary value problem

Consider the Helmholtz equation for the complex valued potential ϕ,

$$\nabla^2\phi + k^2\phi = 0, \tag{1}$$

with real wavenumber k inside or outside finite three dimensional (3D) domain V bounded by closed surface S, subject to mixed boundary conditions

$$\alpha(\mathbf{x})\phi(\mathbf{x}) + \beta(\mathbf{x})q(\mathbf{x}) = \gamma(\mathbf{x}), \quad q(\mathbf{x}) = \frac{\partial\phi}{\partial n}(\mathbf{x}),$$
$$|\alpha| + |\beta| \neq 0, \quad \mathbf{x} \in S. \tag{2}$$

Here and below all normal derivatives are taken, assuming that the normal to the surface is directed outward to V. For external problems ϕ is assumed to satisfy the Sommerfeld radiation condition

$$\lim_{r\to\infty}\left[r\left(\frac{\partial\phi}{\partial r} - ik\phi\right)\right] = 0, \quad r = |\mathbf{x}|. \tag{3}$$

This means that for scattering problems ϕ is treated as the scattered potential.

Note then that there should be some constraints on surface functions $\alpha(\mathbf{x})$, $\beta(\mathbf{x})$, and $\gamma(\mathbf{x})$ for existence and uniqueness of the solution. Particularly, if α and β are constant, this leads to the Robin problem, which degenerates to the Dirichlet or Neumann problem. For $\beta=0$ the case of a "sound-soft" boundary is obtained, and for $\alpha=0$ the "sound-hard" boundary case is obtained.

B. Boundary integral equations

The BEM uses a formulation in terms of boundary integral equations whose solution with the boundary conditions provides $\phi(\mathbf{x})$ and $q(\mathbf{x})$ on the boundary and subsequently determines $\phi(\mathbf{y})$ for any domain point $\mathbf{y}$. This can be done, e.g., using Green's identity

$$\pm\phi(\mathbf{y}) = L[q] - M[\phi], \quad \mathbf{y} \notin S. \tag{4}$$

Here the upper sign in the left hand side should be taken for the internal domain, while the lower sign is for the external domain (this convention is used everywhere below), and L and M denote the following boundary operators:

$$L[q] = \int_S q(\mathbf{x})G(\mathbf{x},\mathbf{y})dS(\mathbf{x}),$$
$$M[\phi] = \int_S \phi(\mathbf{x})\frac{\partial G(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{x})}dS(\mathbf{x}), \tag{5}$$

where G is the free-space Green's function for the Helmholtz equation

$$G(\mathbf{x},\mathbf{y}) = \frac{e^{ikr}}{4\pi r}, \quad r = |\mathbf{x}-\mathbf{y}|. \tag{6}$$

In principle, Green's identity can be also used to provide necessary equations for determination of the boundary values of $\phi(\mathbf{x})$ and $q(\mathbf{x})$, as in this case for smooth S one obtains

$$\pm\tfrac{1}{2}\phi(\mathbf{y}) = L[q] - M[\phi], \quad \mathbf{y} \in S, \tag{7}$$

The well-known deficiency of this formulation is related to possible degeneration of the operators L and $(M-\frac{1}{2})$ at cer-

tain frequencies depending on S, which correspond to resonances of the internal problem for sound-soft and sound-hard boundaries.[20,21] Even though the solution of the external problem is unique for these frequencies, Eq. (7) is deficient in these cases. Moreover, for frequencies in the vicinity of the resonances, the system becomes poorly conditioned numerically. On the other hand, when solving internal problems (e.g., in room acoustics), the nonuniqueness of the solution for the internal problem has a physical meaning, as there are resonances.

In any case, Eq. (7) can be modified to avoid the artifact of degeneracy of boundary operators when solving the correctly posed problems (1)–(3). This can be done using different techniques, including direct and indirect formulations, introduction of some additional field points, etc. A direct formulation based on the integral equation combining Green's and Maue's identities, which is the method proposed by Burton and Miller[20] for sound-hard boundaries, is used. The Maue identity is

$$\pm \tfrac{1}{2} q(\mathbf{y}) = L'[q] - M'[\phi], \quad \mathbf{y} \in S, \tag{8}$$

where

$$L'[q] = \int_S q(\mathbf{x}) \frac{\partial G(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{y})} dS(\mathbf{x}),$$
$$M'[\phi] = \frac{\partial}{\partial n(\mathbf{y})} \int_S \phi(\mathbf{x}) \frac{\partial G(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{x})} dS(\mathbf{x}). \tag{9}$$

Multiplying Eq. (8) by some complex constant λ and summing with Eq. (7), one obtains

$$\pm \tfrac{1}{2}[\phi(\mathbf{y}) + \lambda q(\mathbf{y})] = (L + \lambda L')[q] - (M + \lambda M')[\phi]. \tag{10}$$

Burton and Miller[20] proved that it is sufficient to have $\mathrm{Im}(\lambda) \neq 0$ to guarantee the uniqueness of the solution for the external problem.

C. Combined equation

Using the boundary conditions, the system of equations (2) and (10) can be reduced to a single linear system for some vector of elemental or nodal unknowns $[\psi]$,

$$A[\psi] = c, \tag{11}$$

which is convenient for computations. The boundary operator A and functions ψ and c can be constructed following standard BEM procedures. They can be expressed as

$$A[\psi] = (L + \lambda L')[u'] - (M + \lambda M')[u] \mp \tfrac{1}{2}(u + \lambda u'),$$
$$c = (L + \lambda L')[b'] - (M + \lambda M')[b] \mp \tfrac{1}{2}(b + \lambda b'), \tag{12}$$

where u and u' are related to the unknown ψ, and b and b' are related to the knowns. For example, for the Neumann problem $u=\phi$, $u'=0$, while $b'=q$ and $b=0$.

D. Discretization

Boundary discretization leads to approximation of boundary functions via finite vectors of their surface samples and integral operators via matrices acting on these vectors. For example, if the surface is discretized by a mesh with M panels (elements), $S_{l'}$, and N vertices, $\mathbf{x}_j$, and integrals over the boundary elements are computed, one obtains

$$L[q](\mathbf{x}_l^{(c)}) = \sum_{l'=1}^{M} \int_{S_{l'}} q(\mathbf{x}) G(\mathbf{x},\mathbf{x}_l^{(c)}) dS(\mathbf{x})$$
$$\approx \sum_{l'=1}^{M} L_{ll'} q_{l'}, \quad l = 1, \ldots, M,$$
$$q_{l'} = q(\mathbf{x}_{l'}^{(c)}), \quad L_{ll'} = \int_{S_{l'}} G(\mathbf{x},\mathbf{x}_l^{(c)}) dS(\mathbf{x}), \tag{13}$$

where $\mathbf{x}_{l'}^{(c)}$ is the center of the l'th element, and for computations of matrix entries $L_{ll'}$ one can use well-known quadratures, including those for singular integrals.[21,22] The above equation is for the case of panel collocation, while analogous equations can be derived for vertex collocation. Similar formulas can be used for other operators. Note that to accurately capture the solution variation at the relevant length scales, the discretization should satisfy $kr_{\max} \ll 1$, where $r_{\max}$ is the maximum size of the element. In practice, discretizations that provide several elements per wavelength usually achieve an accuracy consistent with the other errors of the BEM. The above method of discretization with collocation either at the panel centers or vertices was implemented and tested. Discretization of the boundary operators reduces problem (11) to a system of linear equations.

E. Iterative methods

Different iterative methods can be tried to solve Eq. (11), which has a nonsymmetric dense complex valued matrix A. Any iterative method requires computation of the matrix-vector product $A[x]$, where $[x]$ is some input vector. The method used in the present algorithm is the fGMRES method,[19] which has the advantage that it allows use of approximate right preconditioner, which in its turn can be computed by executing of the internal iteration loop using unpreconditioned GMRES.[23] Choice of the preconditioning method must be achieved for a cost that is $O(N)$ or smaller. This flexibility is exploited in the present algorithm.

IV. USE OF THE FAST MULTIPOLE METHOD

The main idea of the use of the FMM for the solution of the discretized boundary integral equation is based on the decomposition of operator A,

$$A = A_{\mathrm{sparse}} + A_{\mathrm{dense}}, \tag{14}$$

where the sparse part of the matrix has only nonzero entries A_{lj} corresponding to the vertices $\mathbf{x}_l$ and $\mathbf{x}_j$, such that $|\mathbf{x}_l - \mathbf{x}_j| < r_c$, where r_c is some distance usually of the order of the distance between the vertices, whose selection can be based on some estimates or error bounds, while the dense part has nonzero entries A_{lj} for which $|\mathbf{x}_l - \mathbf{x}_j| \geqslant r_c$. The use of the FMM reduces the memory complexity of the overall product to $O(N)$ and the computational complexity to $o(N^2)$,

which can be $O(N)$, $O(N\log^{\beta}N)$, $\beta \geqslant 1$, or $O(N^{\alpha})$, $\alpha < 2$, depending on the wavenumber, domain size, effective dimensionality of the boundary, and translation methods used.[12]

A. FMM strategy

The use of the FMM for the solution of boundary integral equations brings a substantial shift in the computational strategy. In the traditional BEM the full system matrix must be computed to solve the resulting linear system either directly or iteratively. The memory needed to store this matrix is fixed and is not affected by the accuracy imposed on the computation of the surface integrals. Even if one uses quadratures with a relatively high number of abscissas and weights to compute integrals over the flat panels in a constant panel approximation, the memory cost is the same, and the relative increase in the total cost is small, as that cost is dominated by the linear system solution.

If one chooses, as done by previous authors using the FMM accelerated BEM (see, e.g., Refs. 5 and 7), to compute nonsingular integrals very accurately in the FMM using expansions of Green's function, such as

$$G(\mathbf{x},\mathbf{x}_l^{(c)}) = ik\sum_{n=0}^{\infty}\sum_{m=-n}^{n} R_n^{-m}(\mathbf{x}-\mathbf{x}_{l'}^{(c)})S_n^m(\mathbf{x}_l^{(c)}-\mathbf{x}_{l'}^{(c)}), \qquad (15)$$

where R_n^m and S_n^m are the spherical basis functions for the Helmholtz equation, then from Eq. (13) the expressions for the matrix entries may be obtained as

$$L_{ll'} = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} C_n^m S_n^m(\mathbf{x}_l^{(c)}-\mathbf{x}_{l'}^{(c)}),$$

$$C_n^m = ik\int_{S_{l'}} R_n^{-m}(\mathbf{x}-\mathbf{x}_{l'}^{(c)})dS(\mathbf{x}). \qquad (16)$$

As the sum is truncated for maximum $n=p-1$, then there are p^2 complex expansion coefficients per element. If this p is the same as the truncation number for the FMM, this requires substantial memory to store Mp^2 complex values.

A different strategy is proposed here. To reduce the memory consumption, one should use schemes where the integrals are computed at the time of the matrix-vector product and only at the necessary accuracy. In the case of the use of higher-order quadratures, one is faced then with a well-known dilemma to either compute integrals in the flat panel approximation with higher-order formulas or just increase the total number of nodes (discretization density) and use lower-order quadrature. In the case of use of the FMM with "on the fly" integral computations, the computational complexity will be almost the same for both ways, while the latter way seems preferable, as it allows the function to vary from point to point and employs better approximation for the boundary (as the vertices are located on the actual surface and variations of the surface normal are accounted for better).

Therefore, in the case of the use of the FMM, one can try to use the following approximation, at least in the far-field (the dense part), for the nonsingular integrals:

$$L_{lj} = s_j G(\mathbf{x}_j,\mathbf{x}_l), \quad M_{lj} = s_j \frac{\partial G}{\partial n_j}(\mathbf{x}_j,\mathbf{x}_l),$$

$$L'_{lj} = s_j \frac{\partial G}{\partial n_l}(\mathbf{x}_j,\mathbf{x}_l), \quad M'_{lj} = s_j \frac{\partial^2 G}{\partial n_l \partial n_j}(\mathbf{x}_j,\mathbf{x}_l), \qquad (17)$$

$$l,j = 1,\ldots,N, \quad \mathbf{x}_l \neq \mathbf{x}_j,$$

where in the case of panel collocation, s_j are the panel areas, $\mathbf{x}_j$ are the centers of panels, and n_j is the normal to the panels. In the case of vertex collocation, these quantities are appropriately modified. For the treatment of the singular integrals ($\mathbf{x}_l=\mathbf{x}_j$), a method described later is used.

For near-field computations, these formulas could be used with a fine enough discretization for the nonsingular integrals, although one may prefer to use higher-order quadrature. Several tests, using for near-field integral representation Gauss quadratures of varying order (in the range of 1–625 nodes per element), showed that approximation (17) used for near field provides fairly good results for good meshes.

B. FMM algorithm

The Helmholtz FMM algorithm employed for matrix-vector products is described in Refs. 12 and 24, with modifications that allow use of different translation schemes for low and high frequencies. Particulars of the algorithm are that a level-dependent truncation number p_l is used and that rectangularly truncated translation operators are employed for multipole-to-multipole and local-to-local translations. These are performed using the RCR-decomposition and result in $O(p^3)$ single translation complexity. The RCR-decomposition is also used for the multipole-to-local translations for levels with $ka_l < (kD)_*$, where a_l is the radius of the circumsphere of a box on level l. For levels corresponding to $ka_l \geqslant (kD)_*$, the multipole expansions are converted to samples of the signature function at a cost of $O(p^3)$, and then diagonal forms of the translation operator $O(p^2)$ are used, and in the downward pass at some appropriate level conversion of the signature function to the local expansion of the required length at a cost of $O(p^3)$ is used. This procedure automatically provides filtering to ensure that the representation has the correct bandwidth. It must be noted that conversions from multipole to local expansions are required only once per box since consolidation of the translated functions is performed in terms of signature functions. This amortizes the $O(p^3)$ conversion cost and makes the scheme faster than the one based on the RCR-decomposition for the same accuracy. The algorithm, in this part, is thus close to the one described in Ref. 17. The difference is that interpolation/anterpolation procedures are unnecessary here. Also, for low-frequency translation the RCR-decomposition for the multipole-to-local translation is used and found to be as efficient as the method based on conversion into exponential forms for moderate p. Particulars of the present implementation include a precomputation of all translation operators, particularly translation kernels, so during the run time of the

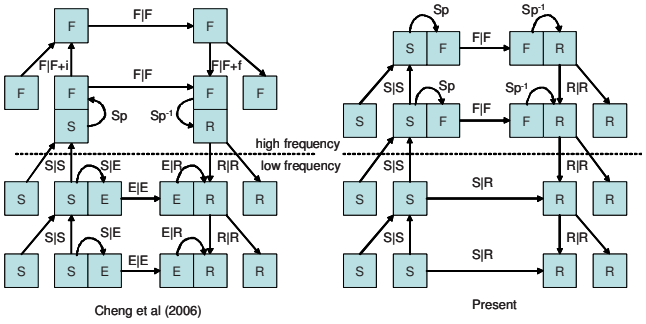


FIG. 1. (Color online) Illustration comparing the wideband FMMs of Cheng *et al.* (Ref. 17) and that presented in this paper for a problem in which the FMM octree has four levels and in which the high-low frequency switch threshold occurs between levels 2 and 3. The left hand side for each algorithm shows the FMM upward pass, while the right hand side shows the FMM downward pass. Each box represents various steps for that level, such as multipole expansion (S), local expansion (R), far-field signature function samples (F), and exponential form for each coordinate direction (E). The "glued" boxes mean that for a given box at that level the two types of expansions are constructed. $S|S$, $R|R$, $S|R$, $E|E$, and $F|F$ denote translation operators acting on the respective representations. **Sp** and $\mathbf{Sp}^{-1}$ denote forward and inverse spherical transform, $S|E$ and $E|R$ are the respective conversion operators. $F|F+i$ and $F|F+f$ mean that the translation is accompanied by use of an interpolation or filtering procedure.

procedure, which is performed many times for the iterative process, only simple arithmetic operations (additions and multiplications) are executed.

1. Comparison of algorithms

Figure 1 illustrates the present 3D Helmholtz FMM algorithm (on the right) and also compares it with that proposed in Ref. 17. These algorithms have in common the separation of the high- and low-frequency regions where different translation methods are used. It is seen that the present algorithm at high frequencies implements the idea used in the algorithm[17] for lower frequencies, while instead of conversion to the exponential form the spherical transform is used to convert the multipole expansion to the signature function representation and back. The signature function representation is omnidirectional and, in contrast to the exponential forms, does not require additional data structures and multiple representations (since translations for this representation are different in each coordinate direction). Also, this approach is valid for the values of p are necessary for the high-frequency region. However, despite the use of these efficient techniques, the present algorithm has a formal translational scaling $O(p^3)$ since in the high-frequency region for the multipole-to-multipole $S|S$ and local-to-local $R|R$ operators the spherical function representations are used.

2. Data structure

The version of the FMM used in this paper employs an octree-based data structure, when the computational domain is enclosed into a cube of size $D\times D\times D$, which is assigned to level 0, and further the space is subdivided by the octree to the level $l_{\max}$. The algorithm works with cubes from level 2 to $l_{\max}$. For generation of the data structure, we use hierarchical box ordering based on the bit interleaving and precompute lists of neighbors and children, which are stored and used as needed. The FMM used skips "empty" boxes at all levels.

3. Level-dependent truncation number

Each level is characterized by the size of the expansion domain, which is the radius a_l of the circumsphere of the boxes at level l. Selection of the truncation number in the algorithm is automated based on an expression of the form $p_l=p(ka_l,\epsilon,\delta)$, where ϵ is the prescribed accuracy and δ the separation parameter [we used $\delta=2$ (see the justification in Ref. 12)]. A detailed discussion and theoretical error bounds can be found elsewhere (see e.g., Refs. 12 and 25). Particularly, the following approximation combining low- and high-frequency asymptotics for monopole expansions can be utilized:[12]

$$p_{\mathrm{lo}}=1-\frac{\log\epsilon(1-\delta^{-1})^{3/2}}{\log\delta},$$

$$p_{\mathrm{hi}}=ka+\frac{(3\log 1/\epsilon)^{2/3}}{2}(ka)^{1/3}, \tag{18}$$

$$p=(p_{\mathrm{lo}}^4+p_{\mathrm{hi}}^4)^{1/4}.$$

It is also shown in Ref. 12 that for the use of the rectangularly truncated translation operators the principal term of the error can be evaluated based on this dependence. The numerical experiments show that the theoretical bound frequently overestimates the actual errors, so some corrections can be also applied. The software developed here, in its automatic setting, computes p_{lo} and p_{hi} and—if it happens that $p-p_{\mathrm{hi}}>p_*(\epsilon)$, where p_* is some number dictated by the overall accuracy requirements—uses $p=p_{\mathrm{hi}}+p_*(\epsilon)$; otherwise, Eq. (18) is used. In fact, we also had some bound for $p_*(\epsilon)$ depending on ka to avoid blowout in computation of functions at extremely low ka.

As previously mentioned, an automatic switch was implemented from the RCR-decomposition to the diagonal forms of the translation operators based on criterion $ka_l\geqslant(kD)_*$. The parameter $(kD)_*$ was based on the error bounds (18) and was selected for the level at which $p-p_{\mathrm{hi}}\leqslant p_{**}$ (we used $p_{**}=2$). This is dictated by the estimation of the threshold at which the magnitude of the smallest truncated term in the translation kernel (26) starts to grow exponentially (see Ref. 12).

Figure 2 illustrates the dependence provided by Eq. (18). We note that FMM with coarse accuracy like $\epsilon=10^{-2}$ can be used for efficient preconditioning. We also can remark that function representation via the multipole expansions and use of the matrix-based translations (such as RCR-decomposition) is not the only choice, and in Refs. 14 and 17 a method based on diagonalization of the translation operators, different from Ref. 16, was developed. This method, however, requires some complication in data structure (decomposition to the x-, y-, and z-directional lists) and is efficient for moderate to large truncation numbers. As we mentioned, the truncation numbers in the low-frequency region can be reduced (plus the BEM itself has a limited accuracy

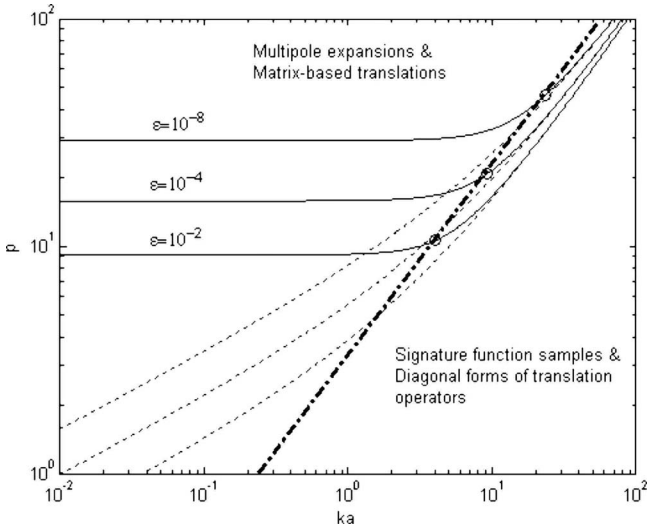


FIG. 2. Dependences of the truncation number p on the dimensionless domain size ka for different prescribed accuracies of the FMM ϵ according to Eq. (18) (δ=2) (solid lines). The dashed lines show the high-frequency asymptotics $p_{\text{hi}}(ka)$. The circles mark the points of switch from function representation via multipole expansions to samples of the far-field signature function and, respectively, the translation method used. The dash-dotted line separates the (ka,p) region into the domains where different function representations are used.

due to flat panel discretization). In this case the efficiency of the matrix-based methods, such as the RCR-decomposition, has comparable or better efficiency. Indeed, function representations via the samples of the far-field signature functions are at least two times larger, which results in larger memory consumption and reduction of efficiency of operations on larger representing vectors.

4. Multipole expansions

Expansions over the singular (radiating) spherical basis functions $S_n^m(\mathbf{r})$ in forms (15) and (16) can be applied to represent the monopole source or respective integrals. In these formulas the singular and regular solutions of the Helmholtz equation are defined as

$$S_n^m(\mathbf{r}) = h_n(kr)Y_n^m(\theta,\varphi), \quad R_n^m(\mathbf{r}) = j_n(kr)Y_n^m(\theta,\varphi),$$
$$n = 0,1,2,\ldots, \quad m = -n,\ldots,n, \tag{19}$$

where in spherical coordinates $\mathbf{r}=r(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$ symbols $h_n(kr)$ and $j_n(kr)$ denote spherical Hankel (first kind) and Bessel functions, and $Y_n^m(\theta,\varphi)$ the spherical harmonics,

$$Y_n^m(\theta,\varphi) = (-1)^m\sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}P_n^{|m|}(\cos\theta)e^{im\varphi},$$
$$n = 0,1,2,\ldots, \quad m = -n,\ldots,n, \tag{20}$$

and $P_n^{|m|}(\mu)$ are the associated Legendre functions consistent with that in Ref. 26, or Rodrigues' formulas

$$P_n^m(\mu) = (-1)^m(1-\mu^2)^{m/2}\frac{d^m}{d\mu^m}P_n(\mu), \quad n \geqslant 0, \quad m \geqslant 0,$$
$$P_n(\mu) = \frac{1}{2^n n!}\frac{d^n}{d\mu^n}(\mu^2-1)^n, \quad n \geqslant 0, \tag{21}$$

where $P_n(\mu)$ are the Legendre polynomials.

In the boundary integral formulation also normal derivatives of Green's function should be expanded (or integrals of these functions over the boundary elements). These expansions can be obtained from expansions of the type of Eqs. (15) and (16) for the monopoles by applying appropriately truncated differential operators in the space of the expansion coefficients,[12] which are sparse matrices and so the cost of differentiation is $O(p^2)$. Indeed if $\{C_n^m\}$ are the expansion coefficients of some function $F(\mathbf{r})$ over basis $S_n^m(\mathbf{r})$, while $\{\hat{C}_n^m\}$ are the expansion coefficients over the same basis of function $\mathbf{n}\cdot\nabla F(\mathbf{r})$ for unit normal $\mathbf{n}=(n_x,n_y,n_z)$, then

$$\hat{C}_n^m = \frac{1}{2}[(n_x+in_y)(b_n^m C_{n-1}^{m+1} - b_{n+1}^{-m-1}C_{n+1}^{m+1}) + (n_x-in_y)(b_n^{-m}C_{n-1}^{m-1}$$
$$- b_{n+1}^{m-1}C_{n+1}^{m-1})] + n_z(a_n^m C_{n+1}^m - a_{n-1}^m C_{n-1}^m),$$
$$m = 0, \pm 1, \pm 2, \ldots, \quad n = |m|, |m|+1, \ldots, \tag{22}$$

where a_n^m and b_n^m are the differentiation coefficients,

$$a_n^m = a_n^{-m} = \sqrt{\frac{(n+1+m)(n+1-m)}{(2n+1)(2n+3)}} \quad \text{for } n \geqslant |m|,$$
$$a_n^m = b_n^m = 0 \quad \text{for } n < |m|,$$
$$b_n^m = \sqrt{\frac{(n-m-1)(n-m)}{(2n-1)(2n+1)}} \quad \text{for } 0 \leqslant m \leqslant n, \tag{23}$$
$$b_n^m = -\sqrt{\frac{(n-m-1)(n-m)}{(2n-1)(2n+1)}} \quad \text{for } -n \leqslant m < 0.$$

5. Translations

Translations of the expansions can be also thought of as applications of matrices to the vectors of coefficients. If translation occurs from level l to l' ($l'=l-1$ for the multipole to multipole, or $S|S$-translation, $l'=l$ for the multipole to local, or $S|R$-translation, and $l'=l+1$ for the local to local, or $R|R$-translation), then $p_{l'}^2$ translated coefficients relate to the p_l^2 original coefficient via the $p_{l'}^2 \times p_l^2$ matrix. Even for precomputed and stored matrices, this requires $O(p^4)$ operations, which is unallowable cost for the translation if using with BEMs.[12] Several methods to reduce this cost are well known. Particularly use of the RCR-decomposition of the $(\mathbf{S}|\mathbf{S})(\mathbf{t})=(\mathbf{R}|\mathbf{R})(\mathbf{t})$ matrices

$$(\mathbf{R}|\mathbf{R})(\mathbf{t}) = \text{Rot}^{-1}(\mathbf{t}/t)\underline{(\mathbf{R}|\mathbf{R})}(t)\text{Rot}(\mathbf{t}/t), \tag{24}$$

where $\mathbf{t}$ is the translation vector, $t=|\mathbf{t}|$, and $\text{Rot}(\mathbf{t}/t)$ is the rotation matrix—which expresses coefficients in the rotated reference frame, whose z-axis is collinear with $\mathbf{t}$, while $\underline{(\mathbf{R}|\mathbf{R})}(t)$ is the coaxial translation operator (along axis z)—reduces the cost of application of all operators to $O(p^3)$. As the geometry of the problem is specified, all these matrices can be precomputed for a cost of $O(p^3)$ operations using recursions[13,12] and can be stored. We note also that the rect-

angular truncation operators $\mathrm{Rot}(\mathbf{t}/t)$ and $\mathrm{Rot}^{-1}(\mathbf{t}/t)$ act on the vectors of length p_l^2 and $p_{l'}^2$, respectively, and produce vectors of the same size, while $(\mathbf{R}|\mathbf{R})(t)$ acts on a vector of size p_l^2 and produces a vector of size $p_{l'}^2$. Therefore, there is no need for any interpolation or filtering as this is embedded into the decomposition. A similar decomposition is applied to the $(\mathbf{S}|\mathbf{R})(t)$ matrix for low frequencies, which provides a numerically stable low-frequency procedure [for levels corresponding to $ka_l < (kD)_*$].

For levels with $ka_l \geqslant (kD)_*$, we use the following decomposition of the translation matrix $(\mathbf{S}|\mathbf{R})(\mathbf{t})$:

$$(\mathbf{S}|\mathbf{R})(\mathbf{t}) = \mathbf{Sp}^{-1}\mathbf{\Lambda}_s(\mathbf{t})\mathbf{Sp}, \tag{25}$$

where $\mathbf{Sp}$ can be thought of as a matrix of size $N_l \times p_l^2$, which performs transform of the expansion coefficients to N_l samples of the far-field signature function (spherical transform), $\mathbf{\Lambda}_s(\mathbf{t})$ is a diagonal translation matrix of size $N_l \times N_l$, and $\mathbf{Sp}^{-1}$ is a matrix of size $p_l^2 \times N_l$, which provides a transform back to the space of the coefficients. The number of samples depends on the truncation number, and it is sufficient to use $N_l = (2p_l - 1)(4p_l - 3)$, where the grid is a Cartesian product of the $2p_l - 1$ Gauss quadrature abscissas with respect to the elevation angle $-1 \leqslant \mu = \cos\theta \leqslant 1$ and $4p_l - 3$ equispaced abscissas with respect to the azimuthal angle $0 \leqslant \varphi < 2\pi$. This grid also can be interpreted as a set of points on the unit sphere $\{\mathbf{s}_j\}$. The entries of the diagonal matrix $\mathbf{\Lambda}_s(\mathbf{t})$ are

$$\mathbf{\Lambda}_{jj}(\mathbf{t}) = \sum_{n=0}^{2p_l-2} i^n(2n+1)h_n(kt)P_n\left(\frac{\mathbf{s}_j \cdot \mathbf{t}}{t}\right), \quad j = 1, \ldots, N_l, \tag{26}$$

which is a diagonal form of the translation operator.[16] The bandwidth of this function, $2p_l - 2$, provides that decomposition (25) of the $p_l^2 \times p_l^2$ translation matrix $(\mathbf{S}|\mathbf{R})(\mathbf{t})$ is exact.[12] Note that for a given grid (which is the same for all translations at level l), the cost of computation of $\mathbf{\Lambda}_s(\mathbf{t})$ for each translation vector $\mathbf{t}$ is $O(p_l^3)$. In the present implementation, all these entries are precomputed and stored, so no computations of $\mathbf{\Lambda}_s(\mathbf{t})$ are needed during the run part of the algorithm. The precomputation part may be sped up by employing a data structure, which eliminates computations of $\mathbf{\Lambda}_{jj}(\mathbf{t})$ for repeated entries $\mathbf{s}_j \cdot \mathbf{t}/t$ and kt for all translations and, in fact, allows a substantial reduction in the cost of the present part of the algorithm.

The operator $\mathbf{Sp}$ can be decomposed into the Legendre transform with respect to $\mu = \cos\theta$ followed by the Fourier transform with respect to φ (e.g., see Refs. 12, 17, and 27). If performed in a straightforward way, each of them requires $O(p^3)$ operations. Despite the fact that there exist algorithms for fast Legendre transform and the fast Fourier transform (FFT) can be employed, which reduces the cost of application of operator $\mathbf{Sp}$ to $O(p^2 \log p)$ or so, for moderate p straightforward methods are much more efficient. Note that the major cost (about 90%) comes from the Fourier transform, so if the FFT is applied efficiently, this speeds up the procedure. Furthermore, the operator $\mathbf{Sp}^{-1}$ can be decomposed into the inverse Fourier transform, diagonal matrix of the Legendre weights, and inverse Legendre transform. The cost of this procedure is the same as that for the computation of the forward transform.

As mentioned earlier, since the same transforms $\mathbf{Sp}$ and $\mathbf{Sp}^{-1}$ should be applied to all expansions at a given level, we can make Eq. (25) more efficient than the RCR-decomposition by first applying the transform $\mathbf{Sp}$ to all box expansions at a given level, then performing all diagonal translations and consolidations, and, finally, applying transform $\mathbf{Sp}^{-1}$ to all boxes.

6. Evaluation of expansions

Finally we mention that for the computation of the BEM operators L' and M', the normal derivative of computed sums at the evaluation point should be taken. As the expansions are available for the sources outside the neighborhood of the evaluation points, this can be performed by the application of the differentiation operator in the coefficient space [see Eq. (22)].

7. Simultaneous matrix-vector products

For an efficient iterative solution of Eq. (10), the FMM can be used to compute in one run the sum of four matrix-vector products together,

$$\Sigma + \lambda\Sigma' = (L_{\mathrm{dense}} + \lambda L'_{\mathrm{dense}})[q] - (M_{\mathrm{dense}} + \lambda M'_{\mathrm{dense}})[\phi], \tag{27}$$

for input vectors q and ϕ. Also, if needed, results for the parts Σ and Σ' can be separated (e.g., for the application of Green's identity alone for the computation of the potential at internal domain points). The dense parts of the matrices correspond to decomposition (14) and, in the case of use of a simple scheme [Eq. (17)] are the matrices with eliminated diagonals.

V. COMPUTATION OF SINGULAR ELEMENTS

Despite the fact that there exist techniques for the computation of the integrals over the singular or nearly singular elements (e.g., with increasing number of nodes and element partitioning or using analytical or semianalytical formulas), these methods can be costly, and below we propose a technique for the approximation of such integrals, which is consistent with the use of the FMM and BEM. This technique is similar to the "simple solution" technique used by some authors to compute the diagonal elements for the BEM for potential problems and for elasticity,[28] except that it is updated with the use of the FMM, and to the case of the Helmholtz equation.

Let $\{\mathbf{x}_j\}$ be a set of points sampling the surface, and U_j^ϵ be a sphere of radius ϵ centered at $\mathbf{x}_j$ and $S_j^\epsilon = S \cap U_j^\epsilon$. The surface operators can be decomposed as

$$L[\sigma] = \int_{S_j^\epsilon} \sigma(\mathbf{x}) G(\mathbf{x},\mathbf{y}) dS(\mathbf{x}) + \int_{S\setminus S_j^\epsilon(\mathbf{y},\epsilon)} \sigma(\mathbf{x}) G(\mathbf{x},\mathbf{y}) dS(\mathbf{x}) = L_j^\epsilon[\sigma] + \tilde{L}_j[\sigma],$$

$$M[\sigma] = \int_{S_j^\epsilon} \sigma(\mathbf{x}) \frac{\partial G(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{x})} dS(\mathbf{x}) + \int_{S\setminus S_j^\epsilon(\mathbf{y},\epsilon)} \sigma(\mathbf{x}) \frac{\partial G(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{x})} dS(\mathbf{x}) = M_j^\epsilon[\sigma] + \tilde{M}_j[\sigma]. \tag{28}$$

Note that for small enough ϵ we have the following approximations of the integrals:

$$L_j^\epsilon[\sigma] \approx \sigma_j l_j^\epsilon(\mathbf{y}), \quad M_j^\epsilon[\sigma] \approx \sigma_j m_j^\epsilon(\mathbf{y}), \tag{29}$$

where functions $l_j^\epsilon(\mathbf{y})$ and $m_j^\epsilon(\mathbf{y})$ are regular inside the domain. Thus, they can be approximated by a set of some basis functions, which satisfy the same Helmholtz equation. To construct such a set and approximation, consider Green's identity for a function, which is regular inside the finite domain (internal problem),

$$\gamma\psi = L\left[\frac{\partial\psi}{\partial n}\right] - M[\psi], \tag{30}$$

where $\gamma=1$ for points inside the domain, $\gamma=1/2$ for points on the boundary, and $\gamma=0$ for points outside the domain. Consider the following test functions:

$$\psi(\mathbf{x}) = e^{ik\mathbf{s}\cdot\mathbf{x}}, \qquad q(\mathbf{x}) = \frac{\partial\psi}{\partial n}(\mathbf{x}) = \mathbf{n}(\mathbf{x})\cdot\nabla e^{ik\mathbf{s}\cdot\mathbf{x}} = ik\mathbf{n}(\mathbf{x})\cdot\mathbf{s}e^{ik\mathbf{s}\cdot\mathbf{x}}, \quad |\mathbf{s}| = 1, \tag{31}$$

which represent plane waves propagating in any direction $\mathbf{s}$. For these functions from Eqs. (28)–(30), one obtains

$$m_j^\epsilon(\mathbf{y}) - ik(\mathbf{n}_j\cdot\mathbf{s}) l_j^\epsilon(\mathbf{y}) = e^{-ik\mathbf{s}\cdot\mathbf{x}_j}\{ik\tilde{L}[(\mathbf{n}\cdot\mathbf{s})e^{ik\mathbf{s}\cdot\mathbf{x}}] - \tilde{M}[e^{ik\mathbf{s}\cdot\mathbf{x}}] - \gamma(\mathbf{y})e^{ik\mathbf{s}\cdot\mathbf{y}}\}. \tag{32}$$

Let $\mathbf{s}_1,\ldots,\mathbf{s}_4$ be four different unit vectors provided that functions $e^{ik\mathbf{s}_\alpha\cdot\mathbf{x}}$, $\alpha=1,\ldots,4$, are linearly independent. Then denoting

$$\omega_{j\alpha}(\mathbf{y}) = e^{ik\mathbf{s}\cdot\mathbf{x}_j}\{ik\tilde{L}[(\mathbf{n}\cdot\mathbf{s}_\alpha)e^{ik\mathbf{s}_\alpha\cdot\mathbf{x}}] - \tilde{M}[e^{ik\mathbf{s}_\alpha\cdot\mathbf{x}}] - \gamma(\mathbf{y})e^{ik\mathbf{s}_\alpha\cdot\mathbf{y}}\}, \quad n_{j\alpha} = \mathbf{n}_j\cdot\mathbf{s}_\alpha, \tag{33}$$

one obtains

$$m_j^\epsilon(\mathbf{y}) - ikn_{j\alpha} l_j^\epsilon(\mathbf{y}) = \omega_{j\alpha}(\mathbf{y}), \quad \alpha = 1,\ldots,4. \tag{34}$$

Obviously the surface operators $L'[\sigma]$ and $M'[\sigma]$ can be similarly decomposed as

$$L'[\sigma] = L_j'^\epsilon[\sigma] + \tilde{L}_j'[\sigma], \quad M'[\sigma] = M_j'^\epsilon[\sigma] + \tilde{M}_j'[\sigma], \qquad L_j'^\epsilon[\sigma] \approx \sigma_j l_j'^\epsilon(\mathbf{y}), \quad M_j'^\epsilon[\sigma] \approx \sigma_j m_j'^\epsilon(\mathbf{y}). \tag{35}$$

Note that these operators are employed only for points on the boundary, and the forms for internal and external problems are either the same or related via a sign change. Thus one can use Maue's identity (8) for the internal problem. In this case using test functions (31), Eqs. (32)–(34) can be modified as follows:

$$m_j'^\epsilon(\mathbf{y}) - ikn_{j\alpha} l_j'^\epsilon(\mathbf{y}) = \omega_{j\alpha}'(\mathbf{y}), \quad \alpha = 1,\ldots,4, \qquad \omega_{j\alpha}'(\mathbf{y}) = e^{-ik\mathbf{s}\cdot\mathbf{x}_j}\{ik\tilde{L}'[(\mathbf{n}\cdot\mathbf{s}_\alpha)e^{ik\mathbf{s}_\alpha\cdot\mathbf{x}}] - \tilde{M}'[e^{ik\mathbf{s}_\alpha\cdot\mathbf{x}}] - \gamma_\alpha(\mathbf{y})e^{ik\mathbf{s}_\alpha\cdot\mathbf{y}}\}, \qquad \gamma_\alpha(\mathbf{y}) = \tfrac{1}{2}ik(\mathbf{n}(\mathbf{y})\cdot\mathbf{s}_\alpha), \quad \mathbf{y}\in S. \tag{36}$$

Then for the chosen set of directions, solution will be provided by the previous equations with $m_j^\epsilon(\mathbf{y})$, $l_j^\epsilon(\mathbf{y})$, and $\omega_{j\alpha}(\mathbf{y})$, replaced with $m_j'^\epsilon(\mathbf{y})$, $l_j'^\epsilon(\mathbf{y})$, and $\omega_{j\alpha}'(\mathbf{y})$, respectively.

The small 4×2 linear systems for each [Eqs. (34) and (36)] can be solved via least squares, and as noted above the FMM provides four simultaneous matrix-vector multiplications, and so four matrix-vector products via the FMM ($\alpha=1,\ldots,4$) are sufficient to get all diagonals.

A. Discretization

The above formulas obviously provide expressions for the diagonal elements of matrices

$$L_{jj} = l_j^\epsilon(\mathbf{x}_j), \quad M_{jj} = m_j^\epsilon(\mathbf{x}_j), \quad L_{jj}' = l_j'^\epsilon(\mathbf{x}_j), \qquad M_{jj}' = m_j'^\epsilon(\mathbf{x}_j), \quad j = 1,\ldots,N. \tag{37}$$

In fact, for the solution of the boundary integral equation, only quantities $L_{jj}+\lambda L_{jj}'$ and $M_{jj}+\lambda M_{jj}'$ are needed. So for given λ the storage can be reduced twice. Also combinations $L_{jj}+\lambda L_{jj}'$ and $M_{jj}+\lambda M_{jj}'$ can be computed instead using the same method as described above.

VI. NUMERICAL EXPERIMENTS

The BEM/FMM was implemented in FORTRAN 95 and was parallelized for symmetric multiprocessing architectures, such as modern multicore PCs, using OpenMP. The results reported below were obtained on a four core PC [Intel Core 2 Extreme QX6700 2.66 GHz processor with 2 $\times$4 Mbytes L2 Cache and 8 Gbytes random access memory (RAM)] running Windows XP-64. Parallelization requires the replication of data among threads and is controlled by the size of the cache. In this implementation, to reduce the size of the stack only the sparse matrix computations and the $S|R$-translations using the RCR-decomposition (i.e., operators employed for levels finer than the high-low-frequency switch level) were parallelized. Such a parallelization strategy was found to be efficient enough (overall parallelization efficiency was about 80%–95%) and enabled computations for kD's up to 500. For $kD<200$ more data can be placed on the available stack, and a complete parallelization with effi-

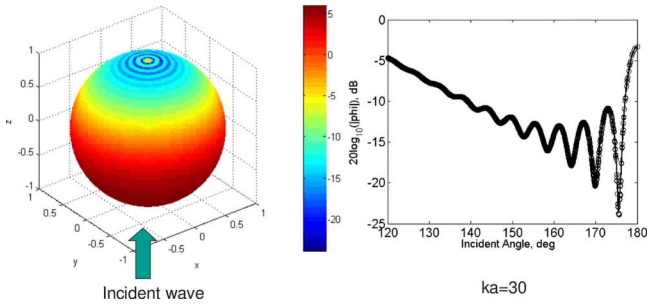


FIG. 3. (Color online) Typical BEM computations of the scattering problem. The graph shows comparison between the analytical solution (38) and BEM solution for the vicinity of the rear point of the sphere for $ka=30$.

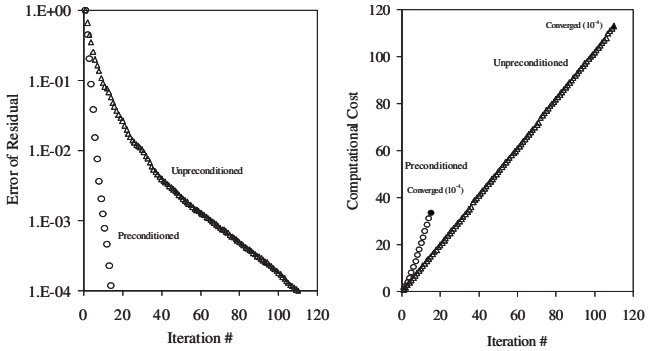


FIG. 4. Left: The absolute error in the residual in the unpreconditioned GMRES (triangles) and in the preconditioned fGMRES (circles) as a function of the number of iterations (outer loop for the fGMRES). Right: the relative computational cost to achieve the same error in the residual for these methods (1 cost unit=1 iteration using the unpreconditioned method). Computations for sphere; $ka=50$ for mesh with 101 402 vertices and 202 808 elements, and $\lambda=6\times10^{-4}i$.

ciency close to 100% was obtained. However, the results reported here use the partial parallelization to enable a uniform comparison of the results obtained over a wide range of kD's.

A. Scattering from a single sphere

The example of an incident plane wave off a single sphere is valuable for tests of the performance of the method since an analytical solution is available in this case. For the incident field $\phi^{\text{in}}(\mathbf{r})=e^{ik\mathbf{s}\cdot\mathbf{r}}$, where $\mathbf{s}$ is the unit vector collinear with the wave vector, the solution for the total (incident plus scattered field) for impedance boundary conditions can be found elsewhere, e.g., in Ref. 12,

$$\phi|_S(\mathbf{s}')=\frac{i}{(ka)^2}\sum_{n=0}^{\infty}\frac{(2n+1)i^nP_n(\mathbf{s}\cdot\mathbf{s}')}{h_n'(ka)+(i\sigma/k)h_n(ka)},\quad \frac{\partial\phi}{\partial n}+i\sigma\phi\Big|_S=0, \tag{38}$$

where a is the sphere radius, $\mathbf{s}'$ is a unit vector pointing to the location of the evaluation surface point, and σ is the boundary admittance, which is zero for sound-hard surfaces and infinity for sound-soft surfaces. Depending on this, one may have for the scattered field a Neumann, Dirichlet, or Robin problem.

The following were varied in the simulations: k, discretization, parameter λ in Eq. (10), the boundary admittance, tolerance, and parameters controlling the FMM accuracy and performance. A typical computational result is shown in Fig. 3.

1. Preconditioning

The best preconditioners found were right preconditioners that compute the solution of the system $A\psi_j=c_j$ at the jth step using unpreconditioned GMRES (inner loop) and low accuracy FMM with lower bounds for convergence of iterations. For example, if in the outer loop of the fGMRES the prescribed accuracy for the FMM solution was 10^{-4} (actual achieved accuracy was $\sim10^{-6}$) and the iterative process was terminated when the residual reaches 10^{-4}, for the inner preconditioning loop we used FMM with prescribed accuracy of 0.2 (actual accuracy was $\sim5\times10^{-3}$) and the iterative process was terminated at a residual value of 0.45. The inner loop solution was much faster than the outer loop, as it is stopped after fewer iterations, and the matrix-vector product was several times faster (lower truncation numbers). The preconditioner reduced by an order of magnitude the number of iterations in the outer loop, more than compensating for the increased cost of inner iterations. Beyond the benefit of improved time, this also has the benefit of reduced memory for large problems. This is because in the GMRES or fGMRES K vectors of length N must be stored, where K is the dimensionality of the Krylov subspace. The iterative process becomes slower if K is restricted, and GMRES is restarted when K vectors have been stored. Hence it is preferable to achieve convergence in $N_{\text{iter}}\leq K$ iterations. In the case of the GMRES-based preconditioner, the storage memory is of order $(K+K')N$, where K' is the dimensionality of the Krylov subspace for preconditioning. Since both K and K' are much smaller than the K required for unpreconditioned GMRES, the required memory for solution is reduced substantially.

Figure 4 shows the convergence of the unpreconditioned GMRES and the preconditioned fGMRES with the FMM-based preconditioner as described above. The computations were made for a sound-hard sphere of radius a, whose surface was discretized by 101 402 vertices and 202 808 triangular elements for the relative wavenumber $ka=50$. The preconditioned iteration converges three times faster in terms of the time required. The matrix-vector product via the low accuracy FMM in preconditioning was approximately seven times faster than that computed with higher accuracy in the outer loop (1.4 s versus 9.75 s). The overall error of the obtained solution was slightly below 5×10^{-4} in the relative L_2-norm.

2. Spurious modes

The test case for the sphere is also good to illustrate the advantage of the Burton–Miller formulation over the formulation based on Green's identity. According to theory,[20,29] the Green's identity formulation may result in convergence to a solution, which is the true solution plus a nonzero solution of the internal problem corresponding to zero boundary conditions at the given wavenumber. Such solutions are not physical since the solution of the external scattering problem is

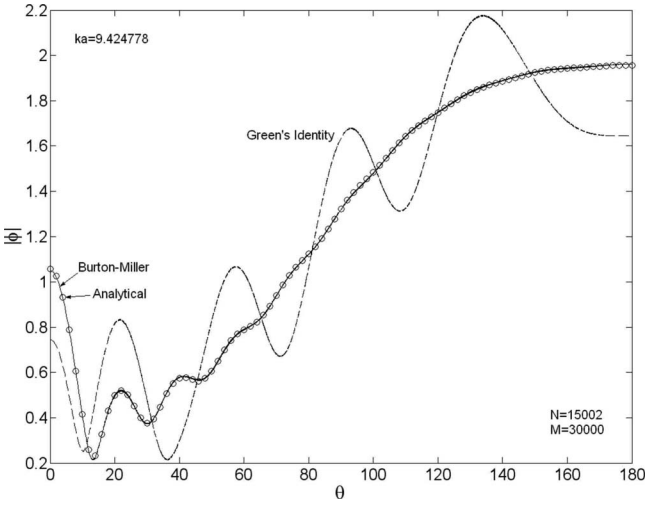


FIG. 5. Solutions of the plane wave scattering problem obtained using BEM with the Burton–Miller and Green's identity formulations and fGMRES iterator for a sphere at resonance $ka=9.424\,778$ (triangular mesh of 15 002 vertices and 30 000 elements). Analytical solution is shown by the circles.

unique, and, therefore, they manifest a deficiency of the numerical method based on the Green's identity.

For a sphere any internal solution can be written in the form

$$\phi^{\text{int}}(\mathbf{r}) = \sum_{n=0}^{\infty} j_n(kr) \sum_{m=-n}^{n} B_n^m Y_n^m(\theta,\varphi), \tag{39}$$

where (r,θ,φ) are the spherical coordinates of $\mathbf{r}$ and B_n^m are arbitrary constants. The set of zeros of the functions $j_n(ka)$ provides a discrete set of values of ka for which $\phi^{\text{int}}|_S=0$, even though ϕ^{int} is not identically zero inside the sphere. The minimum resonant k can be obtained from the first zero of $j_0(ka)$, which is $ka=\pi$. We conducted some numerical tests with Burton–Miller and Green's formulations for a range of ka $(0.01 \leqslant ka \leqslant 50)$ for different resonant values of ka.

Figure 5 provides an illustration for case $ka=3\pi \approx 9.424\,778$, which is the third zero of $j_0(ka)$. When the Burton–Miller formulation is used with some λ with $\text{Im}(\lambda) \neq 0$, a solution consistent with Eq. (38) is obtained, while when computations are performed using Green's formula a wrong solution with the additional spurious modes is achieved. We checked that in this case the converged solution can be well approximated $(\sigma=0)$ by

$$\phi|_S(\mathbf{s}') = B + \frac{i}{(ka)^2} \sum_{n=1}^{\infty} \frac{(2n+1) i^n P_n(\mathbf{s}\cdot\mathbf{s}')}{h_n'(ka)}, \tag{40}$$

where B is some complex constant depending on the initial guess in the iterative process. This shows that, in fact, the zero-order harmonic of the solution, corresponding to the resonating eigenfunction, failed to be determined correctly, which is the expected result. Note that such solutions appear when iterative methods are used, where degeneration of the matrix operator for some subspace does not affect convergence in the other subspaces. If the problem is solved directly, the degeneracy or poor conditioning of the system matrix would result in a completely wrong solution.

In the nonresonant cases, the Green's formulation provided a good solution, while normally the number of iterations was observed to increase as ka comes closer to a resonance. For large ka (40 or more), computations using Green's identity become unstable, which is explainable by density of the zeros of $j_n(ka)$. For the Burton–Miller formulation, the empirically found value of the parameter λ was

$$\lambda = \frac{i\eta}{k}, \quad \eta \in [0.01, 1]. \tag{41}$$

For the case above $\eta=0.03$ provides good results, while for $ka<1$ the number of iterations increases compared to the Green's identity, and when there are certainly no resonances, Green's identity can be recommended. Increase in this parameter usually decreases the accuracy of computations since more weight is put on the hypersingular part of the integral equation, while decrease in the parameter for large ka leads to an increase in the number of iterations. For $\eta \ll 0.01$ the Burton–Miller integral equation shows the problems of spurious modes.

3. Performance

For the FMM, the characteristic scale is usually based on the diagonal (maximum size) of the computational domain, D, and kD is an important dimensionless parameter. Further, one can compute the maximum size of the boundary element, which for a triangular mesh is the maximum side of the triangle, d. This produces another dimensionless parameter, kd. For a fixed body of surface area $S \sim D^2$, the number of elements in the mesh is of the order $N \sim S/d^2 \sim D^2/d^2$. A formal constraint for discretizations used for an accurate solution of the Helmholtz equation is $d/\lambda_a \ll 1$, where $\lambda_a = 2\pi/k$ is the acoustic wavelength. In practice this condition can be replaced with $kd < \chi$, where χ is some constant of order 1, so there are not less than $2\pi/\chi$ mesh elements per wavelength. This number usually varies in range of 5–10. This shows that the total number of elements should be at least $N \sim D^2/d^2 \gtrsim (kD)^2/\chi^2 = O((kD)^2)$.

Figure 6 shows the results of numerical experiments for scattering from a sound-hard sphere, where we fixed parameter $\chi \approx 0.94$ for $kD \geqslant 20$, which provided at least six elements per wavelength and increased the number of elements proportionally to $(kD)^2$ (the maximum case plotted corresponds to $kD=500$ where the mesh contains 1 500 002 vertices and 3 000 000 elements). Additional data on some tests showing the relative error in the L_2-norm, ϵ_2, and peak memory required for the problem solution for the current implementation are given in Table I. Note that in all cases the relative error in the L_∞-norm was less or about 1%. For $kD < 3$, which we characterize as very low-frequency regime, we used a constant mesh with 866 vertices and 1728 elements. Also for this range λ was set to zero, while for other cases λ was computed using Eq. (41) with $\eta=0.03$. The time and memory required for computations substantially depend on the accuracy of the FMM, tolerance for convergence, and settings for the preconditioner. The prescribed FMM error for the outer loop was 10^{-4}, which, in fact, provides accuracy about 10^{-6} in the L_2-norm. This was found from the tests,

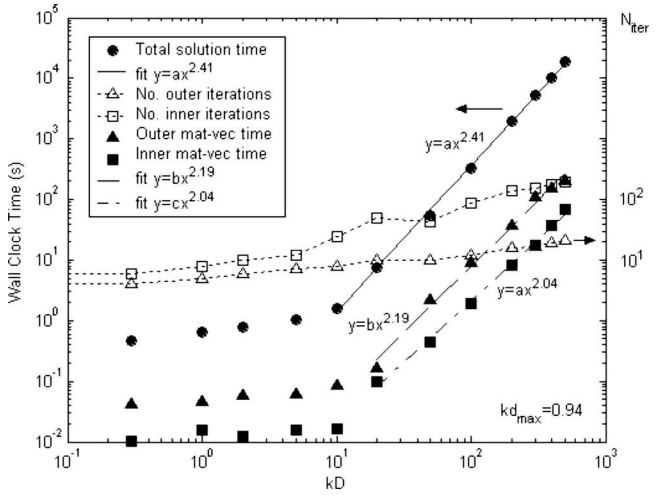


FIG. 6. Wall clock time and the number of iterations for solution of scattering problem from a sound-hard sphere in range $0.1 \leqslant kD \leqslant 500$. The curves that fit the time to the data at high frequencies were obtained using least squares.

where the actual FMM error was computed by comparisons with the straightforward ("exact") method at 100 check points. The prescribed accuracy for the preconditioner was 4.5×10^{-2}, which also substantially overestimates the actual accuracy. Tolerance for convergence for the outer loop was 10^{-4}, while that for the inner loop was 0.45, where the number of iterations was limited by 11. The memory required depends on the dimensionalities of the Krylov subspaces for the outer and inner loops, which were set to 35 and 11, respectively, while the convergence for all cases was within 20 outer iterations, so the memory, in fact, can be reduced. Additional acceleration can be obtained almost for all cases by the storage of the near-field data [matrix A_{sparse}, Eq. (14)]. However, based on the RAM (8 Gbytes) this works only for $N < 5 \times 10^5$. To show a scaling without jumps, such a storage was not used, and the sparse matrix entries were recomputed each time as the respective matrix-product was needed.

a. Computational complexity scaling. The total time required for the solution of the problem is scaled approximately as $O((kD)^{2.4})$, while the times for the matrix-vector multiplication in the outer and in the inner iteration loops of the preconditioned fGMRES scale, respectively, as $O((kD)^{2.2})$ and $O((kD)^{2.04})$. A standard least squares fit for the data at $kD > 10$ was used.

These results require some analysis and explanation since the theoretical expectation of the matrix-vector complexity for the current version of the FMM is $O((kD)^3)$. There are several different reasons that the algorithm is scaled better. First, we note that the complexity of the sparse matrix-vector product, which in a well-balanced algorithm takes a substantial part of the computational time, is scaled as $O(N)$, i.e., as $O((kD)^2)$. Second, for high-frequency computations a switch between the high-frequency and low-frequency representations always happens. As soon as the low-frequency part is limited by condition $ka_l < (kD)_*$, where $(kD)_*$ depends on the prescribed accuracy only, the computational cost for this part becomes proportional to the number of occupied boxes in the low-frequency regime. This number is $O(B_{l_{\max}})$, where $B_{l_{\max}}$ is the number of occupied boxes at the finest level $l_{\max}$, which for simple shapes is $O(N) = O((kD)^2)$. Therefore, the low-frequency part for the fixed prescribed accuracy can be executed in $O((kD)^2)$ time. Therefore, all the $O((kD)^3)$ scaling comes only from the high-frequency part. However, here the most expensive part related to the $S|R$-translations is performed using the diagonal forms with complexity $O((kD)^2)$, while the asymptotic constant related to the $S|S$- and $R|R$-translations and spherical transform is relatively small. Profiling of the algorithm shows that sparse matrix-vector multiplication and $S|R$-translations in the low-frequency part contribute up to 90% to the complexity in the range of kD's studied. Adding all these facts together, we can see that the FMM for problems solved is rather the $O((kD)^2)$ algorithm, and small addition to power 2 is due to $O((kD)^3)$ operations with much smaller asymptotic constant.

According to the above discussion, the complexity of the FMM can be estimated as

$$\text{Time} = A(kD)^3 + B(kD)^2, \tag{42}$$

where A and B are some constants. The least squares fit of high-frequency data on the matrix-vector multiplication in the outer loop with Eq. (42) is as good as fit $O((kD)^{2.19})$, shown in Fig. 6, and provides $B/A = 1.39 \times 10^3$. This number, indeed, is large, and it actually shows that only at $kD \gg 10^3$ the high-frequency asymptotical behavior can be reached for the prescribed accuracy of 10^{-4}. For the low accuracy preconditioner used, this ratio becomes even smaller ($B/A = 1.75 \times 10^3$), which justifies its applicability even for higher kD's.

The overall algorithm scaling as $O((kD)^{2.4})$ now is not surprising since the factor $O((kD)^{0.2})$ can be ascribed to the (slowly) increasing number of iterations. Indeed, Fig. 6 shows that the number of iterations in the outer loop grows slowly.

b. Memory complexity scaling. Theoretically, the memory required to solve the problem should be scaled as $O(N) = O((kD)^2)$. Due to different possibilities for memory management and optimizations, plus realization of trade-off option memory versus speed and limited resources, the actual scaling can deviate from the theoretical one. Table I

TABLE I. Error, memory, and FMM-BEM solution wall clock time for some test cases with a spherical scatterer with known analytical solution.

kD	0.0001	0.01	1	20	50	100	200	300	400	500
N	866	866	866	2402	1.5×10^4	6×10^4	2.4×10^5	5.4×10^5	9.6×10^5	1.5×10^6
ε_2	3.3×10^{-5}	3.3×10^{-5}	1.1×10^{-4}	5.4×10^{-3}	2.25×10^{-4}	4.1×10^{-4}	5.34×10^{-4}	3.6×10^{-4}	4.92×10^{-4}	3.2×10^{-3}
Memory, Mbyte	20	20	32	63	268	461	1144	2861	3755	4748
Time, s	0.438	0.406	0.656	7.59	53.6	327	1990	5150	10 400	19 100

TABLE II. Results from some recent FMM and FMM accelerated BEM publications operating in the high-frequency regime. The performance of the FMM and the FMM-BEM reported in this paper is comparable or superior (see Tables I and IV). Data not reported are indicated with a "–"; for the FMM only (Ref. 17) the values reported can be considered as those applying to one iteration of the solution procedure.

Ref.	kD	N	$N/(kD)^2$	Memory	Error	Time	Iteration	Remarks
10	435	1 046 528	5.53	4800	$>10^{-5}$	54 267	–	Scattering
9	126.64	98 304	6.13	1487	0.03	24 325	59	Internal
9	126.64	49 152	3.06	799.6	0.03	11 848	58	Internal
17	544	646 143	2.09	549	10^{-3}	672	n/a	FMM only
17	544	619 520	2.09	1111	10^{-6}	1832	n/a	FMM only

shows that in the implementation used for this paper, the memory required grows slower than $O((kD)^2)$ in the range studied. Certainly, it should be scaled not less than $O((kD)^2)$ at larger kD since at least several vectors of size N should be stored. The table provides some data, which show that computation of a million size problem with $kD \sim 500$ is reachable on desktop PCs or workstations.

c. Comparison with recent publications. As discussed previously two regimes can be recognized in FMM-BEM simulations, the low and high-frequency regimes. Tables II and III indicate some results from papers published in the past few years. The data here should be compared with those presented in Table I where data for the present code are given. Looking first at the high-frequency regime, there appear to be three relevant papers (Refs. 9, 10, and 17). Reference 17 only has a FMM matrix-vector product, and a detailed comparison of the present algorithm and the one of that paper was provided in Sec. IV. As mentioned previously, the size of the discretization N in a BEM simulation should be at least $O(kD)^2$, although more points may be necessary if the geometry is complex. In each case the comparison is quite favorable for the present code.

The results of the best performing FMM BEM solver[10] which appeared during the time when this paper was under review, is compared with the present work. In this paper the authors provide data on test cases for scattering from a sphere, so a comparison was possible. In this case a Robin problem (2) with $\alpha=1$ and $\beta=-1/(10-10i)$ was solved for a sphere of radius $a=20\lambda$ ($kD=435$). The incident field was generated by a source located at $(0,0,10a)$. Unfortunately there were no data on the iteration process used, and the only notice was that the process converged to the residual 10^{-5}. So the same parameters were used for the present test case. Further, a mesh containing 1 046 528 panels was used. Since this mesh was not available, a mesh of similar size that could be generated was used. This mesh contained 1 130 988 triangles and 565 496 vertices. Another mesh tried contained 2 096 688 panels and 1 048 346 vertices. In the cited paper Green's identity was used, while in the present study the combined Eq. (12) with λ from Eq. (41) is used. An increase in η from the value of 0.03 used for the solution of the Neumann problem to a value of 1.2 improved convergence for the Robin problem. The results are shown in Table IV.

It is seen that the present algorithm performed one order of magnitude faster, while the memory consumption was of the same order. Of course, the present algorithm was parallelized and executed on a four core machine, compared to the one core serial implementation of the Multilevel Fast Multipole Algorithm. However, even assigning a perfect parallelization factor of 4, we see that the same problem can be solved several times faster. The speed-up of the present algorithm can be related to several issues, one of which is the preconditioning and use of the combined equation instead of Green's identity. For the largest case reported in Table IV, the present iteration took 23 outer loop iterations, for which matrix-vector multiplication took about 61 s/iteration, while for the inner loop, which was in any case limited by 11 iterations, the matrix-vector product took about 27 s/iteration.

As far as low-frequency computation comparisons are concerned, comparisons were performed against Refs. 4, 7, and 9. Here again Table I shows that the present code has performance that is comparable or better. Here it is seen that some of the simulations reported in the literature have discretizations that are much finer than what may have been required by the problem.

B. More complex shapes

Many problems in acoustics require computations for substantially complex shapes, which include bioacoustics,

TABLE III. Results from some recent FMM and FMM accelerated BEM publications operating in the low-frequency regime. The performance of the FMM and the FMM-BEM reported in this paper is comparable or superior (see Table I). The parameter $N/(kD)^2$ represents how overdiscretized the problem is vis-à-vis the restrictions imposed by the frequency.

Ref.	kD	N	$N/(kD)^2$	Mem	Time	Iteration	Time/Iteration	Remarks
9	16	1536	6.13	4800	28	5	5.6	Internal room
9	16	98 304	6.13	1487	14	5	2.8	Internal room
14	0.9375	45 056	51 263	…	41 962	452	92.2	Internal L-shape
14	0.9375	704	801	…	25.1	21	1.19	Internal L-shape
7	3.464	200 000	16 668	…	…	8000	…	External scattering

TABLE IV. Comparison of solution times and memory at $kD=435$ reported in Ref. 10 and the present study.

	No. of elements	N (unknowns)	T (s)	M (Gbyte)
10	1 046 528	1 046 528	54 267	4.8
Present (case 1)	1 130 988	565 496	3 820	2.7
Present (case 2)	2 096 688	1 048 346	7 762	2.8

human hearing, sound propagation in dispersed media, engine acoustics, and room acoustics. Many such cases are reported in the literature: e.g., scattering off aircraft,[17] off animals,[17] off engine blocks,[7] and off many scattering ellipsoids.[7] The present algorithm was tested on many such cases to show that the iteration process is convergent for these cases and that solutions can be obtained on bodies with thin and narrow appendages. Figure 7 provides an idea on the sizes and geometries that were tested. Note that modeling of complex shapes requires surface discretizations, which are determined not only by the wavelength-based conditions $\chi \ll 2\pi$ but also by the requirements that the topology and some shape features should be properly represented. Indeed even for a solution of the Laplace equation ($k=0$), the boundary element methods can use thousands and millions of elements that just properly represent the geometry. Usually the same mesh is used for multifrequency analysis, in which case the number of elements is fixed and selected to satisfy criteria for the largest k required. In this case the number of elements per wavelength for small k can be large. Also, of course, discretization plays an important role in the accuracy of computations. So if some problem with complex geometry should be solved with high accuracy, then the number of elements per wavelength can be again large enough.

The last geometry illustrated in Fig. 7 is similar to the case studied in Ref. 7 for $kD=3.464$. Here the performance and accuracy were studied for a range of kD from 0.35 to 175 ($ka=0.01-5$, where a is the largest axis of an ellipsoid). The surface of each ellipsoid was discretized with more than 1000 vertices and 2000 elements to provide an acceptable accuracy of the method even for low frequencies. The convergence for the Neumann, Dirichlet, and Robin problems was very fast (just a few iterations) for small ka, where, as discussed above, the use of low-frequency FMM is necessary. The convergence was not affected by the increase in the number of nodes, although for accuracy better discretizations

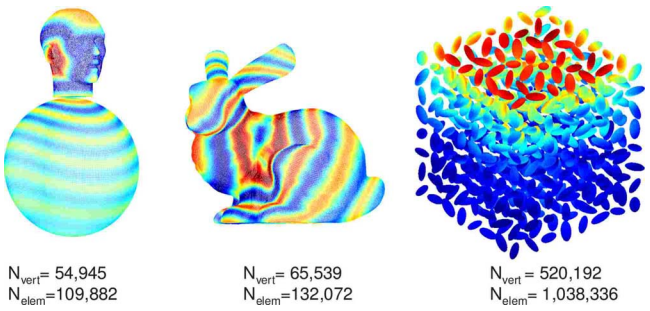


FIG. 7. (Color online) Examples of test problems solved with the present version of the BEM: human head-torso, bunny models (7.85 and 25 kHz acoustic sources located inside the objects, $kD=110$ and 96, respectively), and plane wave scattering by 512 randomly oriented ellipsoids ($kD=29$).

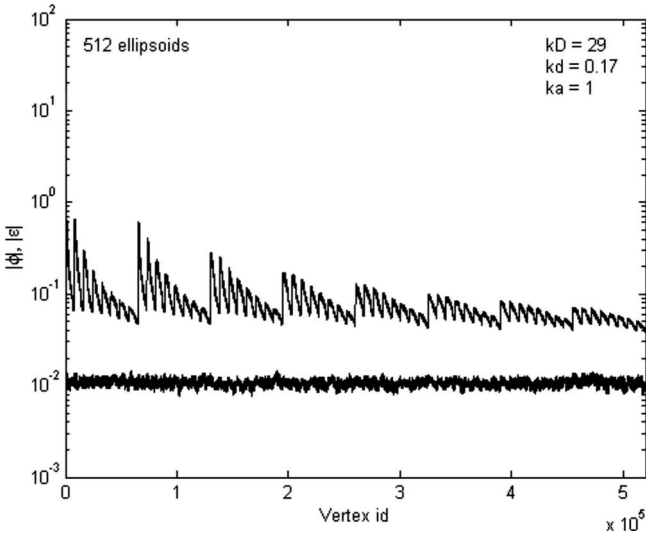


FIG. 8. Solution (the upper curve) and error (the lower curve) in the boundary condition at each vertex for the case of the ellipsoids in Fig. 7.

are preferred. As a test solution we used an analytical solution when a source was placed inside one of the ellipsoids, and surface values and normal derivatives were computed at each vertex location analytically. For larger discretizations that were used, we were able to achieve ~1% relative errors in strong norm (L_∞) for the range of parameters used.

Figure 8 shows an absolute relative error at each vertex,

$$\epsilon_i = \frac{|\phi_i^{(\mathrm{BEM})} - \phi_i^{(\mathrm{an})}|}{|\phi_i^{(\mathrm{an})}|}, \quad i = 1, \ldots, N_{\mathrm{vert}}, \tag{43}$$

where $\phi_i^{(\mathrm{an})}$ and $\phi_i^{(\mathrm{BEM})}$ are the analytical and BEM solutions, and $|\phi_i^{(\mathrm{an})}|$ is the modulus of the solution. The maximum error here was $\max(\epsilon_i)=1.58\%$, which is usually acceptable for physics-based problems and engineering computations.

C. Computational challenges and drawbacks

Despite the fact that the present algorithm can run cases in a wide range of kD up to 500, the numerical stability for larger kD is still a question. This mostly relates to the computation of matrix $S|S$- and $R|R$-translation operators at higher frequencies via the RCR-decomposition. Particularly we should use different recursions for the computation of entries of the rotation operators from those that were acceptable for lower truncation numbers than those described in our earlier paper[13]. Such recursions are available (see p. 336 in Ref. 12) and stable if recursive backpropagation is used. An analysis shows that recursive computation of coaxial translation operators can be also unstable at larger frequencies, and this subject requires additional consideration, which is beyond the present paper.

In terms of practical use of the algorithm, it is still a research question: how to improve preconditioning. For example, the current preconditioner worked well for the external Neumann problem in automatic settings, while the solution of Robin or mixed problems showed a slow-down of the convergence, and there was a need for intervention and tuning of the η parameter and number of inner loop iterations

and accuracy to improve convergence. Further research on the convergence of the iterative method is necessary.

VII. CONCLUSION

A version of the FMM accelerated BEM is presented, where a scalable FMM is used both for dense matrix-vector multiplication and preconditioning. The equations solved are based on the Burton–Miller formulation. The numerical results show scaling consistent with the theory, which far outperforms conventional BEM in terms of memory and computational speed. Realization of a broadband FMM for efficient BEM requires different schemes for the treatment of low- and high-frequency regions and switching from multipole to signature function representation. The tests of the methods for simple and complex shapes show that it can be used for an efficient solution of scattering and other acoustical problems encountered in practice for a wide range of frequencies.

ACKNOWLEDGMENTS

This work was partially supported by a Maryland Technology Development Corporation grant in 2006 (at the University of Maryland) and partially by Fantalgo, LLC. We thank Dr. Vincent Cottoni and Dr. Phil Shorter of the ESI Group U.S. R&D for a careful testing of the software on different meshes before its inclusion in their VA-One product.

[1]L. Greengard, *The Rapid Evaluation of Potential Fields in Particle Systems* (MIT, Cambridge, MA, 1988).
[2]N. Nishimura, "Fast multipole accelerated boundary integral equation methods," Appl. Mech. Rev. **55**, 299–324 (2002).
[3]M. Fischer, "The fast multipole boundary element method and its application to structure-acoustic field interaction," Ph.D. dissertation, Institut A für Mechanik der Universität Stuttgart (2004).
[4]M. Fischer, U. Gauger, and L. Gaul, "A multipole Galerkin boundary element method for acoustics," Eng. Anal. Boundary Elem. **28**, 155–162 (2004).
[5]T. Sakuma and Y. Yasuda, "Fast multipole boundary element method for large-scale steady-state sound field analysis, Part I: Setup and validation," Acust. Acta Acust. **88**, 513–525 (2002).
[6]S. Schneider, "Application of fast methods for acoustic scattering and radiation problems," J. Comput. Acoust. **11**, 387–401 (2003).
[7]L. Shen and Y. J. Liu, "An adaptive fast multipole boundary element method for three-dimensional acoustic wave problems based on the Burton–Miller formulation," Comput. Mech. **40**, 461–472 (2007).
[8]Y. Yasuda and T. Sakuma, "Fast multipole boundary element method for large-scale steady-state sound field analysis, Part II: Examination of numerical items," Acust. Acta Acust. **89**, 28–38 (2003).
[9]Y. Yasuda and T. Sakuma, "An effective setting of hierarchical cell structure for the fast multipole boundary element method," J. Comput. Acoust. **13**, 47–70 (2005).
[10]M. S. Tong, W. C. Chew, and M. J. White, "Multilevel fast multipole algorithm for acoustic wave scattering by truncated ground with trenches," J. Acoust. Soc. Am. **123**, 2513–2521 (2008).
[11]B. Chandrasekhar and S. M. Rao, "Acoustic scattering from rigid bodies of arbitrary shape—double layer formulation," J. Acoust. Soc. Am. **115**, 1926–1933 (2004).
[12]N. A. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions* (Elsevier, Oxford, UK, 2004).
[13]N. A. Gumerov and R. Duraiswami, "Recursions for the computation of multipole translation and rotation coefficients for the 3-D Helmholtz equation," SIAM J. Sci. Comput. (USA) **25**, 1344–1381 (2003).
[14]E. Darve and P. Havé, "Efficient fast multipole method for low-frequency scattering," J. Comput. Phys. **197**, 341–363 (2004).
[15]L. Greengrard, J. Huang, V. Rokhlin, and S. Wandzura, "Accelerating fast multipole methods for the Helmholtz equation at low frequencies," IEEE Comput. Sci. Eng. **5**, 32–38 (1998).
[16]V. Rokhlin, "Diagonal forms of translation operators for the Helmholtz equation in three dimensions," Appl. Comput. Harmon. Anal. **1**, 82–93 (1993).
[17]H. Cheng, W. Y. Crutchfield, Z. Gimbutas, L. F. Greengard, J. F. Ethridge, J. Huang, V. Rokhlin, N. Yarvin, and J. Zhao, "A wideband fast multipole method for the Helmholtz equation in three dimensions," J. Comput. Phys. **216**, 300–325 (2006).
[18]*Fast and Efficient Algorithms in Computational Electromagnetics*, edited by W. C. Chew, J. M. Jin, E. Michelssen, and J. Song (Artech House, Boston, 2001).
[19]Y. Saad, "A flexible inner-outer preconditioned GMRES algorithm," SIAM J. Sci. Comput. (USA) **14**, 461–469 (1993).
[20]A. J. Burton and G. F. Miller, "The application of the integral equation methods to the numerical solution of some exterior boundary-value problems," Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).
[21]L. H. Chen and J. Zhou, *Boundary Element Methods* (Academic, New York, 1992).
[22]S. Kirkup, "The boundary element method in acoustics," http://www.boundary-element-method.com/acoustics/index.htm (Last viewed December 30, 2007).
[23]Y. Saad and M. H. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems," SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput. **7**, 856–869 (1986).
[24]N. A. Gumerov and R. Duraiswami, "Computation of scattering from clusters of spheres using the fast multiple method," J. Acoust. Soc. Am. **117**, 1744–1761 (2005).
[25]E. Darve, "The fast multipole method (I): Error analysis and asymptotic complexity," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **38**, 98–128 (2000).
[26]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, DC, 1964).
[27]J. R. Driscoll and D. M. Healy, Jr., "Computing Fourier transforms and convolutions on the 2-sphere," Adv. Appl. Math. **15**, 202–250 (1994).
[28]E. Lutz, A. R. Ingraffea, and L. J. Gray, "Use of 'simple solutions' for boundary integral equation methods in elasticity and fracture analysis," Int. J. Numer. Methods Eng. **35**, 1737–1751 (1992).
[29]D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory* (Springer-Verlag, Berlin, 1998).

Space domain complex envelopes: Definition and a spatial modulation method

Choon-Su Park[a] and Yang-Hann Kim[b]
Department of Mechanical Engineering, Center for Noise and Vibration Control (NOVIC), Korea Advanced Institute of Science and Technology (KAIST), Science Town, Daejeon 305-701, Korea

(Received 20 March 2008; revised 6 October 2008; accepted 6 October 2008)

When one visualizes a sound field as a means of treating noise sources, a detailed variation of the sound field is not required. It is sufficient to see source locations and overall variation of the field. A complex envelope in space can provide adequate information that one wishes to get because an envelope shows a gross change in signal. In other words, to interpret overall variation of sound fields in terms of a complex envelope is attempted. To achieve this objective, spatial complex envelopes have been defined firstly, and then a spatial modulation method to obtain the envelope has been theoretically developed and verified.
 [DOI: 10.1121/1.3006385]

PACS number(s): 43.60.Ac, 43.60.Gk, 43.60.Jn [EJS] Pages: 206–211

I. MOTIVES AND PROBLEM DEFINITION

The spatial variation of acoustic pressure with respect to time enables us to observe locations and shapes of time-variant as well as steady-state sound sources. The methods for time domain sound visualization predict sound field by using convolution integral or Fourier transform on time and space (for example, see Refs. 1 and 2). Therefore, not only calculation time is long but also the sound field that we obtained has much more information than what we usually want to know. For example, source locations and gross variation of the sound field are sufficient information to use sound field as a method to obtain information for dealing with noise sources.

There are two issues associated with time domain sound visualization: one is processing time issue and the other is the analysis problem.[3] The time domain sound visualization definitely requires as fast calculation time as possible. The analysis problem addresses how we have to observe and analyze the sound field. These two issues motivated us to study envelopes in time domain for reducing the number of data and define envelopes in space domain for visualizing overall variation of sound field, which provides us information that we want to have.

An envelope displays overall fluctuation of the fast varying signal, as illustrated in Fig. 1(a), and the overall variation of a signal can practically tell us whether the machinery works well or not in the case of rotating machinery (e.g., rolling bearings). Therefore, envelope analysis has been used as the fault diagnosis method (for example, see Refs. 4 and 5). Envelopes in time domain are well defined and have been applied in various ways. (e.g., Refs. 6–8) It is, nevertheless, needed to recognize that spatial envelopes of sound field have never been defined mathematically so far, even though spatial envelopes for picture images have been introduced by Oliva and Torralba[9] in 2001. Therefore, it is worth considering complex envelopes of sound field for sound visualization, which can translate a sound field into the information we want to get: source locations and overall fluctuation of sound pressure. Figure 1(b) is an example of a sound field radiated from a monopole source, which oscillates with accompanying slowly varying envelope (black line) in space domain. It is our aims to extract the slowly varying envelope from the sound field and also to propose a method to do that theoretically.

II. SPATIAL COMPLEX ENVELOPES

A. The derivation of a spatial complex envelope

It is known that acoustic waves in space (wavelength, λ) and time (frequency, f) propagate with the relation $\lambda = c/f$ (c: the speed of propagation). This means that a narrow band signal that makes a complex envelope in time also induces an associated complex envelope in space. The spatial complex envelope is therefore expected to be defined along the same line with the temporal complex envelope. Furthermore, using the simplest example (a monopole source), we definitely expect to get a clear understanding on the spatial envelope by explaining the relation between the spatial envelope and the well-known temporal envelope.

If an acoustic monopole source radiating with a center frequency f_c accompanies slowly varying amplitude $\boldsymbol{a}(r,t)$ at position r and time t and also has phases ϕ_T and ϕ_S with respect to time and space, respectively, the pressure $\boldsymbol{p}(r,t)$ can be written as

$$\boldsymbol{p}(r,t) = \boldsymbol{a}(r,t)\frac{e^{\{-j(2\pi f_c t - k_c R + \phi_T + \phi_S)\}}}{R}, \tag{1}$$

where k_c is a wavenumber corresponding to f_c (i.e., $k_c = 2\pi f_c/c$) and R is the distance between the source position and observation positions. The boldfaced font means that the value is complex.

Equation (1) can be divided into the fast varying component and the other remaining terms. The former is related to the oscillation of f_c, and the latter consists of slowly varying amplitudes and phases. That is,

[a]Electronic mail: cpure77@kaist.ac.kr
[b]Electronic mail: yanghannkim@kaist.ac.kr

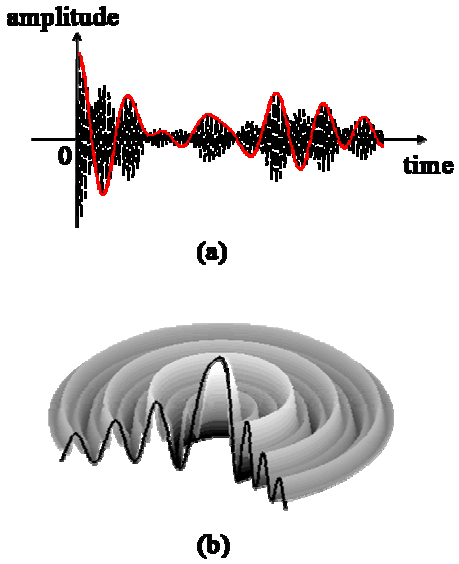


FIG. 1. (Color online) Complex envelopes in time domain and space domain: (a) a temporal complex envelope (solid line) of a time signal and (b) a spatial complex envelope (solid line) of spatially distributed sound pressure field, which radiated from a monopole source.

$$\boldsymbol{p}(r,t) = \frac{\boldsymbol{a}(r,t)e^{-j(\phi_S+\phi_T)}}{R} e^{-j(2\pi f_c t - k_c R)}. \tag{2}$$

Then, a complex envelope on space and time, $\boldsymbol{p}_{\mathrm{CE}}(r,t)$, is given as

$$\boldsymbol{p}_{\mathrm{CE}}(r,t) = \frac{\boldsymbol{a}(r,t)e^{-j(\phi_S+\phi_T)}}{R} \tag{3}$$

by removing the fast varying component in accordance with the concept of the temporal complex envelope analysis, which is used, for instance, in Ref. 8.

The slowly varying amplitude $\boldsymbol{a}(r,t)$, which has a finite number of components within a narrow band can be written as

$$\boldsymbol{a}(r,t) = \sum_{n=1}^{N} c_n e^{-j(2\pi\Delta f_n t - \Delta k_n R)}, \tag{4}$$

where the subscript $n(=1,2,\ldots,N)$ represents the index of frequency in the band, c_n is the coefficient of each frequency component, Δf_n indicates frequency much lower than the center frequency ($\Delta f_n \ll f_c$), and $\Delta k_n (=2\pi\Delta f_n/c)$ is the corresponding wavenumber to the frequency. Then, the complex envelope can be rewritten as

$$\boldsymbol{p}_{\mathrm{CE}}(r,t) = \sum_{n=1}^{N} c_n \frac{e^{-j(2\pi\Delta f_n t - \Delta k_n R + \phi_T + \phi_S)}}{R}. \tag{5}$$

As we can see from the numerator of Eq. (5), the spatial part ($e^{j(\Delta k_n R - \phi_S)}$) of the complex envelope, *spatial complex envelope*, assumes the same form of the temporal part ($e^{-j(2\pi\Delta f_n t + \phi_T)}$). It tells us that the spatial complex envelope is composed of wavenumbers associated with the frequencies of the temporal complex envelope.

However, R, which contains the information of source location, appears in the spatial complex envelope only. This distinguishes the spatial complex envelope from the temporal one. In other words, the temporal complex envelope is completely determined by frequencies in a narrow band, but the spatial complex envelope requires not only wavenumbers associated with the frequencies in the band but also geometrical characteristics of sources.

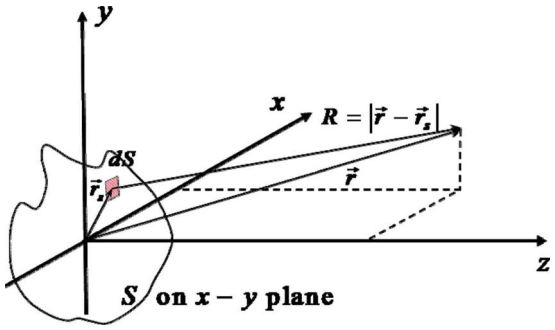


FIG. 2. (Color online) Radiation from a baffled piston of arbitrary shape on the xy-plane. S is the region of shape, and dS is the infinitesimal region for integration. $\vec{r}_s$ and $\vec{r}$ indicate source position and observation position, respectively, and R is the distance between the source and observation point.

The complex envelope of Eq. (5) is the envelope of a monopole source. Therefore, only the information of location R appears in the spatial envelope as the geometrical characteristic. Sound sources, however, cannot be regarded as a single monopole in general. Sources have their own particular shapes or distributions. The *geometric information* that has the information of sources' location, shape, or distribution is one of what we want to get from the spatial complex envelope.

B. The geometric information of spatial complex envelopes

The geometric information not only distinguishes the spatial complex envelope from the temporal one but also provides us the information associated with the geometry of sources in a sound field. It should be therefore considered in detail in order to define the spatial envelope and to find out an appropriate modulation method to get the spatial envelope.

Let us examine the sound field radiated from a baffled piston to observe the features of geometric information for general cases. The sound field generated from a baffled piston can be completely derived from the Helmholtz integral equation by choosing an appropriate Green's function that satisfies the Neumann boundary condition (see Appendix A). Hence, the baffled piston can be a simple but special case, which can serve as a model for practical sound radiators. A sound field from a vibrating piston of a constant normal velocity u_0 in the region S on the x-y plane (see Fig. 2) and also with the slowly varying amplitude of Eq. (4) can be given as

$$\boldsymbol{p}(x,y,z,t) = \frac{j\rho_0 c u_0}{2\pi} \sum_{n=1}^{N} c_n k_n e^{-j2\pi\Delta f_n t} e^{-j2\pi f_c t} \times \int_S \frac{e^{j\Delta k_n R} e^{jk_c R}}{R} dS, \tag{6}$$

where $\rho_0 c$ is the characteristic impedance of the media. The two components of the spatial complex envelope, wavenum-

bers and geometric information, are presented in the integral formula. The wavenumbers are presented as e^{jk_nR}/R ($k_n=k_c+\Delta k_n$, $n=1,2,\ldots,N$), and the geometric information is presented as the integral region S. If a window function which has the same shape of the piston is brought in, then Eq. (6) can be rewritten as

$$\boldsymbol{p}(x,y,z,t)=\frac{j\rho_0 cu_0}{2\pi}\sum_{n=1}^{N}c_nk_ne^{-j2\pi\Delta f_nt}e^{-j2\pi f_ct}\times\int_{-\infty}^{\infty}\frac{e^{j\Delta k_nR}e^{jk_cR}}{R}w(r_s)dS, \tag{7}$$

where $w(r_s)$ is the window function. By employing the function, the integral in Eq. (6) can be changed into the convolution integral between e^{jk_nR}/R and $w(r_s)$.

The complex envelope of the baffled piston would also be obtained by removing the fast varying components $e^{-j(2\pi f_ct-k_cR)}$ as we did for the case of a monopole source. To remove $e^{-j2\pi f_ct}$ or make $e^{-j2\pi f_ct}$ unity by multiplying $e^{j2\pi f_ct}$ is the complex modulation,[10] which can be done in a time domain without any difficulty. However, to make e^{jk_cR} unity in a space domain is not directly obtained because we cannot determine R in advance. To know R beforehand means that we are already aware of where sources are or how sources are distributed. In addition, e^{jk_nR}/R (related to wavenumbers) and $w(r_s)$ (related to geometric information) are convolved with each other in the space domain as shown in Eq. (7). It is therefore impossible to change wavenumbers without any change in the geometric information in the space domain. These make us observe the problem in other domains, where wavenumbers for spatial data can be changed without regard to the geometric information. The wavenumber domain can be a possible candidate because the convolution integral is changed into multiplication by the spatial Fourier transform. Then, the two components, geometric information and wavenumber, are presented separately in the wavenumber domain.

III. SPATIAL COMPLEX ENVELOPE IN WAVENUMBER DOMAIN

Let us start with the simplest case again to find out the characteristics of the spatial complex envelope in the wavenumber domain. We begin with a monopole source, which has a slowly varying amplitude and zero initial phase for simplicity. That can be given by

$$p_{\mathrm{CE}}(r,t)=\frac{\cos(2\pi f_mt-k_mR)}{R}, \tag{8}$$

where $R=\sqrt{(x-x_s)^2+(y-y_s)^2+z^2}$, and $(x_s, y_s,0)$ is the location of the monopole source. Then the pressure $\boldsymbol{p}(r,t)$ of the monopole source, which has a center frequency f_c and a complex envelope $p_{\mathrm{CE}}(r,t)$, is made up of two frequency components, $f_1(=f_c+f_m)$ and $f_2(=f_c-f_m)$. The wavenumber spectra are obtained by using two dimensional (2D) spatial Fourier transform, defined in Appendix B, with respect to each frequency. That is,

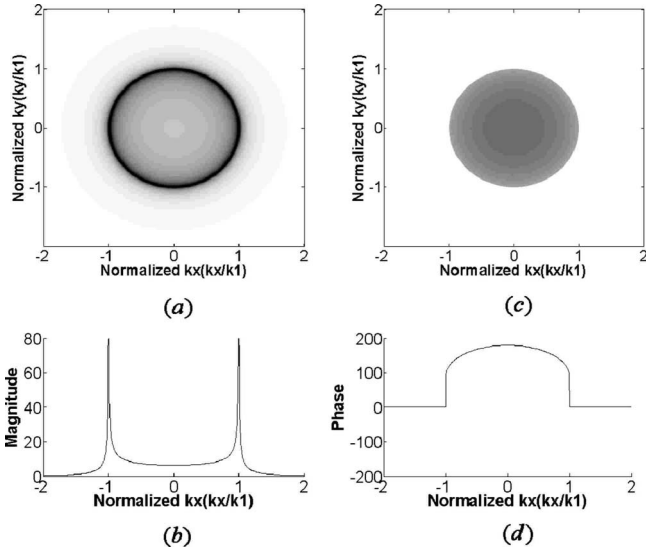


FIG. 3. The magnitude and phase of $e^{jzk_{1,z}}/k_{1,z}(k_{1,z}=\sqrt{k_1^2-k_x^2-k_y^2})$ for the wavenumber k_1. The wavenumber components on k_x and k_y are normalized with regard to k_1. The radius of circle is directly related to the wavenumber k_1. (a) The magnitude of the wavenumber spectrum. (b) The cross section of the magnitude of the spectrum along $k_y=0$. (c) The phase (degree) of the spectrum. (d) The cross section of the phase of the spectrum along $k_y=0$.

$$\boldsymbol{P}(k_x,k_y,z;f_1)=j\pi\frac{e^{jz\sqrt{k_1^2-k_x^2-k_y^2}}}{\sqrt{k_1^2-k_x^2-k_y^2}}e^{-j(x_sk_x+y_sk_y)} \tag{9}$$

for the frequency f_1 and

$$\boldsymbol{P}(k_x,k_y,z;f_2)=j\pi\frac{e^{jz\sqrt{k_2^2-k_x^2-k_y^2}}}{\sqrt{k_2^2-k_x^2-k_y^2}}e^{-j(x_sk_x+y_sk_y)} \tag{10}$$

for the frequency f_2. The associated mathematical derivation is well presented in Ref. 11. Each spectrum [Eqs. (9) and (10)] is made up of two terms: $e^{jzk_{i,z}}/k_{i,z}$ ($k_{i,z}=\sqrt{k_i^2-k_x^2-k_y^2}$, $i=1, 2$) and $e^{-j(x_sk_x+y_sk_y)}$. The former has magnitude and phase associated with the z-directional component of each wavenumber ($k_{i,z}$). Figure 3 shows the magnitude and phase for the wavenumber k_1. The radius of a circle shown in Fig. 3(a) is directly related to the wavenumber k_1, and the phase is distributed with a convex shape within the circle, as shown in Figs. 3(c) and 3(d). The latter conveys the information of source location as phase variation in the wavenumber domain. The two components of the sound field are presented separately in the wavenumber domain.

In addition, the wavenumber spectra of the complex envelope $p_{\mathrm{CE}}(r,t)$ [Eq. (8)] are also obtained similarly as

$$\boldsymbol{P}_{\mathrm{CE}}(k_x,k_y,z;f_m)=j\pi\frac{e^{jz\sqrt{k_m^2-k_x^2-k_y^2}}}{\sqrt{k_m^2-k_x^2-k_y^2}}e^{-j(x_sk_x+y_sk_y)} \tag{11}$$

for the frequency f_m and

$$\boldsymbol{P}_{\mathrm{CE}}(k_x,k_y,z;-f_m)$$

$$= \begin{cases} -j\pi \dfrac{e^{jz\sqrt{(-k_m)^2-k_x^2-k_y^2}}}{\sqrt{(-k_m)^2-k_x^2-k_y^2}} e^{-j(x_s k_x+y_s k_y)}, & 0<\sqrt{k_x^2+k_y^2}<k_m \\ \pi \dfrac{e^{-z\sqrt{k_x^2+k_y^2-(-k_m)^2}}}{\sqrt{k_x^2+k_y^2-(-k_m)^2}} e^{-j(x_s k_x+y_s k_y)}, & \sqrt{k_x^2+k_y^2}>k_m, \end{cases} \tag{12}$$

for the frequency $-f_m$. It should be noted that the wavenumber spectrum for the minus frequency, $-f_m$, has 180° out of phase with respect to the spectrum for a positive frequency (see Appendix C). By comparing Eqs. (11) and (12) [the wavenumber spectra of $p_{\mathrm{CE}}(r,t)$] with Eqs. (9) and (10) [the wavenumber spectra of $\boldsymbol{p}(r,t)$], the wavenumbers are different with the extent of k_c in spite of the same geometric information ($e^{-j(x_s k_x+y_s k_y)}$). This clearly shows us that a spatial complex envelope from the sound field can be obtained by changing the wavenumbers of the sound field into the wavenumbers of the spatial envelope.

Moreover, the wavenumber spectra for the baffled piston are also examined in the same vein with the monopole case. Equation (7) with two frequencies, $\Delta f_1=f_m$ and $\Delta f_2=-f_m$, can be written as

$$\boldsymbol{p}(x,y,z,t) = \frac{j\rho_0 c u_0}{2\pi} \sum_{n=1}^{2} c_n k_n e^{-j2\pi\Delta f_n t} e^{-j2\pi f_c t} \times \int_{-\infty}^{\infty} \frac{e^{j\Delta k_n R} e^{jk_c R}}{R} w(x_s,y_s,0)\,dx_s dy_s. \tag{13}$$

By 2D spatial Fourier transform, the wavenumber spectrum for the frequency $f_1(=f_c+\Delta f_1)$ is obtained as

$$\boldsymbol{P}(k_x,k_y,z;f_1) = \rho_0 c u_0 c_1 k_1 \frac{e^{jz\sqrt{k_1^2-k_x^2-k_y^2}}}{\sqrt{k_1^2-k_x^2-k_y^2}} \boldsymbol{W}(k_x,k_y,0), \tag{14}$$

where $\boldsymbol{W}(k_x,k_y,0)$ is obtained by a 2D spatial Fourier transform of $w(x_s,y_s,0)$. The wavenumber spectrum is similar to the monopole case, and the geometric information $\boldsymbol{W}(k_x,k_y,0)$ also appears separately from $e^{jz\sqrt{k_1^2-k_x^2-k_y^2}}/\sqrt{k_1^2-k_x^2-k_y^2}$. This allows us to alter the wavenumbers, leaving the geometric information unchanged in cases of the baffled piston as well.

IV. A MODULATION METHOD FOR SPATIAL COMPLEX ENVELOPE

The analogy between temporal and spatial complex envelopes implies that the concept of modulation for spatial data is also similar to the temporal one. In other words, as frequencies are moved by the temporal complex modulation, wavenumbers should be changed with respect to the frequencies shifted. That is, spatial modulation methods for obtaining a spatial complex envelope are nothing but to convert Eq. (9) into Eq. (11) and Eq. (10) into Eq. (12), respectively. A simple and straightforward way to change the wavenumbers, for instance, from k_1 [Eq. (9)] into k_m [Eq. (11)] without change in geometric information is to multiply Eq. (9) by

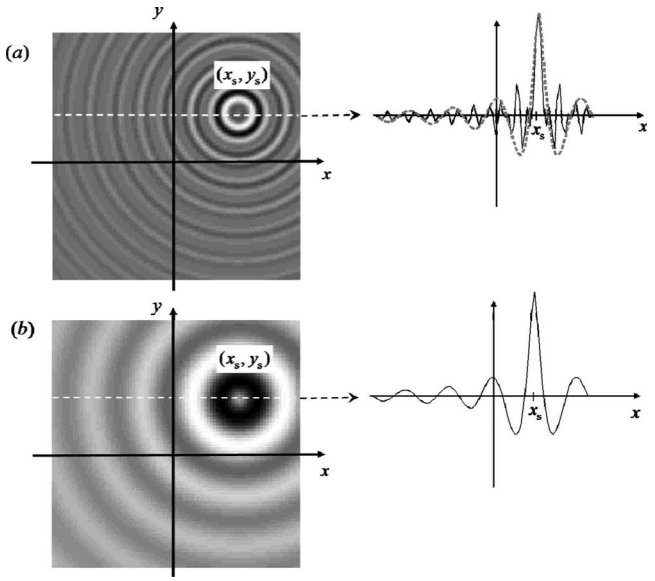


FIG. 4. A sound pressure field radiated from a monopole source at $(x_s,y_s,0)$ and its spatial complex envelope at an instance. (a) A sound field $\boldsymbol{p}(r,t)$ and its cross section with respect to $y=y_s$. The dash line indicates slowly varying amplitude, namely, spatial complex envelope. (b) The spatial complex envelope $[p_{\mathrm{CE}}(r,t)]$ of the sound field $\boldsymbol{p}(r,t)$ and its cross section with respect to $y=y_s$.

$$\boldsymbol{M}_{k_1\to k_m}(k_x,k_y,z) = \frac{\sqrt{k_1^2-k_x^2-k_y^2}}{e^{jz\sqrt{k_1^2-k_x^2-k_y^2}}} \times \frac{e^{jz\sqrt{k_m^2-k_x^2-k_y^2}}}{\sqrt{k_m^2-k_x^2-k_y^2}} \tag{15}$$

in the wavenumber domain. Equation (12) can also be obtained similarly from Eq. (10) by multiplying a modulator for the minus frequency $-f_m$ (see Appendix D). Consequently, the wavenumber spectra [Eqs. (11) and (12)] for the complex envelope [Eq. (8)] are perfectly reconstructed from multiplying Eqs. (9) and (10) by Eqs. (15) and (D1), respectively. Figure 4 shows the pressure field [Fig. 4(a)] at an arbitrary z radiated from a monopole at $(x_s, y_s, 0)$ and its complex envelope [Fig. 4(b)] at the same plane obtained by the temporal and spatial modulations. The right figures of each 2D sound field show the cross section of the spatial variation along $y=y_s$.

Furthermore, as we can see from Eqs. (9) and (14), the wavenumber spectrum of the baffled piston case is quite similar to the monopole case. Hence, the spatial modulation method on the baffled piston source must be the same with the monopole case. The wavenumber spectrum of another wavenumber $-k_m$ can also be obtained from the spectrum of k_2. Then, the complex envelope of the baffled piston can also be obtained as

$$\boldsymbol{p}(x,y,z,t) = \frac{j\rho_0 c u_0}{2\pi} \sum_{n=1}^{2} c_n k_n e^{-j2\pi\Delta f_n t} \int_{-\infty}^{\infty} \frac{e^{j\Delta k_n R}}{R} w(r_s)\,dS \tag{16}$$

by inverse Fourier transform and modulation on space and time. The fast varying terms $(e^{-j(2\pi f_c t-k_c R)})$ associated with the center frequency disappeared in Eq. (16) compared with Eq. (13). This is the very complex envelope of the baffled piston for both time and space.

The principle of the proposed modulation method is to change only the component related to wavenumbers regardless of the geometric information in the wavenumber domain. The modulation method is therefore available for any sound fields (including sound fields governed by Rayleigh's second integral equation[12]) as long as the wavenumber and geometric information of the sound fields are presented separately in the wavenumber domain. It is also noteworthy that the spatial envelope should be obtained for the same plane of the sound field. Since the wavenumber spectrum is changed by the spatial modulator $\boldsymbol{M}$, the sound field is inevitably changed when we obtain the spatial envelope in other planes using backward propagation after the modulation.

V. CONCLUSION

We have presented complex envelopes in the space domain and proposed a method to get the spatial envelopes in the wavenumber domain. Even though the spatial complex envelopes are the extension of temporal complex envelopes to the acoustic sound field, they have distinct geometric information not in temporal envelopes. A modulation method to get the spatial envelope of a sound field was proposed based on the observation of the spectra between the sound field of a monopole source and its spatial complex envelope. The modulation method is also ascertained by the baffled piston source, which is a simple but sufficient practical model as a general sound radiator.

ACKNOWLEDGMENTS

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

APPENDIX A: DERIVATION OF EQUATION (6) FROM HELMHOLTZ INTEGRAL EQUATION

The Helmholtz integral equation[12] is defined as

$$\boldsymbol{p}_1(x,y,z;f_1) = \int_S \left[j\rho_0 c k_1 \boldsymbol{G}(r|r_s)u(r_s) - p(r_s)\frac{\partial}{\partial n}\boldsymbol{G}(r|r_s)\right] dS \tag{A1}$$

for a single frequency, f_1. This integral equation is inherently derived for a single frequency. What we are interested in, however, is a narrow band, which is a signal that has several frequencies. When we have a sound source that has N numbers of frequencies in a band and the same boundary conditions, then the pressure field induced by the source can be written as

$$\boldsymbol{p}(x,y,z;f_n) = \sum_{n=1}^{N} c_n \delta(f-f_n) \int_S \left[j\rho_0 c k_n \boldsymbol{G}(r|r_s)u(r_s) - p(r_s)\frac{\partial}{\partial n}\boldsymbol{G}(r|r_s)\right] dS, \tag{A2}$$

where $n(=1,2,3,\ldots,N)$ represents the index of frequency in the band and c_n is the Fourier coefficient of each frequency.

Pressure field in time domain $[\boldsymbol{p}(x,y,z,;t)]$ can be obtained by Fourier transform as

$$\boldsymbol{p}(x,y,z;t) = \int_{-\infty}^{\infty} \boldsymbol{p}(x,y,z;f_n) e^{j2\pi f_n t} df_n. \tag{A3}$$

By putting Eq. (A2) into Eq. (A3), the pressure field can be rewritten as

$$\begin{aligned}
\boldsymbol{p}(x,y,z;t) &= \int_{-\infty}^{\infty} \left\{ \sum_{n=1}^{N} c_n \delta(f-f_n) \int_S \left[j\rho_0 c k_n \boldsymbol{G}(r|r_s)u(r_s) - p(r_s)\frac{\partial}{\partial n}\boldsymbol{G}(r|r_s)\right] dS \right\} e^{j2\pi f_n t} df_n \\
&= \sum_{n=1}^{N} c_n \int_S \left[j\rho_0 c k_n \boldsymbol{G}(r|r_s)u(r_s) - p(r_s)\frac{\partial}{\partial n}\boldsymbol{G}(r|r_s)\right] dS \\
&\quad \times \int_{-\infty}^{\infty} \delta(f-f_n) e^{j2\pi f_n t} df_n \\
&= \sum_{n=1}^{N} c_n \int_S \left[j\rho_0 c k_n \boldsymbol{G}(r|r_s)u(r_s) - p(r_s)\frac{\partial}{\partial n}\boldsymbol{G}(r|r_s)\right] dS \cdot e^{-j2\pi f_n t}.
\end{aligned} \tag{A4}$$

Here, we can select an appropriate Green's function for $\partial G(r|r_s)/\partial n = 0$, which satisfies the Neumann boundary condition. The Green's function can be obtained from the summation of the free-space Green's function as

$$\boldsymbol{G}_N(r|r_s) = \frac{e^{jkR}}{2\pi R}, \tag{A5}$$

where $R = \sqrt{(x-x_s)^2+(y-y_s)^2+z^2}$.[13]

Consequently, Eq. (A4) is given by

$$\boldsymbol{p}(x,y,z,t) = \frac{j\rho_0 c}{2\pi} \sum_{n=1}^{N} c_n k_n e^{-j2\pi f_n t} \int_S u(r_s) \frac{e^{jk_n R}}{R} dS. \tag{A6}$$

If the normal velocity is constant (u_0) on the region S and the frequency and wavenumber are expressed as $f_n = f_c + \Delta f_n$ and $k_n = k_c + \Delta k_n$, Eq. (A6) is rewritten as

$$\boldsymbol{p}(x,y,z,t) = \frac{j\rho_0 cu}{2\pi}\sum_{n=1}^{N} c_n k_n e^{-j2\pi\Delta f_n t} e^{-j2\pi f_c t} \times \int_S \frac{e^{j\Delta k_n R} e^{jk_c R}}{R} dS, \quad \text{(A7)}$$

which is the same as Eq. (6).

APPENDIX B: 2D SPATIAL FOURIER TRANSFORM PAIR

$$\boldsymbol{P}(k_x,k_y,z;f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{p}(x,y,z;f) e^{-j(k_x x + k_y y)} dxdy, \quad \text{(B1)}$$

$$\boldsymbol{p}(x,y,z;f) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{P}(k_x,k_y,z;f) e^{j(k_x x + k_y y)} dk_x dk_y. \quad \text{(B2)}$$

APPENDIX C: A WAVENUMBER SPECTRUM FOR A MINUS FREQUENCY

The pressure distribution made by a monopole with a unit magnitude and a minus wavenumber $-k_m$ is expressed as

$$\boldsymbol{p}(x,y,z;-f_m) = \frac{e^{-jk_m\sqrt{x^2+y^2+z^2}}}{\sqrt{x^2+y^2+z^2}}. \quad \text{(C1)}$$

By 2D spatial Fourier transform, the wavenumber spectrum is obtained as

$$\boldsymbol{P}(k_x,k_y,z;-f_m) = 2\pi\begin{cases} -j\dfrac{e^{jz\sqrt{k_m^2-k_x^2-k_y^2}}}{\sqrt{k_m^2-k_x^2-k_y^2}}, & 0 < \sqrt{k_x^2+k_y^2} < k_m \\ \dfrac{e^{-z\sqrt{k_x^2+k_y^2-k_m^2}}}{\sqrt{k_x^2+k_y^2-k_m^2}}, & \sqrt{k_x^2+k_y^2} > k_m \end{cases}. \quad \text{(C2)}$$

Compared with the wavenumber spectrum of a positive wavenumber, each is a complex conjugate. This makes the phase of the minus wavenumber spectrum upside down with respect to the positive one as

$$\arg\{\boldsymbol{P}(k_x,k_y,z;-f_m)\} = \tan^{-1}\left[\frac{\cos(z\sqrt{k_m^2-k_x^2-k_y^2})}{\sin(z\sqrt{k_m^2-k_x^2-k_y^2})}\right]. \quad \text{(C3)}$$

Then, by the relation of tangent and cotangent, it can be written as

$$\arg\{\boldsymbol{P}(k_x,k_y,z;-f_m)\} = -z\sqrt{k_m^2-k_x^2-k_y^2}. \quad \text{(C4)}$$

This result is opposite of the value with the phase of the positive wavenumber, $z\sqrt{k_m^2-k_x^2-k_y^2}$.

APPENDIX D: A SPATIAL MODULATOR FOR THE MINUS FREQUENCY

Particularly, a spatial modulator changing the wavenumber spectrum for the frequency f_2 into that of frequency $-f_m$ can be written as

$$\boldsymbol{M}_{k_2\to -k_m}(k_x,k_y,z) = \begin{cases} -\dfrac{e^{jz\sqrt{(-k_m)^2-k_x^2-k_y^2}}}{\sqrt{(-k_m)^2-k_x^2-k_y^2}} \times \dfrac{\sqrt{k_2^2-k_x^2-k_y^2}}{e^{jz\sqrt{k_2^2-k_x^2-k_y^2}}}, & 0 < \sqrt{k_x^2+k_y^2} < k_2 \\ \dfrac{e^{-z\sqrt{k_x^2+k_y^2-(-k_m)^2}}}{\sqrt{k_x^2+k_y^2-(-k_m)^2}} \times \dfrac{\sqrt{k_x^2+k_y^2-k_2^2}}{e^{-z\sqrt{k_x^2+k_y^2-k_2^2}}}, & \sqrt{k_x^2+k_y^2} > k_2 \end{cases}. \quad \text{(D1)}$$

[1] M. Forbes, S. Letcher, and P. Stepanishen, "A wave-vector, time-domain method of forward projecting time-dependent pressure fields," J. Acoust. Soc. Am. **90**, 2782–2793 (1991).
[2] J. Hald, "Time domain acoustical holography," Proceedings of Inter-Noise 95, Newport Beach, CA (1995) Vol. **2**, pp. 1349–1354.
[3] Y.-H. Kim, "Can we hear the shape of a noise source?," Plenary Lecture, The 18th International Congress on Acoustic, Kyoto International Conference Hall, Japan, (April 4–9, 2004) pp. 3357–3370.
[4] R. H. Lyon, *Machinery Noise and Diagnostics* (Butterworth, Washington, DC, 1987) Chap. 6.
[5] Y.-H. Kim, W.-S. Cheung, and Y.-K. Kwak, "Complex envelope of the vibration signature of ball bearing system and its application to fault detection," Fourth International Congress on Experimental Mechanics, Portland, OR (1988) pp. 1067–1071.
[6] J. Dugundji, "Envelopes and pre-envelopes of real waveforms," IRE Trans. Inf. Theory **4**, 53–57 (1958).
[7] S. O. Rice, "Envelopes of narrow-band signals," Proc. IEEE Vol. **70**, 692–699 (1982).
[8] S. J. Roome, "Analysis of quadrature detectors using complex envelope notation," IEE Proc. F, Radar Signal Process. **136**, 95–100 (1989).
[9] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the Spatial envelope," Int. J. Comput. Vis. **42**, 145–175 (2001).
[10] D. G. Childers and M. Pao, "Complex demodulation for transient wavelet detection and extraction," IEEE Trans. Audio Electroacoust. **AU-20**, 295–308 (1972).
[11] H.-S. Kwon, "Sound visualization by using enhanced planar acoustic holographic reconstruction," Ph.D thesis, KAIST (1997), Chap. 2 and Appendix.
[12] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic p, New York, 1999) Chap. 1, pp. 1–7.
[13] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications*, (Acoustical Society of America, 1989), Chap. 5, pp. 213–217.

Basin-scale time reversal communications

H. C. Song,[a] W. A. Kuperman, and W. S. Hodgkiss
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

(Received 11 January 2008; revised 15 October 2008; accepted 16 October 2008)

During November 1994, broadband acoustic signals were transmitted from a 75-Hz source to a 20-element, 700-m vertical array at approximately 3250 km range in the eastern North Pacific Ocean as part of the acoustic engineering test (AET) of the acoustic thermometry of ocean climate program [Worcester *et al.*, J. Acoust. Soc. Am. **105**, 3185–3201 (1999)]. The AET tomography signal can be treated as a binary-phase shift-keying communication signal with an information rate of 37.5 bits/s. With the multipath arrivals spanning 5–8 sec, these data represent an extreme case of intersymbol interference. The AET array data are processed using time reversal combined with frequent channel updates to accommodate channel variations over the 20-min long reception, followed by a *single* channel decision-feedback equalizer. The almost error-free performance using all 20 array elements demonstrates the feasibility of time reversal communications at basin scale. Further, comparable performance of single receive element communications integrating over multiple transmissions indicates that the ocean provided temporal diversity that is as effective as the spatial diversity provided by the array.
 [DOI: 10.1121/1.3021435]

PACS number(s): 43.60.Dh, 43.60.Gk, 43.60.Fg [DRD] Pages: 212–217

I. INTRODUCTION

The basic problem encountered in ocean acoustic communications is the time-dependent, highly dispersive nature of the channel. While procedures have been developed for line-of-sight and some multipath ocean environments to ranges typically much less than 100 km,[1–3] very long range propagation utilizing the deep sound channel is particularly problematic. For example, data from the acoustic engineering test (AET) of the acoustic thermometry of ocean climate project[4] clearly highlight these difficulties. During the AET, the propagation of 37.5 Hz bandwidth pulses at a carrier frequency of 75 Hz was observed at long range. The effective pulse length of $1/37.5=25$ ms at the transmitter exhibits a time spread at the receiver of greater than 5 sec at the range of 3250 km,[5] or about 200 times the length of the original "symbol."

The AET was originally conducted as a tomography experiment in which the early arriving, weaker but visible and identifiable wavefronts were used for inversion of ocean properties (see Fig. 2). Tomographers typically discard a majority of the energy that is concentrated near the sound channel axis because the arrival structure cannot be related to a background ocean acoustic model. However, it is just this complex, high-energy late-arrival structure that can be utilized by acoustic time reversal methods to enable very long range acoustic communications without the necessity of employing computationally formidable multichannel equalizers.[5]

In this paper, we demonstrate using the AET array data that time reversal combined with frequent channel updates to accommodate channel variations over the 20-min long reception, followed by a *single* channel decision-feedback equalizer (DFE) provides almost error-free performance using all 20 array elements. Further, using data from a single receive element with multiple receptions provides similar almost error-free performance as in the single reception array case, but of course with the bit rate reduction associated with the number of repetitions integrated.

The paper is organized as follows. Section II reviews the time reversal communications. Section III describes the tomography experiment conducted during November 1994. Section IV investigates the temporal variations of the channel. The proposed time reversal receiver and the resulting performance are presented in Section V.

II. TIME REVERSAL COMMUNICATIONS

Over the past several years, time reversal (or phase conjugation in the frequency domain) communications has been discussed extensively in the literature. The approach exploits spatial diversity to achieve spatial and temporal focusing in complex multipath environments. The preliminary system concept has been demonstrated using experimental data collected in shallow water for both active[6] and passive[7] time reversal communications. Mathematically, the two approaches are equivalent with the communications link being in the opposite direction by invoking spatial reciprocity of wave propagation.[8,9] Yang[10] addressed the difference between passive phase conjugation (PPC) and the decision-feedback equalizer[11] approach and Preisig[12] compared the PPC and channel-estimate based DFE (CE-DFE).

While the temporal focusing (pulse compression) achieved by the time reversal approach reduces intersymbol interference (ISI) significantly, there always is some residual ISI that results in saturation of the performance at high SNR.[13] As a result, the time reversal approach has been combined with adaptive channel equalization, which simulta-

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

neously removes the residual ISI and compensates for slow channel variations to yield better performance than is achieved by time reversal alone.[14–17] Indeed, Song and Kim[18] have confirmed that the combination offers nearly optimal performance using the theoretical performance bounds derived in Ref. 12. Furthermore, comparison between theory and data[19] suggests that the theoretical performance can provide a useful upper bound for predicting performance of time reversal communications. The combined approach is referred to as the correlation-based DFE by Yang.[17] Since the time reversal approach collapses multiple channels into a single channel, the complexity of the successive DFE remains unchanged as the number of array elements increases, resulting in low computational complexity.

Normally, the time reversal approach assumes that the channel is time invariant or slowly varying. Passive time reversal applies spatiotemporal matched filtering to combine multichannel data, using either measured channel responses from channel probes or initial channel estimates from training symbols at the beginning of the data packet. In dynamic ocean environments, however, the channel can vary substantially over time and the performance of time reversal communications will deteriorate due to the mismatch between the actual and assumed channel impulse responses. A simple approach to avoid the mismatch is frequent transmission of a channel probe signal at the expense of data rate.[7] A more elaborate approach is to track continuously the channel using previously detected symbols (decision-directed mode) prior to matched filtering without compromising the data rate. In fact, time reversal alone with frequent channel updates has been proposed and referred to as decision-directed passive phase conjugation (DDPPC).[20] More recently it has been demonstrated that continuous channel updates prior to time reversal combining followed by a single channel DFE significantly improve the performance in a time-varying channel due to environmental fluctuations.[21] Separately, joint time reversal combining with channel updates and multichannel equalization has been discussed with emphasis on the use of low complexity multichannel algorithms.[22]

The combination of robustness from the self-averaging time reversal process[23] and low computational complexity from both a single channel DFE and a significantly reduced number of taps due to temporal compression appears promising for challenging underwater acoustic communications.[3] To date, however, time reversal communications has been demonstrated mostly in shallow water environments on the order of 100-m depth and up to 20-km range.[8] Interestingly, Dowling[9] suggested in the early 1990s that time reversal pulse compression would be beneficial for communications in deep water where the multipath spread is relatively long by analyzing deep-ocean propagation measurements (IWAC'90) conducted at 272 km range at 460 Hz.

In this paper, we investigate the feasibility of time reversal communications at long ranges in deep water. This is partly motivated by an earlier effort[5] where acoustic tomography transmissions over a 3250 km path in the North Pacific Ocean[4] at 75 Hz were examined using an adaptive multichannel DFE with integrated phase tracking and Doppler compensation. We will analyze the same data set using a temporally adaptive time reversal approach, i.e., multichannel time reversal combining with continuous channel updates followed by a single channel DFE.[8,21]

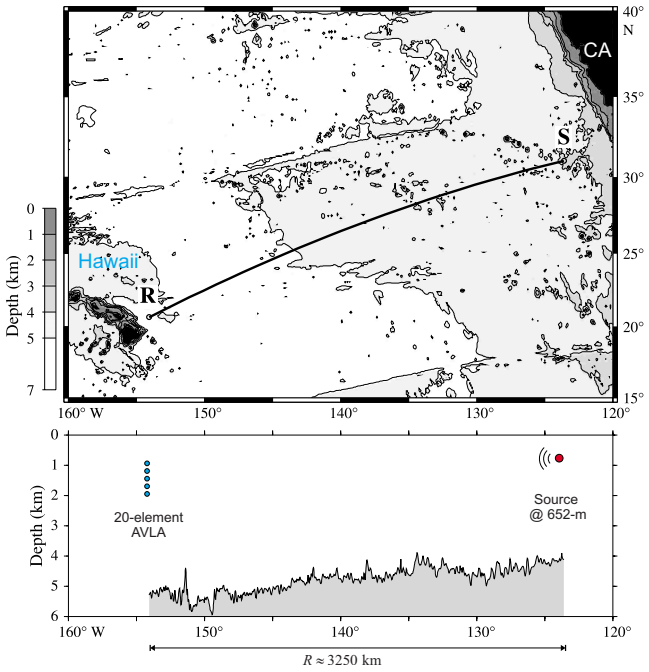


FIG. 1. (Color online) Acoustic path from the 75-Hz source *S* suspended from R/P FLIP off southern California to the vertical receiving array *R* moored east of Hawaii at approximately 3250 km range along with the bathymetry (reproduction of Fig. 1 in Ref. 4).

III. ACOUSTIC ENGINEERING TEST (AET)

As a preliminary test of the acoustic thermometry of ocean climate (ATOC) project, the acoustic engineering test was conducted for 7 days in November 1994 in the eastern North Pacific Ocean as shown in Fig. 1. A detailed description of the experiment can be found in Ref. 4 and here we review the main parameters relevant for our discussion.

The 75-Hz acoustic source *S* was suspended at 652-m depth, near the sound-channel axis, from the floating instrument platform R/P FLIP moored off San Diego, California in 4000 m deep water. The source level was 195 dB re 1 μPa at 1 m. The receiver denoted by *R* was an autonomous vertical line array moored just east of Hawaii in water about 5300 m deep. The vertical receiving array consists of 20 hydrophones at 35-m spacing (1.75 wavelengths at 75 Hz) between 900- and 1600-m depth, below the sound-channel axis (650-m). The receiving array is located 3250 km away from the source. The geodesic acoustic path is depicted in Fig. 1 along with the bathymetry, indicating no significant feature. Sound speed profiles between the source and receiving array can be found in Fig. 5 of Ref. 4.

The tomography source transmitted a phase-modulated maximal-length shift-register sequence (*m*-sequence) at the carrier frequency of 75 Hz with a bandwidth of $W = 37.5$ Hz. The *m*-sequence contains 1023 binary digits (or chips) and the period of the sequence is 27.28 s with each digit consisting of two carrier cycles. The number of periods for each transmission was a mix of 20, 40, or 80 correspond-

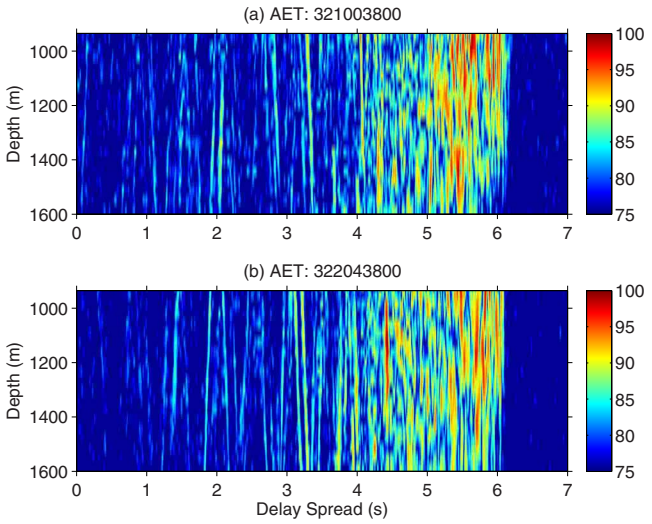


FIG. 2. (Color online) Acoustic intensity (dB) as a function of time delay spread and hydrophone depth for two ATOC AET receptions on (a) JD321 (321003800) and (b) JD322 (322043800). Early arriving wavefronts are visible but weak, while a majority of the energy is concentrated on the later arrivals in the last 2–3 sec.

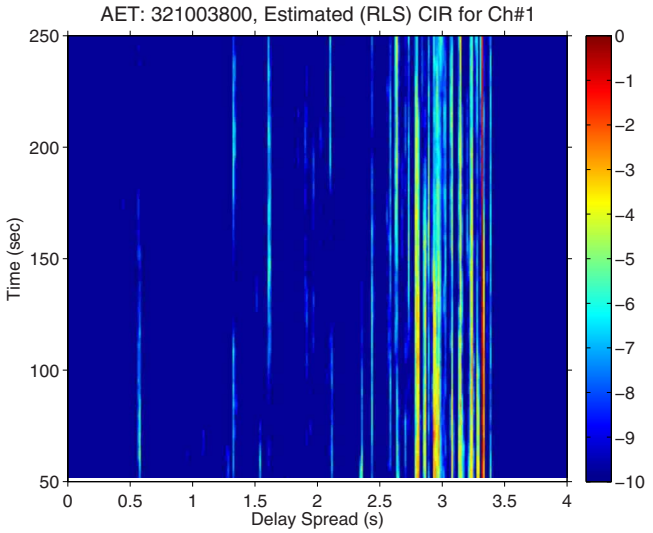


FIG. 3. (Color online) Temporal evolution of the channel response for Ch#1 at the top of the vertical array (935-m) on JD321 reception shown in Fig. 2(a).

ing to approximately 10-, 20-, or 40-min, with 2 or 4 hours between transmissions. The sampling frequency at the receiver was $f_s=300$ Hz.

Figure 2 shows two 20-min AET receptions in time/depth format on JD321 and JD322 (28 h apart), respectively, displaying distinctive arrival patterns. The delay spread is about 6 sec. The early arriving wavefronts are visible but weak while a majority of the energy is concentrated on the later arrivals associated with low-order modes propagating near the sound-channel axis. The later arrivals with significant energy will be utilized for acoustic communications as opposed to the early identifiable rays used for typical acoustic tomography data analysis.[4]

The binary *m*-sequence can be interpreted as a binary-phase shift-keying (BPSK) communication signal with an information rate of 37.5 bits/s. Alternatively, it can be treated as a direct sequence (DS) or pseudonoise (PN) spread spectrum signal[24] where each 1023 chip sequence represents a single information bit with a much slower data rate,[25] i.e., 1 bit every 27.28 s. In this paper, we will analyze the *m*-sequence as a BPSK communication signal with a data rate of 37.5 bits/s. For communications processing, the received data are complex demodulated to baseband and then sampled at twice the symbol rate (i.e., 75 Hz) facilitating use of a fractionally sampled DFE in Sec. V.

IV. CHANNEL ESTIMATION

Acoustic tomography processing coherently combines multiple periods of the *m*-sequence in order to extract the low-level signal at long ranges[4] as presented in Fig. 2. Now that the tomography signal is treated as a BPSK communication signal, it is possible to track the channel continuously at the symbol rate (37.5 Hz) so that we can evaluate how rapidly the channel is evolving. In the analysis that follows, the 40-period, 20-min long transmission on JD321 shown in Fig. 2(a) will be examined.

Using the known sequence of symbols, the temporal evolution of the channel response over the first 5-min period is shown in Fig. 3 for a single receiver (Ch#1) at the top of the vertical array (935-m). Note that about 3-s delay spread captures all the significant arrivals within 10 dB from the maximum. This truncated version of the channel response will be used for time reversal processing in Sec. V. Coarse synchronization of the start time was accomplished by correlating with one period of the *m*-sequence. Channel estimation is carried out using an exponentially weighted recursive least squares (RLS) algorithm[26] with forgetting factor of $\lambda=0.9995$. Signal fading is visible in the early identifiable arrivals while the final 1 s of the arrival exhibits a highly variable complicated structure associated with scattering of near-axial energy by internal waves.[4] The 3-s delay spread corresponds to intersymbol interference of approximately 110 symbols at the symbol rate (37.5 Hz).

To further examine the temporal variations of the channel, the channel response can be correlated with the initial channel estimate, equivalent to the generalized $q(t)$ function[8,27] for a single channel (i.e., $M=1$) such that

$$q(t;\tau)=g(-t;0)*g(t;\tau), \tag{1}$$

where $g(t;\tau)$ is the response of the channel at delay time t to an impulse applied at time τ and $*$ denotes convolution. The magnitude (correlation) and phase of the complex baseband version of $q(0;\tau)$ function is displayed in Fig. 4. It should be mentioned that the phase tracking result shown in Fig. 4 (lower panel) is similar to the one presented in Ref. 5. The initial phase slope up to 300 s corresponds to a Doppler shift of $f_d=0.0015$ Hz.

As expected, the channel correlation steadily deteriorates and reduces to 3 dB at around 800 s (or 13-min). Incidentally, Worcester *et al.*[4] reported that the coherence time for acoustic tomography is on the order of 13 min (about 28 periods). Our analysis of other AET transmissions, however, indicates that the coherence time measured by the 3-dB down criterion could be much shorter. For example, the data

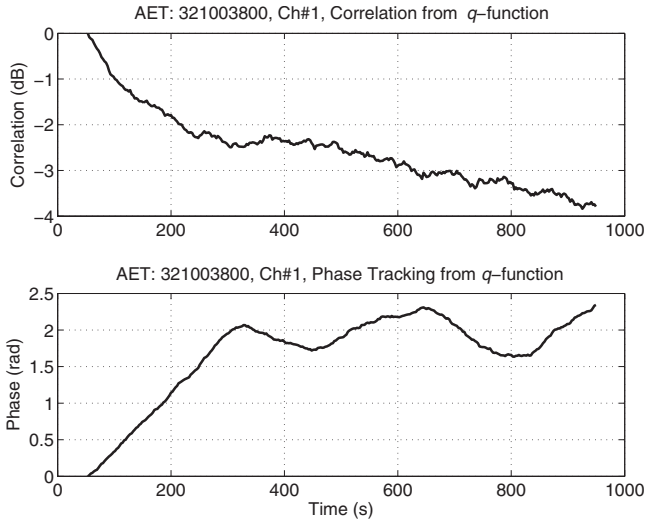


FIG. 4. The magnitude (correlation) and phase of the complex-basebanded version of the q function versus time.

corresponding to Fig. 2(b) are found to have the coherence time less than 5-min (12 periods). Thus, the plots in Fig. 2 are obtained from coherent averaging of just 12 periods of the m-sequence after appropriate Doppler processing.[4]

Of course the 3-dB loss is not acceptable for this low-energy signal and requires more frequent channel updates to minimize the mismatch between the actual and assumed channels over the 20-min long reception. In fact, the channel updates will be performed every one or two periods for best performance as shown in Sec. V.

V. RECEIVER AND PERFORMANCE

The proposed receiver for processing the AET data is illustrated in Fig. 5 where passive time reversal multichannel

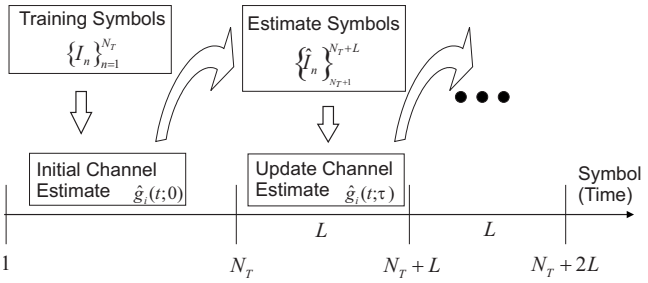


FIG. 6. Procedure for channel updates. Initial channel estimates from the N_T training symbols are applied as matched filters to the next block of L symbols for time reversal demodulation using the receiver in Fig. 5. The newly estimated L symbols are used for updating the channel estimates, which are then applied to the next block of L symbols.

combining is followed by a single channel decision-feedback equalizer.[8] The spatiotemporal matched filter or demodulation filter corresponds to the channel response $G_i(f) = G_0^*(f)H_i^*(f)$ where $G_0(f)$ is the signal transmitted (i.e., rectangular envelope) and $H_i(f)$ is the channel impulse response, since the received data are a convolution of the two signals. In our approach, phase tracking is conducted prior to the DFE so that the DFE focuses mainly on eliminating the residual ISI. We employ a decision-feedback carrier phase estimate based on the maximum likelihood (ML) criterion.[24]

To accommodate the channel variations addressed in Sec. IV, frequent channel updates will be incorporated as illustrated in Fig. 6. The initial channel estimates from the N_T training symbols are applied as matched filters to the next block of L symbols for time reversal demodulation using the receiver in Fig. 5. The estimated L data symbols can be used to update the channel estimates, which then are applied to the next block of L symbols for demodulation, and so on. For the channel estimates, we will use the least mean square (LMS)

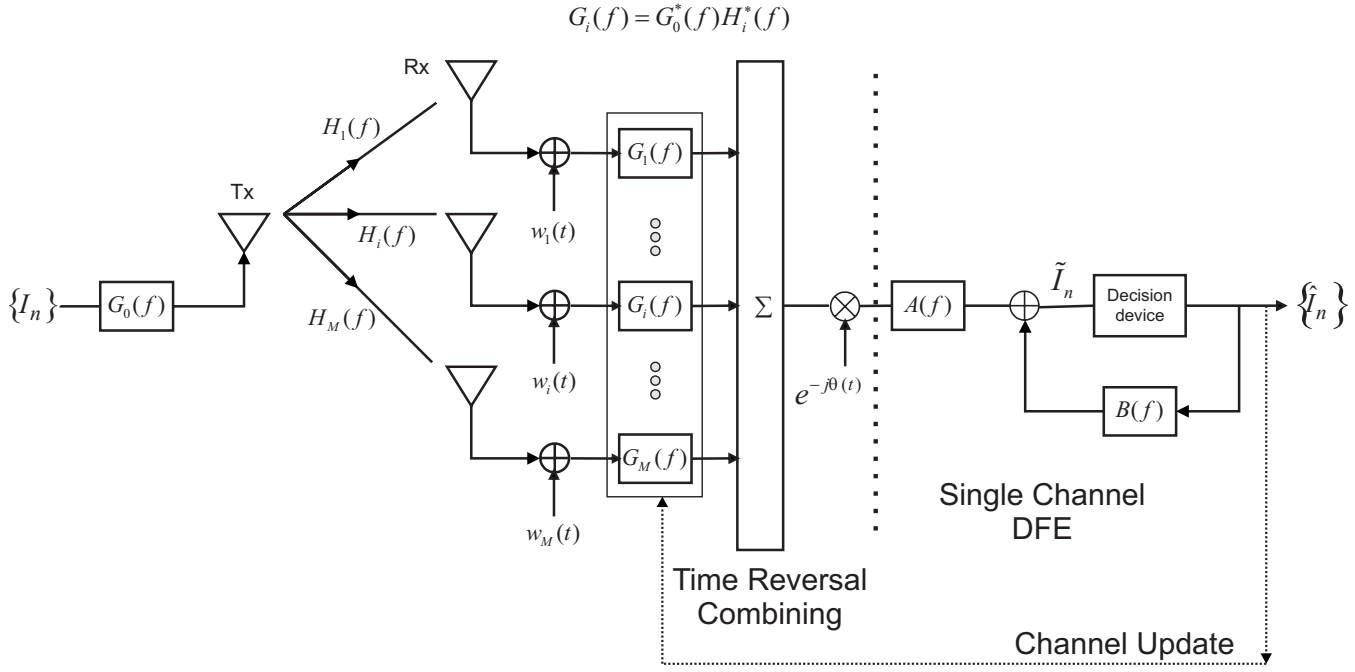


FIG. 5. Block diagram for passive time reversal communications with channel updates. Multichannel time reversal combining is followed by a single channel decision-feedback equalizer (DFE). Note that phase tracking is performed prior to the DFE. For time-varying environments, frequent channel updates using previously detected symbols can be incorporated (see Fig. 6).

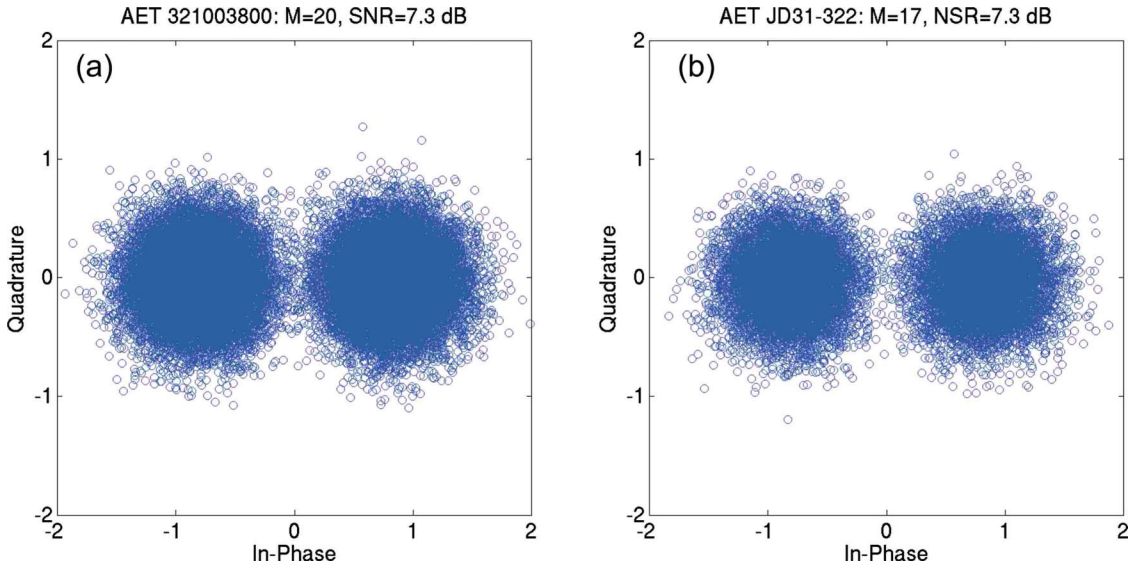


FIG. 7. (Color online) Performance of time reversal communications in terms of scatter plot: (a) vertical array diversity from $M=20$ elements and (b) temporal diversity combined from $M=17$ transmissions over a 2-day period (JD321–322) for a single element (Ch#1). The output SNR is 7.3 dB for both cases.

algorithm for faster execution[24] while the single channel fractionally sampled DFE ($K=2$) employs the RLS algorithm in our analysis. This hybrid approach allows for keeping the computational complexity minimal.

The performance of time reversal communications with continuous channel update applied to the data [JD321003800, Fig. 2(a)] is displayed as a scatter plot in Fig. 7(a). The output SNR is 7.3 dB using all 20 elements of the vertical array ($M=20$) with a bit error rate (BER) of 49/34782. For the same data set, Ref. 5 employed a multichannel DFE with integrated phase tracking and Doppler compensation and reported 6.3 dB output SNR. Note that explicit Doppler processing (resampling) was not required in our approach although phase tracking was carried out by a decision-directed carrier phase estimate. The number of fractionally spaced feed-forward taps in the equalizer is chosen to be 60 corresponding to about 0.8 s, which is much shorter than the 3-s delay spread (see Fig. 3) resulting from temporal compression of time reversal. The number of feedback taps in the equalizer is set to 30. The LMS step size for channel estimation is $\Delta=0.00025$ and the RMS forgetting factor for the adaptive DFE is $\lambda=0.9995$. The number of training symbols and update block size are $N_T=3\times1023$ (3 periods) and $L=2\times1023$ (2 periods), respectively.

In terms of computational complexity, the total number of parameters for the adaptive DFE is 90(=60+30) in addition to the channel estimation computations. Since the fast LMS algorithm is employed for channel estimation, it takes about 5-min on a conventional desktop workstation to process the 20-min long transmission using all 20 array elements. In contrast, the multichannel DFE approach in Ref. 5 involved 3600 parameters when using all 20 elements and, thus, used the LMS algorithm for the DFE. As pointed out in Sec. II, the impact of the number of receivers in the time reversal approach is minimal due to the subsequent processing by a single channel DFE.

The performance of time reversal communications for a number of different transmissions over a 2-day period (JD321–JD322) is presented in Fig. 8 in terms of BER when $M=20$ (circles) and $M=10$ (triangles). Transmission 1 refers to the data analyzed above [JD321003800, Fig. 2(a)]. The BERs are on the order of 10^{-3} using $M=20$ and increase to approximately 10^{-2} when using only 10 elements ($M=10$). The best performance with a BER of 1/34782 and 9.3 dB output SNR using $M=20$ is obtained from transmission 5 (JD322043800) shown in Fig. 2(b).

To this point, time reversal communications at basin scale has been demonstrated exploiting the spatial diversity of a vertical array for individual 20-min long transmissions. It might also be possible to achieve similar diversity from temporal variations of the channel, provided that the channel responses are sufficiently uncorrelated with each other. Recall from Sec. IV that the channel coherence time for the

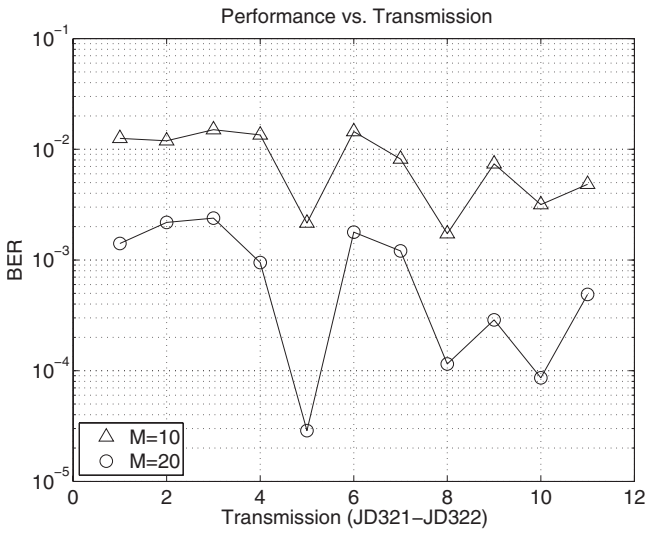


FIG. 8. Performance of time reversal communications in terms of BER for different transmissions over a 2-day period (JD321–322) using vertical array diversity when $M=20$ (circles) and $M=10$ (triangles). Transmission 1 refers to JD321003800 [Fig. 2(a)]. The best performance is obtained from transmission 5 (JD322043800) displayed in Fig. 2(b).

AET data is in the vicinity of 5–13 min. To examine temporal diversity, we have selected a single element (Ch#1) and combined a total of 17 transmissions made over a 2-day period (JD321–JD322) with transmissions separated by 2 or 4 hr (two of them are shown in Fig. 2). In other words, this is single receive element communications with diversity being obtained from the multiple transmissions separated much longer than the coherence time. The resulting scatter plot is shown in Fig. 7(b) with a comparable performance of 7.3 dB and BER of 16/19368. Some of the 17 transmissions were only 10-min long and, thus, we processed just the first 10-min of data for each transmission. Note that all the parameters employed are identical except that the update block size is reduced by a half, i.e., $L=1\times1023$ (one period) to compensate for some fast-varying receptions.

VI. SUMMARY AND CONCLUSIONS

During November 1994, broadband acoustic signals were transmitted from a 75-Hz source to a 20-element, 700-m vertical array at approximately 3250 km range in the eastern North Pacific Ocean as part of the acoustic engineering test of the acoustic thermometry of ocean climate project. The AET tomography signal consisted of repetitions of a *m*-sequence containing 1023 digits with two carrier cycles per digit, which was treated as a BPSK communication signal with an information rate of 37.5 bits/s. Although the multipath arrivals spanned 5–8 sec, most of the energy was contained by the later arrivals associated with low-order modes in the last 2–3 sec. The AET array data were processed using time reversal multichannel combining with frequent channel updates to accommodate channel variations over the 20-min receptions, followed by a single channel fractionally sampled DFE. The excellent performance using all 20 elements (vertical spatial diversity) with low computational complexity demonstrated the feasibility of time reversal communications even at basin scale. In addition, the comparable performance of single receive element communications using multiple transmissions illustrated that temporal diversity is as effective as spatial diversity.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Grant No. N00014-06-1-0128. We are grateful to Matthew Dzieciuch for providing the AET data and associated tomography analysis software.

[1]V. Capellano, "Performance improvements of a 50 km acoustic transmission through adaptive equalization and spatial diversity," in *Proc. of Oceans'97*, 569–573 (1997).
[2]A. Plaisant, "Long range acoustic communications," in *Proc. of Oceans'98*, 472–476 (1998).
[3]D. Kilfoyle and A. Baggeroer, "The state of the art in underwater acoustic telemetry," IEEE J. Ocean. Eng. **25**, 4–27 (2000).
[4]P. Worcester, B. Cornuell, M. Dzieciuch, W. Munk, B. Howe, J. Mercer, R. Spindel, J. Colosi, K. Metzger, T. Birdsall, and A. Baggeroer, "A test of basin-scale acoustic tomography using a large aperture vertical array at 3250-km range in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[5]L. Freitag and M. Stojanovic, "Basin-scale acoustic communications: A feasibility study using tomography m-sequences," in *Proc. MTS/IEEE OCEANS'01*, 2256–2261 (2001).
[6]G. Edelmann, T. Akal, W. Hodgkiss, S. Kim, W. Kuperman, H. Song, and T. Akal, "An initial demonstration of underwater acoustic communication using time reversal mirror," IEEE J. Ocean. Eng. **27**, 602–609 (2002).
[7]D. Rouseff, D. Jackson, W. Fox, C. Jones, J. Ritcey, and D. Dowling, "Underwater acoustic communications by passive-phase conjugation: Theory and experimental results," IEEE J. Ocean. Eng. **26**, 821–831 (2001).
[8]H. C. Song, W. S. Hodgkiss, W. A. Kuperman, T. Akal, and M. Stevenson, "Multiuser communications using passive time reversal,"IEEE J. Ocean. Eng. **32**, 915–926 (2007).
[9]D. R. Dowling, "Acoustic pulse compression using passive phase-conjugate processing," J. Acoust. Soc. Am. **95**, 1450–1458 (1994).
[10]T. C. Yang, "Differences between passive-phase conjugation and decision-feedback equalizer for underwater acoustic communications," IEEE J. Ocean. Eng. **29**, 472–487 (2004).
[11]M. Stojanovic, J. A. Capitovic, and J. G. Proakis, "Adaptive multi-channel combining and equalization for underwater acoustic communications," J. Acoust. Soc. Am. **94**, 1621–1631 (1993).
[12]J. C. Preisig, "Performance analysis of adaptive equalization for coherent acoustic communications in the time-varying ocean environment," J. Acoust. Soc. Am. **118**, 263–278 (2005).
[13]M. Stojanovic, "Retrofocusing techniques for high rate acoustic communications," J. Acoust. Soc. Am. **117**, 1173–1185 (2005).
[14]G. Edelmann, H. Song, S. Kim, W. Hodgkiss, W. Kuperman, and T. Akal, "Underwater acoustic communication using time reversal," IEEE J. Ocean. Eng. **30**, 852–864 (2005).
[15]H. C. Song, W. S. Hodgkiss, W. A. Kuperman, M. Stevenson, and T. Akal, "Improvement of time reversal communications using adaptive channel equalizers," IEEE J. Ocean. Eng. **31**, 487–496 (2006).
[16]H. C. Song, W. S. Hodgkiss, W. A. Kuperman, W. Higley, K. Raghukumar, T. Akal, and M. Stevenson, "Spatial diversity in passive time reversal communications," J. Acoust. Soc. Am. **120**, 2067–2076 (2006).
[17]T. C. Yang, "Correlation-based decision-feedback equalizer for underwater acoustic communications," IEEE J. Ocean. Eng. **30**, 865–880 (2005).
[18]H. C. Song and S. M. Kim, "Retrofocusing techniques in a waveguide for acoustic communications (L)," J. Acoust. Soc. Am. **121**, 3277–3279 (2007).
[19]H. C. Song, W. S. Hodgkiss, and S. M. Kim, "Performance prediction of passive time reversal communications (L)," J. Acoust. Soc. Am. **122**, 2517–2518 (2007).
[20]J. Flynn, J. Ritcey, D. Rouseff, and W. Fox, "Multichannel equalization by decision-directed passive phase conjugation: Experimental results," IEEE J. Ocean. Eng. **29**, 824–836 (2004).
[21]A. Song, M. Badiey, H. C. Song, W. S. Hodgkiss, M. Porter, and KauaiEx, "Impact of ocean variability on coherent underwater acoustic communications during the Kauai experiment (KauaiEx)," J. Acoust. Soc. Am. **123**, 856–865 (2008).
[22]J. Gomes, A. Silva, and S. Jesus, "Adaptive spatial combining for passive time-reversed communications," J. Acoust. Soc. Am. **124**, 1038–1053 (2008).
[23]M. Fink, "Time-reversed acoustics," Sci. Am. November, 91–97 (1999).
[24]J. Proakis, *Digital Communications* (McGraw-Hill, New York, 2001).
[25]T. Yang and W.-B. Yang, "Performance analysis of direct-sequence spread-spectrum underwater acoustic communications with low signal-to-noise ratio input signals," J. Acoust. Soc. Am. **123**, 842–855 (2008).
[26]S. Haykin, *Adaptive Filter Theory* (Prentice-Hall, Upper Saddle River, NJ, 2002).
[27]T. Yang, "Temporal resolutions of time-reversed and passive-phase conjugation for underwater acoustic communications," IEEE J. Ocean. Eng. **28**, 229–245 (2003).

The prolate spheroidal wave functions as invariants of the time reversal operator for an extended scatterer in the Fraunhofer approximation

Jean-Luc Robert
Philips Research North America, 345 Scarborough Road, Briarcliff Manor, New York 10510

Mathias Fink
Laboratoire Ondes et Acoustique, ESPCI, Université Paris 7, 10 rue Vauquelin, 75005 Paris, France

(Received 6 February 2008; revised 17 October 2008; accepted 20 October 2008)

The decomposition of the time reversal operator, known by the French acronym DORT, is widely used to detect, locate, and focus on scatterers in various domains such as underwater acoustics, medical ultrasound, and nondestructive evaluation. In the case of point-scatterers, the theory is well understood: The number of nonzero eigenvalues is equal to the number of scatterers, and the eigenvectors correspond to the scatterers Green's function. In the case of extended objects, however, the formalism is not as simple. It is shown here that, in the Fraunhofer approximation, analytical solutions can be found and that the solutions are functions called prolate spheroidal wave-functions. These functions have been studied in information theory as a basis of band-limited and time-limited signals. They also arise in optics. The theoretical solutions are compared to simulation results. Most importantly, the intuition that for an extended objects, the number of nonzero eigenvalues is proportional to the number of resolution cell in the object is justified. The case of three-dimensional objects imaged by a two-dimensional array is also dealt with. Comparison with previous solutions is made, and an application to super-resolution of scatterers is presented.
 [DOI: 10.1121/1.3023060]

PACS number(s): 43.60.Fg, 43.20.Fn, 43.60.Tj, 43.80.Qf [TDM] Pages: 218–226

I. INTRODUCTION

The DORT method[1,2] (French acronym for decomposition of the time reversal operator) is a time reversal based method that can detect point-scatterers in a medium and extract their Green's function. The Green's functions can then be used to focus on the scatterers without knowledge of the medium. Applications of the method include autofocusing in inhomogeneous media, adaptive correction of aberration in medical images, detection of scatterers, or super-resolution.

The method consists first in a building the so-called *transfer matrix* $\boldsymbol{K}(\omega)$, at a given temporal frequency ω. The matrix coefficient $K_{ij}(\omega)$ is the signal received by array element i after a pulse has been transmitted by element j, as seen in Fig. 1. Practically $K_{ij}(\omega)$ is obtained by taking the Fourier transform (FT) of the broadband signal received by array element i. We omit the argument ω in the following. The time reversal operator is $\boldsymbol{K}\boldsymbol{K}^{\boldsymbol{H}}$, the product of the transfer matrix by its Hermitian transpose. In the DORT method, the singular value decomposition (SVD) of $\boldsymbol{K}$ is performed. It yields $\boldsymbol{K}=\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\boldsymbol{H}}$. U and V are unitary matrices containing the singular vectors, and $\boldsymbol{S}$ is a diagonal matrix containing the singular values. The singular vectors of $\boldsymbol{K}$ are the invariants of the time reversal operator.

In the case of rigid point-scatterers, the theory is well understood: The number of nonzero singular values is equal to the number of scatterers, and the singular vectors correspond to the scatterers Green's functions.[2] Focusing on each scatterer is achieved by backpropagation of the corresponding singular vectors, or Green's function.

A scatterer is considered pointlike if its size is much smaller than the resolution cell $\lambda Z/L$, where L is the size of the imaging aperture, and Z is the depth of the scatterer.

However, in media such as biological tissue, scatterers can rarely be considered as point-scatterers. These media are essentially constituted from incoherent scatterers (speckle) or *extended* coherent scatterers. Example of coherent scatterers includes vessels, microcalcifications, or kidney stones. These scatterers are typically the size of one or a few resolution cells.

The goal of this paper is to study the invariants of the time reversal operator in the case of such extended objects. Speckle has been addressed in another paper.[3]

Most of the results derived in this paper are already familiar to researchers in the field of scattering theory. That is because the time reversal operator is equivalent to a linear mapping from incident fields to scattered fields. The singular representation of the linear mapping has been extensively studied in electromagnetism[4] or optics.[5] The problem is also related to finding the number of degrees of freedom of an image.[6,7] Those papers, especially the one from Pierri *et al.*,[4] provides a complete mathematical treatment of the problem. However, the formalism in those papers is distant enough from our present problem to justify a new paper. None of them is dealing with acoustics fields or the time reversal operator formalism.

In Sec. II, a theory for small rigid objects is presented. The equations are written and solved for the case of the Fraunhofer approximation. In this case, the solutions are known as the prolate spheroidal functions.[8–10]

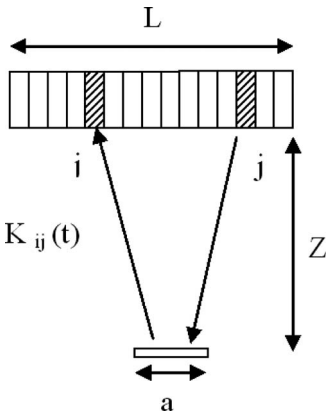


FIG. 1. Setup showing the array and the extended scatterer whose length is a. The coefficient $K_{i,j}$ of the transfer matrix is acquired by transmitting a pulse from element j and receiving with element i.

The theoretical solutions are compared to simulation results presented in Sec. III. In this section, we also consider the problem of Green's function estimation and focusing. We present a study identifying under which conditions good focusing can be achieved from the backpropagation of the first singular vector.

A theory of DORT with cylindrical objects has been proposed by Aubry *et al.*[11] They proposed a solution based on Hermite–Gaussian modes for scatterers much larger than the resolution cell and a solution based on Legendre polynomials for scatterers whose size is about the resolution cell. Our analysis is based on slightly different approximations and on flat objects, yielding the prolate spheroidal functions as solutions in both categories of scatterers. In Sec. IV, we will show that asymptotically our solutions approach the solutions of Aubry *et al.*

Finally, an application to super-resolution is presented in Sec. V.

II. THEORY

A. Expression of the time reversal operator for an extended object

1. Continuous formalism

For any given frequency, the signal received by the array, $\boldsymbol{R}(\omega)=[R_1(\omega),R_2(\omega),\ldots,R_M(\omega)]^T$ [$R_i(\omega)$ defined as the signal received by the ith element of the receive array, M being the number of elements, and T designating the transpose operation] is linked to the transmit signal $\boldsymbol{E}(\omega)=[E_1(\omega),E_2(\omega),\ldots,E_M(\omega)]^T$ by the transfer matrix[1,2] $\boldsymbol{K}$:

$$\boldsymbol{R}=\boldsymbol{K}\boldsymbol{E}. \tag{1}$$

The matrix coefficient $K_{ij}(\omega)$ is the signal received by array element i after a pulse has been transmitted by element j, as seen in Fig. 1.The time reversal operator is defined as $\boldsymbol{K}\boldsymbol{K}^{\boldsymbol{H}}$.

Equation (1) can be rewritten with explicit summation

$$R_i=\sum_{j=1}^{M}K_{ij}E_j, \tag{2}$$

where M is the number of elements. Using a continuous formalism is easier in this case. We introduce X_T and X_R, as azimuth of a point of the array in transmission and in reception, respectively. In the continuous formalism, X_T and X_R replace, respectively, the discrete indices j and i. Equation (2) becomes

$$R(X_R)=\int_{-L/2}^{L/2}K(X_R,X_T)E(X_T)dX_T, \tag{3}$$

where L is the length of the array and $K(X_R,X_T)$ is called the *kernel* of the integral equation, and is the equivalent of the matrix $\boldsymbol{K}$ in discrete formalism. The matrix coefficients are given by $\boldsymbol{K}_{ij}=K(X_i,X_j)$, where X_i is the azimuth position of element i. The discrete and continuous problems can be considered as equivalent if the sampling requirements are met, which is the case in most arrays.

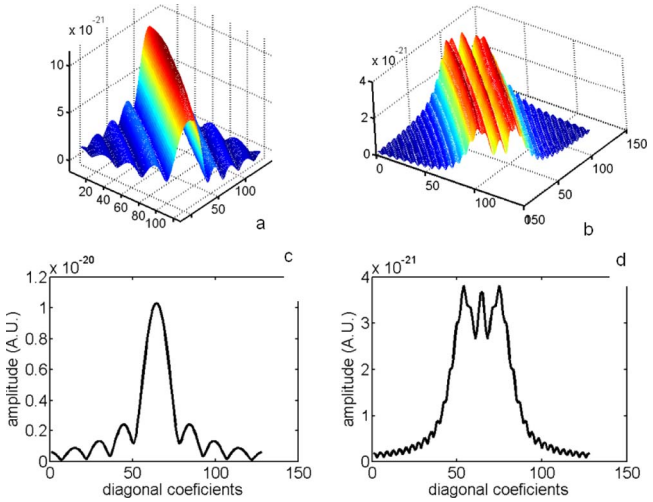


FIG. 2. (Color online) (a) Amplitude of the coefficients of the K matrix for an object of size a, such as $a^2<\lambda Z$ (Fraunhofer approximation). In this case, the amplitude is given by the FT of the object. For a rectangular object, this yields a sinc function. (b) Same plot for a larger object ($a^2=64$ mm, while $\lambda Z=10$ mm) The phase term in the Fresnel transform can no longer be neglected. Bottom: Amplitude along the diagonal in the two cases. The sinc pattern (c) and Fresnel transform of a square aperture pattern (d) are noticed.

2. Kernel in the Fresnel approximation

The kernel can easily be derived in the case of the Fresnel approximation. This is done in Appendix A. For an object whose scattering distribution is $d(x)$, we find [Eq. (A5)] that the kernel is, modulo a phase factor that depends only on Z,

$$K(X_R,X_T)=e^{j(\pi X_R^2/\lambda Z)}e^{j(\pi X_T^2/\lambda Z)}\times\mathrm{FT}\{d(x)e^{j(\pi x^2/\lambda Z)}\}\left[\frac{X_R+X_T}{\lambda Z}\right], \tag{4}$$

where FT stands for the Fourier transform. Equation (4) is related to the Fresnel transform of the object scattering distribution (where the Fresnel transform of a function $d(x)$ is defined as $\int d(x)e^{j\pi((X-x)^2/\lambda Z)}dx=e^{j(\pi X^2/\lambda Z)}\int d(x)e^{j(\pi x^2/\lambda Z)}\times e^{-j2\pi(xX/\lambda Z)}\,dx$). Basically, the amplitude of the kernel is given by the Fresnel transform of the object. The amplitude depends only on X_R+X_T, which gives the particular symmetry of the matrix with respect to the diagonal (anti-Toeplitz) (see Fig. 2).

3. Kernel for objects of size $a^2 < \lambda Z$

For an object small enough, $\mathrm{FT}\{d(x)e^{j(x^2/\lambda Z)}\}[(X_R+X_T)/\lambda Z]$ may be approximated by $\mathrm{FT}\{d(x)\}[(X_R+X_T)/\lambda Z]$. The variation of the parabolic phase term is negligible over a short distance. This approximation is known as the Fraunhofer approximation and will be considered true for the remaining of the derivations. We are assuming the Fraunhofer approximation in the object plane only, and not for the array plane. The condition for the object size is $a^2 < \lambda Z$. A typical depth in medical ultrasound is $Z=50$ mm, and typical wavelengths range from $\lambda=0.2$ mm (linear arrays) to $\lambda=0.5$ mm (phased arrays). In simulation, this approximation seems to be reasonable for scatterers whose size is $a<4$ mm (linear arrays) or $a<10$ mm (phased); this includes most coherent scatterer sizes in medical ultrasound, for example.

Under the Fraunhofer approximation, Eq. (4) becomes

$$K(X_R,X_T) = e^{j(\pi X_R^2/\lambda Z)}e^{j(\pi X_T^2/\lambda Z)}\mathrm{FT}\{d(x)\}\left[\frac{X_R+X_T}{\lambda Z}\right]. \quad (5)$$

The amplitude of the kernel is given by the FT of the object.

We also assume that the object has a rectangular scattering distribution (1 inside the object and 0 outside the object) and is centered on the axis. Its FT is then a sinc function. The expression of the kernel simplifies to

$$K(X_R,X_T) = e^{j(\pi X_R^2/\lambda Z)}e^{j(\pi X_T^2/\lambda Z)}\ \mathrm{sinc}\left[\frac{\pi a(X_R+X_T)}{\lambda Z}\right]. \quad (6)$$

In practice, the transfer matrix $\boldsymbol{K}$ is measured. As mentioned above, it is the discrete equivalent of the kernel. The matrix coefficients are therefore given by Eq. (6) under the Fraunhofer approximation. The amplitude of the coefficients of $\boldsymbol{K}$ is shown in Fig. 2 for two objects. The first one met the Fraunhofer approximation, and the sinc pattern is well noticeable. The second object is larger. In this case, the phase term in Eq. (4) can no longer be neglected and the amplitude is given by the Fresnel transform of the rectangular object.

B. Invariants of the time reversal operator: The prolate spheroidal functions

Finding the invariants of the time reversal operator is equivalent to computing the singular functions of the kernel $K(X_R,X_T)$ (in the continuous formalism) The singular functions of the kernel are a set of functions $E(X)$ and $R(X)$ constituting an orthonormal basis of the transmit (respectively, receive) arrays, and such that

$$\mu R(X_R) = \int_{-L/2}^{L/2} K(X_R,X_T)E(X_T)dX_T, \quad (7)$$

where μ is the associated singular value. Using Eq. (6), we obtain

$$\mu R(X_R)e^{j(-\pi X_R^2/\lambda Z)} = \int_{-L/2}^{L/2} \mathrm{sinc}\left[\frac{\pi a(X_R+X_T)}{\lambda Z}\right]E(X_T)e^{j(\pi X_T^2/\lambda Z)} \times dX_T - L/2 < X_R < L/2. \quad (8)$$

As the sinc kernel is Hermitian, Eq. (8) can be solved by choosing a function, $W(X,)$ such as

$$W(X) = R(X)e^{j(-\pi X^2/\lambda Z)} = E(-X)e^{j(\pi X^2/\lambda Z)}. \quad (9)$$

Equation (8) then becomes

$$\mu W(X_R) = \int_{-L/2}^{L/2} \mathrm{sinc}\left[\frac{\pi a(X_R+X_T)}{\lambda Z}\right]W(-X_T)dX_T = \int_{-L/2}^{L/2} \mathrm{sinc}\left[\frac{\pi a(X_R-X_T)}{\lambda Z}\right]W(X_T)dX_T - L/2 < X_R < L/2. \quad (10)$$

Equation (10) is a well-studied second order Fredholm equation, with Hermitian kernel, and the solutions have been derived by Slepian and Pollack.[8] They are called the prolate spheroidal wave-functions. These functions have also been studied in optics[9,10,12] in problems closely related to the one addressed in this paper. In another paper,[4] studying the number of degrees of freedom of radiated fields, they are used to solve problems mathematically similar to the present one, both in the Frauhhoffer and in the Fresnel approximation. The prolate spheroidal wave-functions depend on only one parameter $N=La/(\lambda Z)$, which can be interpreted as the number of resolution cell in the objects. Boyd[13] provided a good mathematical overview of the solution.

The main result is that the number of nonzero singular values is proportional to the parameter N, which in our case is the number of resolution cells that fit in the object. More precisely, the N first singular values are equal, and the following singular values drop rapidly to zero. This is an intuitive result. Indeed, as a generalization of the DORT theory for pointlike objects, we could expect the number of nonzero singular values to be the number of resolved points in the object. Thus the number of nonzero singular values is a measure of the object size.

The singular vectors (the prolate spheroidal functions) have the interesting property to be the invariants of the finite FT (FT limited to a finite aperture) In other words, eigenvectors in the array space or in the object space have the same shape, since the field in the object plane is given by the FT of the emitted signal [Eq. (A1)].

The prolate spheroidal functions are also the natural basis of the space of simultaneously band-limited and time-limited functions. No functions can be completely band limited and time limited simultaneously, but the prolates are the closest functions to fit this description. They are the most concentrated simultaneously in time and frequency. This is why they are of such interest in information theory, signal processing, and optics. The consequence of this property in our application is that they maximize the concentration of

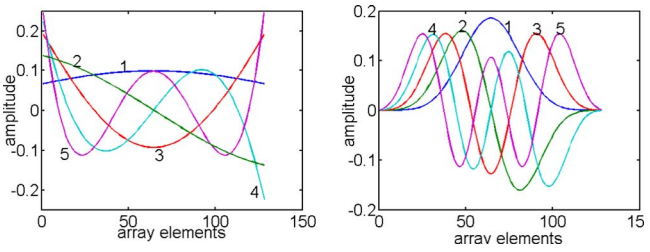


FIG. 3. (Color online) Five first spheroidal prolate functions amplitude for $N=1$ (left) and $N=10$ (right).

energy simultaneously in the array plane and the object plane. The five first prolate spheroidal functions are plotted in Fig. 3 for $N=1$ and $N=10$.

The invariants of the time reversal operator are derived from the prolate functions using Eq. (9):

$$R^i(X) = W^i(X)e^{j(\pi X^2/\lambda Z)},$$

$$E^i(X) = W^i(-X)e^{j(-\pi X^2/\lambda Z)} = W^i(X)e^{j(-\pi X^2/\lambda Z)}, \tag{11}$$

where $W^i(X)$ is the ith prolate function, and the last equality is due to the symmetry of the functions.

If the rectangular object is not centered on the axis, but still close enough so that the Fraunhofer approximation holds, the kernel in Eq. (6) contains a linear phase term $\exp(j(2\pi x_0(X_R+X_T)/\lambda z))$ corresponding to the offset between the object's center at $x=x_0$ and the axis. The solutions in Eq. (11) become $R^i(X)=W^i(X)e^{j(2\pi(X^2+x_0X)/\lambda Z)}$ and $E^i(X)=W^i(X)e^{j(-2\pi(X^2+x_0X)/\lambda Z)}$.

The phase of the singular vectors corresponds to a focusing on the object center (geometric delay).

These results are valid for small objects in the Fraunhofer approximation, such as $a^2<\lambda Z$. Solutions for larger objects, in the more general Fresnel approximation, are derived in Refs. 4 and 5. It is shown analytically that the number of nonzero singular values is still proportional to the number of resolution cell. The singular vectors can still be expressed using the prolate functions plus an additional phase term. However, the analytical form of the singular vectors is not very important in this case, because for large objects, the nonzero singular values are nearly identical and therefore the observed invariants are often a mixture of several theoretical singular vectors. In previous work,[4,14] it is shown in simulation that the result for the number of nonzero singular values being equal to the number of resolution cells can be generalized to the near-field as well. This is linked to more fundamental results about the number of degrees of freedom of the image of an object through a finite aperture.[6,7,15]

III. RESULTS

A. Results with a one-dimensional array

1. Object with $N=10$

An extended object has been simulated with field II.[16] The parameters were the following: $\lambda=0.2$ mm (center frequency), $a=4$ mm, $Z=50$ mm, and $L=25$ mm, which yields $N=La/\lambda Z=10$.

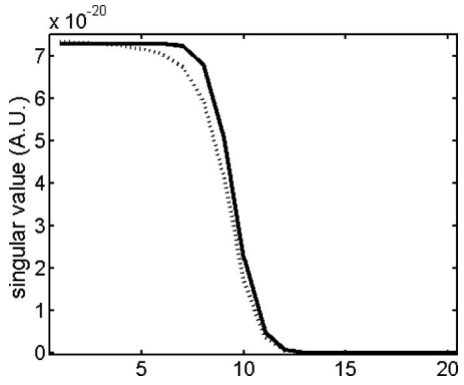


FIG. 4. Analytical (solid) and simulated (dashed) singular values for an object with $N=10$. The difference is likely due to the element directivity.

The theoretical singular values are compared with the simulation results in Fig. 4. There is a good agreement between theory and simulation. The number of nonzero singular values is about N, as predicted in Sec. II. In simulation, the singular values start decreasing earlier. This is due to the fact that the directivity of the elements was not taken into account in the theory. The directivity could be taken into account using perturbation theory.[11]

It is also interesting to look at the evolution of the singular values with the frequency in Fig. 5. The number of nonzero singular values increases versus the wavelength. This is shown in Fig. 5 (right) where the singular values have been normalized by the first one to remove the effect of the bandwidth. This is a typical property of prolate functions.[10] This is in agreement with the fact that resolution increases with frequency. Thus more resolution cells fit in the object.

A few singular functions obtained in simulation are compared to the prolate functions (analytical solutions) in Fig. 6. In practice, it is not always easy to observe the prolate with a rectangular object, as the first few singular vectors have nearly identical singular values. The shape of the singular functions is also in agreement with experimental results obtained by Aubry *et al.*[11]

2. Scatterers whose size is about the resolution cell ($N=1$) "resonance region"

Another important case are the scatterers whose sizes are close to the imaging wavelengths. This is called the *resonance region*. They are particularly hard to characterize in an

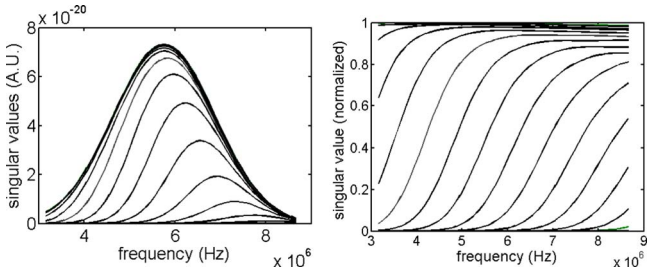


FIG. 5. (Color online) Left panel: Singular values as a function of the frequency. The bell shape is due to the bandwidth of the transducer. On the right panel, the singular values have been normalized to get rid of the bandwidth effect. As the frequency increases, more points are resolved in the object, and therefore the number of nonzero singular values increases.

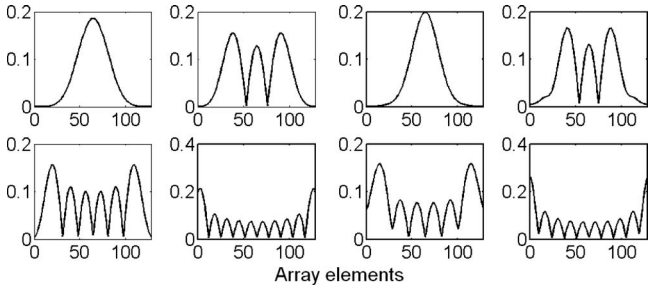


FIG. 6. Left: Absolute value of the amplitude of the prolate functions for $N=10$ (first, third, sixth, and tenth functions are shown). Right: Amplitude of the time reversal invariants obtained in simulation for an object such as $N=10$.

image. This can be the case for the microcalcifications. For the example of $N=1.2$, Figs. 7 and 8 show a very good agreement between simulation and theory for the singular values (Fig. 7) and for the amplitude of the singular vectors (Fig. 8).

B. Application to Green's function estimation and focusing

As seen in Eq. (11), the solution in homogeneous media is the homogeneous Green's function of the object center modulated by the prolate functions. The amplitude modulation originates from the interferences among several points of the object. The same result can be generalized to a heterogeneous medium, if the isoplanatic patch (the region where all the points *see* the same heterogeneity[17]) is about the size of the object or larger. In the heterogeneous media case, the invariants become

$$R^i(X) = W^i(X)e^{j(\pi X^2/(\lambda Z)+\phi(X))},$$

$$E^i(X) = W^i(X)e^{j(-\pi X^2/(\lambda Z)-\phi(X))}, \quad (12)$$

where $\phi(X)$ is the additional phase term due to the propagation through the heterogeneities. Indeed the signals received by different elements propagate through different media with different speeds of sound.[18,19] Equation (12) can be seen as the Green's function in heterogeneous media modulated by the prolate functions.

The estimation of Green's function in heterogeneous media is an important problem, since the Green's function can be used to focus on the corresponding point. Backpropagating a point Green's function can be seen as time reversing the signal coming from the point. When only point-scatterers are present, the DORT method is very efficient in determining Green's functions and focusing[1,2,20] without any knowledge on the medium. In speckle, the focused DORT (FDORT) method can be used.[3] Improving the focusing leads to better image quality, with better resolution and contrast. This is particularly important in medical imaging. This area is known as aberration correction.[3,21–23]

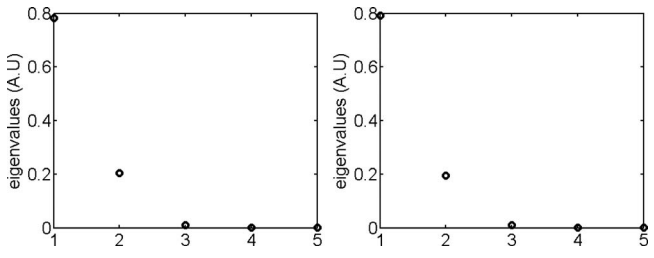


FIG. 7. Analytical (left) and simulated (right) singular values for an object in the so-called *resonance region* (the object is about the size of the resolution cell) with a parameter $N=1.2$ in this case. One singular value is dominant but a second singular value has also a relatively large value. This second singular value is important in order to differentiate the object from a point scatterer.

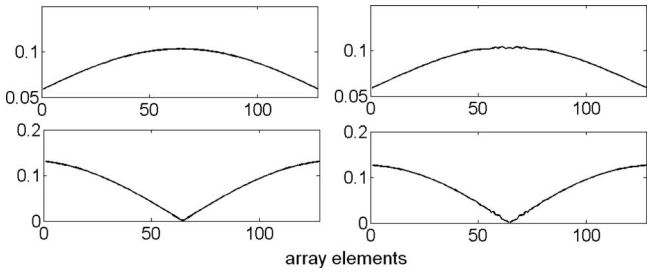


FIG. 8. Absolute value of the amplitude of the first two invariants for $N=1.2$. The first invariant is on top. Left: Analytical solutions (the two first prolate functions) Right: Simulation results. The agreement is very good.

As a result of the amplitude modulation, the Green's function cannot be estimated as precisely from extended objects as from point-scatterers. For example, if one considers the first invariant, one will have a good estimate of the Green's function for the central elements in the array where the amplitude is important, but no estimate for the outer elements where the amplitude is zero.

If the object is relatively small, e.g., one or two resolution cells, the width of the amplitude function is still close to the array width, as seen in Fig. 8 (top) and the effect is negligible. The first invariant can be considered almost as the Green's function of a point scatterer located at the center of the object. The focusing is illustrated in Fig. 9, where the field resulting from the backpropagation of the four first in-

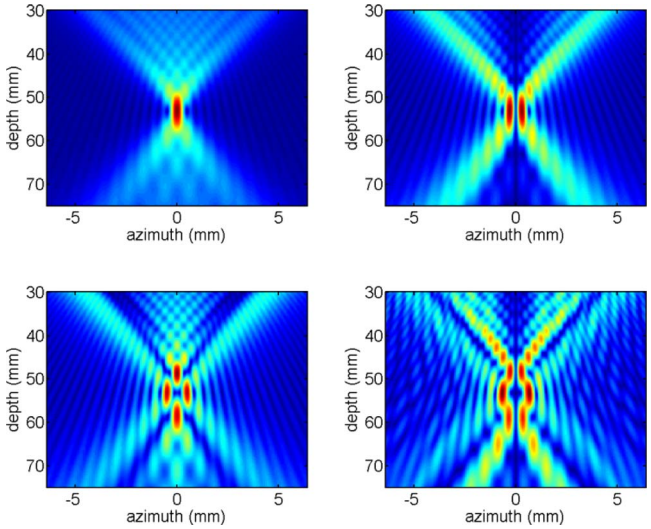


FIG. 9. (Color online) Field resulting from the backpropagation of the first four invariants for the small object (resonance region) The first invariant (top left) yields a focusing comparable to the Green's function of a point. The other invariants cannot be used for focusing because of the lobes.

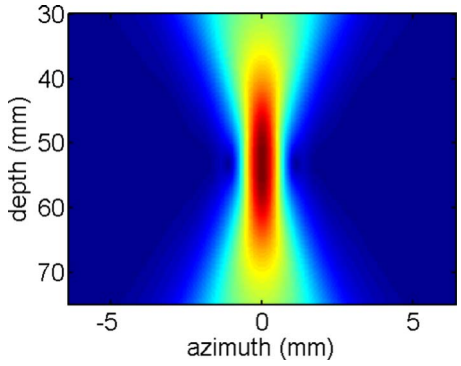


FIG. 10. (Color online) Backpropagation of the first invariants for the object with $N=10$. Because of the narrow amplitude of the first prolate function, the resolution is poor.

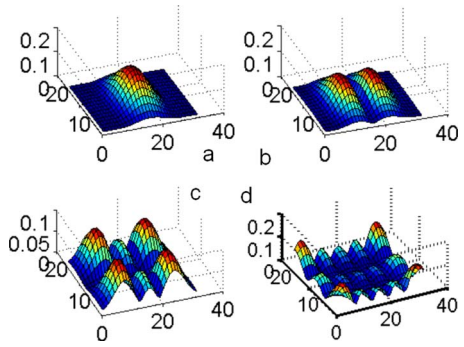


FIG. 11. (Color online) Selected invariants for a simulated rectangular plate imaged with a 2D array. The object size was roughly five resolution cells in azimuth ∗3 resolution cells in elevation. The invariants can be described by a Cartesian product of prolate functions. Here the modes are shown: (a) (1,1), (b) (1,2), (c) (2,3), and (d) (3,5).

variants for the small objects are shown. The first invariant leads to a sharp focus. The same focusing ability is observed in the presence of a phase aberration.

For a larger object, a loss of resolution occurs, because of the narrow amplitude factor seen in Fig. 6. As a result, the focal spot obtained by time reversal of the first invariant is broad. The field resulting from the backpropagation of the first invariant is shown in Fig. 10. The larger the object, the broader the focal spot will be.

Moreover, if the object is not flat but has a complicated shape, the kernel of Eq. (7) is no longer a sinc, and the solution W will likely be complex, and not real like the prolate. In this case, the phase of the invariants will be equal to the phase of the Green's function plus the phase of W, according to Eq. (11). The phase estimate is then biased, which cannot be easily compensated.

To conclude, an extended object can still be used to estimate Green's function if its size is relatively small, but larger objects, especially if they have a complex shape, may introduce an error in the estimate.

C. Results with a 2D array

Most of the new medical imaging ultrasound probes are two-dimensional arrays of transducers (2D array) that image a three-dimensional (3D) volume. It is then useful to look at the analytical solutions for some objects extended in the two dimensions (azimuth and elevation).

1. DORT with a 2D array

With a one-dimensional (1D) array, the emitted signal was described by a vector $\boldsymbol{E}$ whose coefficients correspond to each element of the array. With a 2D array, the vector needs *a priori* to be replaced by a matrix with coefficients $E_{i,j}$ with i referring to the azimuth and j referring to the elevation. Let I and J be the numbers of elements in the azimuth and elevation dimensions, respectively. For the DORT method, it is easier to index each pair of element (i,j) by a single number α, $1 \leq \alpha \leq I \cdot J$. Thus the emitted signal can still be represented by a vector with $I \cdot J$ coefficients E_α. The same is done in receive, and a transfer matrix K^{2D} whose size is $I \cdot J \times I \cdot J$ can be built. The SVD of K^{2D} is then performed. The singular vectors have $I \cdot J$ coefficients and can be remapped to the $I \times J$ 2D array.

2. Invariants

The solutions have a simple form if the object scattering distribution is separable in Cartesian coordinates, $d(x,y) = d(x)d(y)$. This is the case, for example, for a wire aligned with one of the array dimensions (we have then $d(x,y) = \delta(x)1(y)$, where $\delta(x)$ is the Dirac function) or for a rectangular plate $(d(x,y) = \text{box}(x)\text{box}(y)$, where box is the box function) In the case where the scattering distribution is separable, the kernel is also separable (this results from the separability of the FT). Due to symmetry, the solutions are separable as well, and we have $R(X,Y) = R_X(X)R_Y(Y)$, where R_X and R_Y are solutions to the respective 1D equations along each array dimension, and $E(X,Y) = E_X(X)E_Y(Y)$. This is derived in Appendix B.

The number of singular values is $N_x N_y$ (where $N_x = L_x . a_x / \lambda Z$ with L_x and a_x are, respectively, the array and object sizes in the x dimension and N_x is the number of resolution cell in the x dimension), i.e., the number of 2D resolution cells in the object.

Selected invariants for a 2D rectangular plate imaged with a 2D array are shown in Fig. 11.

IV. COMPARISON WITH PREVIOUS WORK

Aubry *et al.*[11] proposed a solution based on Hermite–Gaussian modes for scatterers much larger than the resolution cell (when the edge effects of the array are negligible), and a solution based on Legendre polynomials for scatterers whose size are about the resolution cell (in this case, the edge effect is dominant).

We can find these functions as asymptotic limits of our solutions. We consider here the 1D case for simplicity. For a small parameter N (the number of resolution cell in the object), or near the edges, the prolate functions can be approximated by Legendre polynomials.[13]

Similarly, for large N, the prolate functions can be well approximated by Hermite–Gaussian modes, with a width equal to $1/\sqrt{N}$.[13]

$$W^i(X) \approx e^{-NX^2/2} H^i(\sqrt{N}X),$$

where H^i is the ith Hermite polynomial.

The interest in the prolate functions is that they avoid the transition between the Legendre polynomials and the Hermite–Gaussian mode; there is a continuum. However, the solutions based on Hermite–Gaussian modes seem to be valid on a larger range of scatterers that do not necessarily meet the Fraunhofer approximation. They cover the solutions derived in Appendix B. In addition, the Hermite–Gaussian modes in Ref. 11 can describe a cylindrical object, and not only a flat object as in our case.

The prolate functions are better suited for scatterers in the Fraunhofer regimen, in particular, in the transition region between Legendre and Hermite–Gaussian solutions. Another advantage of our solution is the result about the number of nonzero eigenvalues. This could not be derived directly in Ref. 11 because the finite size of the array was not taken into account.

V. APPLICATION TO SUPER-RESOLUTION IMAGING

The problem with imaging scatterers whose size is about the resolution cell is that they are difficult to differentiate from pure point-scatterers on a conventional image. This is backed up by the fact that they have one dominant singular value. However, the other singular values are small but not completely equal to zero and carry information on the object size and shape. The second singular value is clearly nonzero in Fig. 7, but even the following singular values are nonzero when plotted on a logarithmic scale. The aim of inverse methods is to use and amplify the information in the other singular vectors. However, the noise limits the possible resolution because the information contained in the singular vectors associated with too small singular values is corrupted by the noise.[12] Also, the accuracy of the model of the medium used in the reconstruction seriously limits the resolution.[24] The inverse methods do not allow to reconstruct the missing information lost during the propagation (high spatial frequencies) and do not violate the law of physics. Rather, they work by assuming additional information on the object; in general, the assumption is that the edges are sharp. For the algorithm considered below, the assumption is that the variance is minimal.

Most inverse methods are based on the decomposition of the covariance matrix, and a link between covariance matrices and time reversal operator has been done by Gruber *et al.*[25] Therefore, the methods can be used directly with the time reversal invariants. One of the algorithm is the minimum variance, also known as Capon, algorithm. An expression of the minimum variance algorithm using DORT's eigenvector is given in Ref. 24:

$$I_{ML}(P) = \frac{1}{\sum_{i=1}^{M} \frac{1}{\mu^i} |\langle U^i | G(P) \rangle|}, \tag{13}$$

where μ^i is the ith singular values, U^i is the associated singular vectors, and $G(P)$ is the Green's function of a point P, computed from a model of the medium.

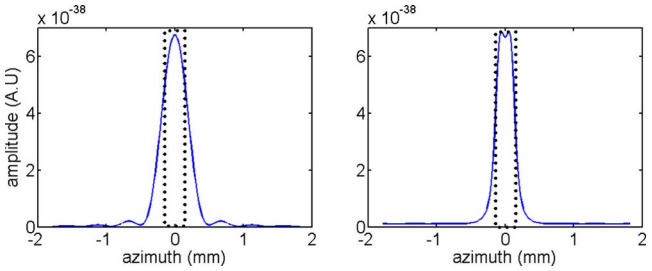


FIG. 12. (Color online) Profile of the object at the center frequency. It is the object convolved by the beam pattern (sinc). It is then a low-pass filtered version of the object (propagation act as a low-pass filter). Right: Profile obtained with the minimum variance algorithm. The real object is in dashed line.

As an illustration, the minimum variance method has been used with a small scatterer ($N=1.2$), in a homogeneous medium. The profile at the focal depth has been reconstructed using the usual backpropagation method, and with the minimum variance method (Fig. 12). Then the same result is shown for a point-scatterer (Fig. 13). A constant term was added to each singular values in Eq. (13) to decrease the sensitivity to very low singular values (as the inverse of the singular values is taken, low singular values have a high impact on the result) and therefore to improve the robustness. The point- and extended-scatterers cannot be distinguished on the conventional profile but are clearly resolved with the minimum variance method.

These methods are good for increasing the resolution in the azimuth dimension, but not in the axial dimension (they are monochromatic methods). This is why some researchers use them with an enclosing array, or cavity, in a setup similar to tomography. Mast *et al.*[26] and Waag *et al.*[27] used a formalism similar to ours (but in spherical coordinate and plane wave approximations). Their inverse method is also based on variance minimization and accounts for multiple scattering. More recent work account for heterogeneities.[28] Lewis *et al.*[29] used an advanced inverse scattering method from the electromagnetic inverse community to determine the shape of microcalcifications and differentiate between clutter and single microcalcifications. Another method used to achieve super-resolution is the MUSIC algorithm.[7,14,24] In Ref. 14, it is applied to a similar problem of imaging an extended object.

VI. CONCLUSION

The invariants of the time reversal operator for a small flat object, whose size a is such as $a^2 < \lambda Z$, are the prolate

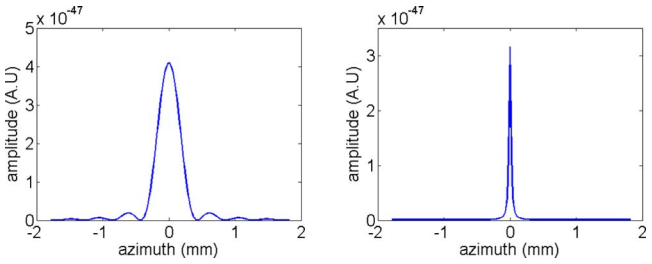


FIG. 13. (Color online) Same as Fig. 12, but for a point scatterer. With the conventional profile, it is very difficult to distinguish between the point and the extended object. With the minimum variance method, the difference is obvious, and it is easier to assess the size of the scatterer.

spheroidal functions. The number of nonzero singular values is roughly equal to the number of resolution cells in the object and therefore is a measure of the object size.

If the object is small enough (a few resolution cells), the first invariant is a good estimate of the Green's function and can be used for focusing. For larger objects, the estimate is only good for central elements and a loss of resolution occurs.

It has been shown that the solutions previously proposed by Aubry *et al.*[11] are asymptotic limits of the prolate functions. The prolate solutions are particularly well suited for the Fraunhofer regimen. The Hermite–Gaussian modes of Aubry *et al.*[11] may be better suited for larger objects, and when the objects are cylindrical.

Super-resolution methods such as the Capon estimator can be used to enhance the information contained in the invariants associated with low singular values. This can be used to differentiate between a point-scatterer and an object whose size is about the wavelength.

The problems described here have been well studied in scattering theory, in electromagnetism[4] or optics. The time reversal operator is equivalent as a forward linear mapping from incident fields to scattered fields. It is related to the more general problem of degrees of freedom. Further results can be found in the corresponding literature.[4–7,15]

The derivations in this paper assumed a rigid scatterer. For elastic scatterer, it has been shown,[30,31] in the case of pointlike objects, that additional singular vectors arises, corresponding to dipolar Green's functions. This is probably the case for extended objects as well. In Ref. 14, the MUSIC algorithm is modified to account for both monopolar and dipolar Green's functions.

APPENDIX A: DERIVATION OF THE KERNEL IN THE FRESNEL APROXIMATION

X_T and X_R are the azimuths of a point of the array, in transmission and in reception, respectively. x is the azimuth of a point in the object plane. Let $d(x)$ be the scattering distribution of the object, located at depth Z. The setup is shown in Fig. 1. First, a signal $E(X_T)$ is transmitted by the array. We consider here that $x \ll Z$ and $X \ll Z$. Therefore, the Fresnel approximation (see Chap. 4 in Ref. 32) can be used to express the pressure field $P_i(x)$ received by the object as the Fresnel transform of $E(X_T)$:

$$P_i(x) = \frac{e^{j\pi(Z^2+x^2/\lambda Z)}}{Z}\int_{\text{aperture}} E(X_T)e^{j(\pi X^2/\lambda Z)}e^{-j2\pi(xX/\lambda Z)}dX_T$$

$$= Ke^{j(\pi x^2/\lambda Z)}\text{FT}\{E(X_T)e^{j(\pi X_T^2/\lambda Z)}\}\left[\frac{x}{\lambda Z}\right], \tag{A1}$$

where $K=\frac{e^{j\pi(Z^2/\lambda Z)}}{Z}$ depends only on Z, and $\text{FT}\{f(X)\}[x/\lambda Z]$ stands for the FT of a function $f(X)$ at frequency $x/\lambda Z$. Interpreting the finite integral as a FT is valid if $E(X_T)$ is considered to be equal to zero outside of the aperture. The pressure backscattered by the object, expressed in the object plane at depth Z, is given by

$$P_b(x) = d(x)P_i(x), \tag{A2}$$

and finally the signal received by the array is given by the Fresnel transform of $P_b(x)$:

$$R(X_R) = Ke^{j(\pi X_R^2/\lambda Z)}\text{FT}\{P_b(x)e^{j(\pi x^2/\lambda Z)}\}\left[\frac{X_R}{\lambda Z}\right]$$

$$= K^2e^{j(\pi X_R^2/\lambda Z)}\text{FT}\left\{d(x)e^{j(\pi x^2/\lambda Z)}\text{FT}\{E(X_T)\right.$$

$$\left.\times e^{j(\pi X_T^2/\lambda Z)}\}\left[\frac{x}{\lambda Z}\right]\right\}\left[\frac{X_R}{\lambda Z}\right]. \tag{A3}$$

Using the relationship relating the FT of a product of functions to the convolution product of the FT of the functions

$$\text{FT}\{f(x)g(x)\}[X] = \text{FT}\{f(x)\}[X] \otimes \text{FT}\{g(x)\}[X],$$

where $\otimes$ is the convolution product, with $f(x) = d(x)e^{j(\pi x^2/\lambda Z)}$ and $g(x) = \text{FT}\{E(X_T)e^{j(\pi X_T^2/\lambda Z)}\}\left[\frac{x}{\lambda Z}\right]$, and noting that

$$\text{FT}\left\{\text{FT}\{E(X_T)e^{j(\pi X_T^2/\lambda Z)}\}\left[\frac{x}{\lambda Z}\right]\right\}\left[\frac{X_T}{\lambda Z}\right]$$

$$= E(-X_T)e^{j(\pi X_T^2/\lambda Z)}$$

yields

$$R(X_R) = e^{j(\pi X_R^2/\lambda Z)}\int_{-L/2}^{L/2}\text{FT}\{d(x)e^{j(\pi x^2/\lambda Z)}\}\left[\frac{X_R+X_T}{\lambda Z}\right]$$

$$\times E(X_T)e^{j(\pi X_T^2/\lambda Z)}dX_T$$

$$-L/2 < X_R < L/2 \tag{A4}$$

the right-hand inequality for X_R, as well as the bounds for the integration, coming from the fact that no signal is received (respectively, transmitted) outside of the aperture. Finally,

$$K(X_R,X_T) = e^{j(\pi X_R^2/\lambda Z)}e^{j(\pi X_T^2/\lambda Z)}$$

$$\times\text{FT}\{d(x)e^{j(\pi x^2/\lambda Z)}\}\left[\frac{X_R+X_T}{\lambda Z}\right]. \tag{A5}$$

The factor K of Eq. (A3) has been omitted for clarity purposes.

APPENDIX B: INVARIANTS FOR 3D OBJECTS WITH 2D ARRAYS

1. Equation for the 2D problem

In two dimensions, the singular value equation becomes

$$\mu R(X_R,Y_R)$$

$$= e^{j\pi(X_R^2+Y_R^2)/\lambda Z}\int_{\text{array}}\int \text{kernel2D}\left[\frac{X_R+X_T}{\lambda Z},\frac{Y_R+Y_T}{\lambda Z}\right]$$

$$\times E(X_T,Y_T)e^{j\pi(X_T^2+Y_T^2)/\lambda Z}dX_TdY_T$$

$$-L/2 < X_R,Y_R < L/2, \tag{B1}$$

where kernel 2D is the 2D Fresnel transform of the object or the FT in the Fraunhofer approximation.

$$\text{kernel 2D} = \text{FT}\left\{d(x,y)e^{j\frac{\pi(x^2+y^2)}{\lambda Z}}\right\}\left[\frac{X}{\lambda Z},\frac{Y}{\lambda Z}\right]. \quad \text{(B2)}$$

2. Analytical solutions for separable kernels

Solutions of Eq. (B1) have a simple form if the object reflectivity function is separable in Cartesian coordinates $d(x,y)=d(x)d(y)$. In this case, the kernel is also separable (property of the FT). For symmetry reason, the solutions are separable, too. $R(X,Y)=R_X(X)R_Y(Y)$ and $E(X,Y)=E_X(X)E_Y(Y)$. Equation (B1) finally becomes

$$\mu R_X(X_R)e^{j\pi X_R^2/\lambda Z}R_Y(Y_R)e^{j\pi Y_R^2/\lambda Z} = \int \text{kernel}_X\left[\frac{X_R+X_T}{\lambda Z}\right]E_X(X_T)e^{j\pi X_T^2/\lambda Z}dX_T \times \int \text{kernel}_Y\left[\frac{Y_R+Y_T}{\lambda Z}\right]E_Y(Y_T)e^{j\pi Y_T^2/\lambda Z}dY_T, \quad \text{(B3)}$$

and the equations for the X and Y quantities can be solved independently. For X, it yields

$$\mu R_X(X_R)e^{j\pi\frac{X_R^2}{\lambda Z}} = \int \text{kernel}_X\left[\frac{X_R+X_T}{\lambda Z}\right]E_X(X_T)e^{j\pi X_T^2/\lambda Z}dX_T,$$

with $\text{kernel}_X=\text{FT}\{d(x)e^{j(\pi x^2/\lambda Z)}\}[X/\lambda Z]$. That is exactly the equation we have for a 1D object with reflectivity $d(x)$. There is a similar equation for Y.

The solutions of Eq. (B3) are then the products of the solutions of the 1D equations. The singular value is the product of the singular values of the two 1D problems.

[1] J.-L. Robert, M. Burcher, C. Cohen-Bacrie, and M. Fink, "Time reversal operator decomposition with focused transmission and robustness to speckle noise: Application to microcalcification detection," J. Acoust. Soc. Am. **119**, 3848–3859 (2006).
[2] C. 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).
[3] J. L. Robert and M. Fink, "Green's function estimation in speckle using the decomposition of the time reversal operator," J. Acoust. Soc. Am. **123**, 866–877 (2008).
[4] R. Pierri and F. Soldovieri, "On the information content of the radiated fields in the near zone over bounded domains," Inverse Probl. **14**, 321–337 (1998).
[5] F. Gori and G. Grella, "The converging prolate spheroidal functions and their use in Fresnel optics," Opt. Commun. **45**, 5–10 (1983).
[6] G. Toraldo di Francia, "Degrees of freedom of an image," J. Opt. Soc. Am. **59**, 799–804 (1969).
[7] F. Gori and G. Guattari, "Shannon number and degrees of freedom of an image," Opt. Commun. **7**, 163–165 (1973).
[8] D. Slepian and H. O. Pollack, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, I," Bell Syst. Tech. J. **40**, 43–64 (1961).
[9] J. W. Goodman, "Statistical properties of speckle pattern," in *Laser Speckle and Related Phenomena*, edited by J. C. Dainty (Springer-Verlag, Berlin, 1975), pp. 9–75.
[10] J. W. Goodman, *Statistical Optics* (Wiley Classics Library, New York, 1985).
[11] A. Aubry, J. d. Rosny, J.-G. Minonzio, C. Prada, and M. Fink, "Gaussian beams and Legendre polynomials as invariants of the time reversal operator for a large rigid cylinder," J. Acoust. Soc. Am. **120**, 2746–2754 (2006).
[12] B. R. Frieden, *Evaluation, Design, and Extrapolation Methods for Optical Signals Based on Use of the Prolate Functions*, Progress in Optics (North Holland, Amsterdam, 1971), Vol. **IX**, pp. 311–407.
[13] J. P. Boyd, "Prolate spheroidal wavefunctions as an alternative to Chebyshev and Legendre polynomials for spectral element and pseudospectral algorithms," J. Comput. Phys. **199**, 688–716 (2004).
[14] S. Hou, K. Solna, and H. Zhao, "A direct imaging method using far-field data," Inverse Probl. **23**, 1533–1546 (2007).
[15] D. Gabor, "Light and information," *Progress in Optics* (North-Holland, Amsterdam, 1961), pp. 109–153 (2007).
[16] J. A. Jensen, "Field: A program for simulating ultrasound systems," Med. Biol. Eng. Comput. **34**, 351–353 (1996).
[17] J. J. Dahl, M. S. Soo, and G. E. Trahey, "Spatial and temporal aberrator stability for real-time adaptive imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1504–1517 (2005).
[18] L. M. Hinkelman, T. D. Mast, L. A. Metlay, and R. C. Waag, "The effect of abdominal wall morphology on ultrasonic pulse distortion. Part I. Measurements," J. Acoust. Soc. Am. **104**, 3635–3649 (1998).
[19] T. D. Mast, L. M. Hinkelman, M. J. Orr, and R. C. Waag, "The effect of abdominal wall morphology on ultrasonic pulse distortion. Part II. Simulations," J. Acoust. Soc. Am. **104**, 3651–3664 (1998).
[20] C. Prada, S. Manneville, D. Spoliansky, and M. Fink, "Decomposition of the time reversal operator: Detection and selective focusing on two scatterers," J. Acoust. Soc. Am. **99**, 2067–2076 (1996).
[21] M. O'Donnell and S. W. Flax, "Phase-aberration correction using signals from point reflectors and diffuse scatterers: Measurements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 768–774 (1988).
[22] S. W. Flax and M. O'Donnell, "Phase-aberration correction using signals from point reflectors and diffuse scatterers: Basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 758–767 (1988).
[23] R. C. Waag and J. P. Astheimer, "Statistical estimation of ultrasonic propagation path parameters for aberration correction," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 851–869 (2005).
[24] C. Prada and J. L. Thomas, "Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix," J. Acoust. Soc. Am. **114**, 235–243 (2003).
[25] F. K. Gruber, E. A. Marengo, and A. J. Devaney, "Time-reversal imaging with multiple signal classification considering multiple scattering between the targets," J. Acoust. Soc. Am. **115**, 3042–3047 (2004).
[26] T. D. Mast, A. I. Nachman, and R. C. Waag, "Focusing and imaging using eigenfunctions of the scattering operator," J. Acoust. Soc. Am. **102**, 715–725 (1997).
[27] R. C. Waag, D. L. Liu, T. D. Mast, and A. I. Nachman, "Imaging with eigenfunctions of the scattering operator," J. Acoust. Soc. Am. **101**, 3091 (1997).
[28] R. Waag, "Image reconstruction using eigenfunctions of a scattering operator," J. Acoust. Soc. Am. **120**, 3023 (2006).
[29] M. A. Lewis, P. Liu, E. Richer, P. P. Antich, and S. G. Johnson, "Linear sampling method for acoustic inverse scattering in breast microcalcification detection," J. Acoust. Soc. Am. **120**, 3025 (2006).
[30] J.-G. Minonzio, C. Prada, D. Chambers, D. Clorennec, and M. Fink, "Characterization of subwavelength elastic cylinders with the decomposition of the time-reversal operator: Theory and experiment," J. Acoust. Soc. Am. **117**, 789–798 (2005).
[31] D. H. Chambers and A. K. Gautesen, "Time Reversal for a single spherical scatterer," J. Acoust. Soc. Am. **109**, 2616–2624 (2001).
[32] J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, New York, 1996).

A phase and space coherent direct imaging method

Songming Hou[a)]
Mathematics and Statistics, Louisiana Tech University, Ruston, Louisiana 71272

Kai Huang[b)]
Department of Mathematics, Florida International University, Miami, Florida 33199

Knut Solna[c)] and Hongkai Zhao[d)]
Department of Mathematics, University of California at Irvine, Irvine, California 92697

(Received 23 May 2008; revised 20 October 2008; accepted 1 November 2008)

A direct imaging algorithm for point and extended targets is presented. The algorithm is based on a physical factorization of the response matrix of a transducer array. The factorization is used to transform a passive target problem to an active source problem and to extract principal components (tones) in a phase consistent way. The multitone imaging function can superpose multiple tones (spatial diversity/aperture of the array) and frequencies (bandwidth of the probing signal) based on phase coherence. The method is a direct imaging algorithm that is simple and efficient since no forward solver or iteration is needed. Robustness of the algorithm with respect to noise is demonstrated via numerical examples.
 [DOI: 10.1121/1.3035835]

PACS number(s): 43.60.Tj, 43.60.Pt [OAS] Pages: 227–238

I. INTRODUCTION

In reflection seismology, ultrasound imaging in medical applications, detection of defects in nondestructive testing, underground mine detection and target detection using radar or a sonar system, and so on, one seeks to identify the location and shape of some scatterers by sending probing waves and measuring the scattered waves, e.g., using scattering relations. This is in general an ill-posed (nonlinear) inverse problem. Imaging the whole medium using a general inverse problem approach may be too complicated and too expensive to be practical in many applications, for instance, if the imaging domain is large compared to the wavelength. If the background medium is homogeneous and some simple boundary condition is satisfied at the boundary of the target, the inverse problem can be turned into a geometric problem, that is, the problem of determining the shape of the target from the scattered wave field pattern. Using nonlinear optimization approach in this case is still difficult and computationally expensive.

Direct imaging methods, which are not based on nonlinear optimization and hence do not require forward solver or iterations, have attracted a lot of attention recently. If the targets are small compared to the array resolution, the location information can be obtained while the geometry information is not resolved. Several matched filter type of algorithms have been developed for imaging or locating point targets, for example, the multiple signal classification (MUSIC) algorithm.[1–4] Under the assumption of point targets the response matrix (defined in Sec. II) has a simple structure. This structure is used in MUSIC and has also been exploited to focus a wave field on selected scatterers using iterated time reversal.[5–10] The iterated time reversal procedure corresponds to the power method for finding the dominant singular vectors for the response matrix. However, with the point target assumption, physical properties and the geometry of the target are neglected. More importantly an extended target is not a superposition of point targets. For extended targets the response matrix has a more complicated structure. Recently a few MUSIC type of algorithms[11–14] have been developed to image the location and shape of extended targets. A crucial step is to use resolution and noise level based thresholding to determine how many singular vectors of the response matrix span the signal space.

Although the generalized MUSIC algorithm for a single frequency is capable of imaging different types of targets with efficiency, robustness, and accuracy, provided full aperture data are given, for limited aperture the results are typically not very good. Multiple frequencies should be used to complement the lack of spatial aperture.

The MUSIC algorithm is based on the singular value decomposition (SVD) of the response matrix. This decomposition allows for an arbitrary complex phase; therefore, combining different frequencies in a phase coherent way is not direct. In this paper, we propose a multitone imaging algorithm that makes use of coherent information in both phase and space. In particular, we take advantage of phase coherence from multiple frequency data to improve both resolution of robustness of the imaging procedure. The crucial points in our multitone algorithm are (1) physically based factorization of the response matrix that transforms a passive target detection problem to an active source detection problem and (2) a phase coherent imaging function that can superpose multiple tones and multiple frequencies to take advantage of both spatial diversity (aperture) of the array

[a)]Electronic mail: shou@latech.edu
[b)]Electronic mail: khuang@fiu.edu
[c)]Electronic mail: ksolna@math.uci.edu
[d)]Electronic mail: zhao@math.uci.edu

and/or the bandwidth of the probing signal. The proposed method can be parallelized easily since the evaluation of the imaging function at different grids is independent.

The outline of this paper is as follows. In Sec. II we describe how to locate point targets using a method that we call the multitone algorithm. In Sec. III, we generalize the method to imaging of extended targets. Numerical experiments are presented in Sec. IV.

II. RESPONSE MATRIX AND IMAGING POINT TARGETS

Our imaging setup uses an array of transmitters that can send out probing waves into the region of interest and an array of receivers that can record scattered waves. Our measurement data are the response matrix whose elements are the inter-responses between array elements. The arrays can enclose the region of interest (full aperture) or can have partial aperture. For simplicity, we first consider an active array when the array of transmitters and the array of receivers coincide, moreover, time harmonic waves. Assume that there are N transducers, which can function both as a transmitter and as a receiver, and that are located at $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_N$. The P_{ij} element of the response matrix P is the received signal at transducer j for a probing pulse sent out from transducer i. Assume that there are M point targets located at $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_M$ with reflectivity $\tau_1,\ldots,\tau_M$. The response matrix in the Born approximation has the following simple structure using spatial reciprocity:

$$P_{ij} = \sum_{m=1}^{M} \tau_m G^0(\boldsymbol{x}_m,\boldsymbol{\xi}_i) G^0(\boldsymbol{\xi}_j,\boldsymbol{x}_m)$$

$$= \sum_{m=1}^{M} \tau_m G^0(\boldsymbol{\xi}_i,\boldsymbol{x}_m) G^0(\boldsymbol{\xi}_j,\boldsymbol{x}_m),$$

where $G^0(\boldsymbol{x},\boldsymbol{y})$ is free space Greens function and we suppress the dependence on frequency. In matrix form we have

$$P = \sum_{m=1}^{M} \tau_m \mathbf{g}_m \mathbf{g}_m^T,$$

where

$$\mathbf{g}_m = [G^0(\boldsymbol{\xi}_1,\boldsymbol{x}_m), G^0(\boldsymbol{\xi}_2,\boldsymbol{x}_m), \ldots, G^0(\boldsymbol{\xi}_N,\boldsymbol{x}_m)]^T,$$

$m=1,2,\ldots,M$, are called illumination vectors, each of which corresponds to the received signals at the array for a point source at $\boldsymbol{x}_m$. For an active array, the response matrix is square and symmetric with rank M in general. If the targets are well resolved by the transducer array, i.e., the separation distance between the targets is larger than the resolution of the array, we have that the point spread function

$$\Gamma(\boldsymbol{x}_m,\boldsymbol{x}_{m'}) = \overline{\mathbf{g}}_m^T \mathbf{g}_{m'} \approx 0 \quad \text{if } m \neq m',$$

which means the wave field corresponding to the time reversal of a point source at one target is almost zero at all other targets. Hence $\hat{\mathbf{g}}_m = \mathbf{g}_m / \|\mathbf{g}_m\|$ and its complex conjugate $\overline{\hat{\mathbf{g}}}_m$ can be regarded as the left and right singular vectors for the response matrix P. In general this one to one correspondence does not exit. However, one can show that $\mathbf{g}_m$, $m=1,2,\ldots,M$ still span the signal space of P even if multiple scattering among point targets is present based on the Foldy–Lax formulation.[1]

Remark. Here we assume the simplest model for point scatters. In general, a point scatterer may induce both monopole and dipole for the scattered field. For example, the scattered field for an acoustic point scatterer is the sum of monopole (by contrast in compressibility) and dipole (by contrast in density).[15,16] Our formulation and imaging function only use the monopole component which works for hard scatterers. We should be able to modify our imaging function to take into account dipoles which will be discussed in our future work.

To motivate our imaging algorithm consider first the case with a point source at the mth scatterer location; the vector of observations at the transducer array is then

$$\mathbf{g}(\boldsymbol{x}_m) = [G^0(\boldsymbol{\xi}_1,\boldsymbol{x}_m), G^0(\boldsymbol{\xi}_2,\boldsymbol{x}_m), \ldots, G^0(\boldsymbol{\xi}_N,\boldsymbol{x}_m)]^T.$$

Phase conjugation at the mirror and backpropagation to the imaging domain correspond to forming the imaging function $I_m(\boldsymbol{x}) = \mathbf{g}(\boldsymbol{x}_m)^H \mathbf{g}(\boldsymbol{x})$, where $\boldsymbol{x}$ is a search point in the domain and the superscript H denotes the transpose and complex conjugate. Note that physical *time reversal* corresponds to phase conjugation in frequency domain and then backtransformation to time domain. In the inverse problem setting, although $\boldsymbol{x}_m$ is unknown, an estimate of $\hat{\mathbf{g}}(\boldsymbol{x}_m)$ can be obtained (up to a constant phase) via the SVD of the response matrix. The imaging function will peak at the source location $\boldsymbol{x}_m$ due to phase coherence; $I_m(\boldsymbol{x}_m) = \|\mathbf{g}(\boldsymbol{x}_m)\|^2$. In particular, if we use normalized $\hat{\mathbf{g}}(\boldsymbol{x})$ and $\hat{\mathbf{g}}(\boldsymbol{x}_m)$ in the above imaging function, it is an optimal matched filter.[17–19] Classic Rayleigh resolution theory gives that $I_m(\boldsymbol{x})$ will be supported in the neighborhood of the source-point $\boldsymbol{x}_m$ with a lateral resolution of order $\lambda L/a$. Here $\lambda = c_0/\omega$ is the wavelength, L is the distance from the array to the source, a is the aperture of the array, and c_0 denotes the speed of propagation.

We compute the SVD of the response matrix to extract dominant singular vectors (tones). This matrix factorization corresponds to turning passive targets into imaging sources for the scattered wave. However, the SVD of a matrix is unique up to a complex phase, e.g., if the following is a SVD of P:

$$P = \sum_m \sigma_m \mathbf{u}_m \mathbf{v}_m^H,$$

where $\mathbf{u}_m(\mathbf{v}_m)$ are the unit left (right) singular vectors and σ_m are the singular values of P, then $e^{i\theta_m}\mathbf{u}_m(e^{i\theta_m}\mathbf{v}_m)$ are also left (right) singular vectors for arbitrary θ_m, $m=1,2,\ldots$. To overcome the arbitrary phase in the SVD, we propose the following modified imaging function for each pair of left and right singular vectors $\mathbf{u}_m$ and $\mathbf{v}_m$, which we call a tone of the response matrix:

$$I_m(\boldsymbol{x}) = [\hat{\mathbf{g}}^H(\boldsymbol{x})\mathbf{u}_m][\hat{\mathbf{g}}^H(\boldsymbol{x})\overline{\mathbf{v}}_m].$$

First, this imaging function removes the phase ambiguity of the SVD of the response matrix. Second, for well resolved point targets,

$$I_m(\boldsymbol{x}) = [\hat{\mathbf{g}}^H(\boldsymbol{x})\hat{\mathbf{g}}(\boldsymbol{x}_m)]^2.$$

Note that by "squaring" in this way, instead of using norm square, we maintain the phase information, e.g., the phase information is just linearly doubled. Next, we superpose the dominant tones for the different frequencies to obtain the general form of the multitone imaging function:

$$I^M(\boldsymbol{x}) = \sum_{\omega} \alpha(\omega) \sum_{m=1}^{M^\omega} [\hat{\mathbf{g}}^H(\boldsymbol{x};\omega)\mathbf{u}_m^\omega][\hat{\mathbf{g}}^H(\boldsymbol{x};\omega)\bar{\mathbf{v}}_m^\omega]. \quad (1)$$

We remark that for an active array with the transmitters and receivers coinciding, the response matrix P is complex symmetric and can be factorized as $P = U\Sigma U^T$. The imaging function then becomes

$$I^M(\boldsymbol{x}) = \sum_{\omega} \alpha(\omega) \sum_{m=1}^{M^\omega} [\hat{\mathbf{g}}^H(\boldsymbol{x};\omega)\mathbf{u}_m^\omega]^2. \quad (2)$$

In the general case when the transmitters and receivers do not coincide, e.g., there are s transmitters located at $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_s$ and there are r receivers located at $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_r$, the response matrix is of dimension $s \times r$. The ijth element P_{ij} then records the response at jth receiver for a signal sent out from ith transmitter. Define the illumination vector with respect to the receiver array and transmitter array, respectively, as

$$\mathbf{g}_r(\boldsymbol{x}) = [G^0(\boldsymbol{\eta}_1,\boldsymbol{x}), G^0(\boldsymbol{\eta}_2,\boldsymbol{x}), \ldots, G^0(\boldsymbol{\eta}_r,\boldsymbol{x})]^T$$

and

$$\mathbf{g}_s(\boldsymbol{x}) = [G^0(\boldsymbol{\xi}_1,\boldsymbol{x}), G^0(\boldsymbol{\xi}_2,\boldsymbol{x}), \ldots, G^0(\boldsymbol{\xi}_s,\boldsymbol{x})]^T.$$

The response matrix in the Born approximation has the following form in the case of M point targets located at $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ with reflectivity $\tau_1, \ldots, \tau_M$:

$$P = \sum_{m=1}^{M} \tau_m \mathbf{g}_s(\boldsymbol{x}_m)\mathbf{g}_r^T(\boldsymbol{x}_m).$$

Thus, the column and row space of P is spanned by $\mathbf{g}_s(\boldsymbol{x}_m)$ and $\mathbf{g}_r^T(\boldsymbol{x}_m)$, respectively. Accordingly the multitone imaging function is constructed as

$$I^M(\boldsymbol{x}) = \sum_{\omega} \alpha(\omega) \sum_{m=1}^{M^\omega} [\hat{\mathbf{g}}_s^H(\boldsymbol{x};\omega)\mathbf{u}_m^\omega][\hat{\mathbf{g}}_r^H(\boldsymbol{x};\omega)\bar{\mathbf{v}}_m^\omega]. \quad (3)$$

Here $\hat{\mathbf{g}}$ denotes the normalized illumination vector.

The frequency weight function $\alpha(\omega)$ can in principle be chosen to reflect the signal to noise ratio (SNR) of different frequencies. However, here we will not discuss this issue and use a uniform weighting. Note second that M^ω is the number of significant tones which may vary with frequency. If there are M point targets that are well resolved by all the frequencies used, then $M^\omega = M$. In general, e.g., for extended targets, M^ω may be proportional to the resolution of frequency ω.[11,20] In particular, when there is strong noise present, e.g., low SNR, M^ω is an important thresholding (regularization) parameter.[11] An important strength of the multitone algorithm is that it is quite robust with respect to the choice of M^ω. In particular, when the noise level is low, we can choose it to coincide with the smaller dimension of the response matrix. This is not the case for, for instance, the MUSIC algorithm, described below, whose imaging result depends more sensitively on the thresholding.

We summarize by stating that two important features of the multitone imaging algorithm are as follows:

1. The SVD factorization of the response matrix turns a passive target detection problem into an active source detection problem. The principal component (tone) decomposition of the response matrix takes the full array into account simultaneously and extracts dominant information or "tones" via the SVD, giving a robust imaging scheme.
2. The imaging function exploits coherent phase information via superposition of complex tones.

Next, we compare our multitone imaging algorithm with two other popular imaging algorithms. For simplicity we discuss the active array case.

1. MUSIC. The MUSIC imaging function[4] is based on the projection to the signal space spanned by dominant singular vectors, which is equivalent to the following quantity:

$$\sum_{m=1}^{M} |\hat{\mathbf{g}}^H(\boldsymbol{x})\mathbf{u}_m|^2,$$

where M is the dimension of the signal space that is determined according to the resolution and/or the SNR analysis.[11,20] However, phase information is lost after projection and hence it is difficult to superpose multiple frequencies based on phase coherence. Thus, travel time information is not effectively utilized in this implementation of the MUSIC imaging functional.
2. Kirchhoff migration. The Kirchhoff migration is a time domain method which exploits travel time information between pairs of transducers. After Fourier transform it can be approximated in frequency domain[21] by the following:

$$\tilde{\mathbf{g}}^H(\boldsymbol{x})P\bar{\tilde{\mathbf{g}}}(\boldsymbol{x}) = \tilde{\mathbf{g}}^H(\boldsymbol{x})\left[\sum_{m=1}^{N} \sigma_m \mathbf{u}_m \mathbf{v}_m^H\right]\overline{\tilde{\mathbf{g}}(\boldsymbol{x})},$$

where $\tilde{\mathbf{g}}(\boldsymbol{x})$ is Green's function without the spatial decaying factor $(1/\|\boldsymbol{x}\|)$ and N is the number of transducers. The above formula is similar to Eq. (1) for each frequency except the following two main differences: (1) our imaging function does not contain the weighting by the singular values and (2) our imaging function introduces a thresholding/regularization based on resolution and/or SNR. These two differences mean that the multitone imaging algorithm only separates signal space from noise space and treats all dominant signals, i.e., dominant singular vectors, equally. The motivation comes from the following observations: a point target partially blocked by other targets will contribute a singular vector with smaller singular values. So our multitone imaging function will increase the visibility of partially blocked targets compared to Kirchhoff migration, as shown in Figs. 2 and 3 in Sec. IV A. Also an extended target is not

a superposition of point targets. For example, it is illustrated in Ref. 11 that each singular vector does not correspond to a point on the boundary of an extended target. The geometry of the boundary is embedded in the signal space spanned by the dominant singular vectors collectively. Hence the multitone imaging function will serve to give a uniform illumination of the visible parts of the boundary.

In summary our multitone imaging algorithm takes advantages of both approaches in a natural way. Like MUSIC, our algorithm is based on the SVD of the response matrix and a resolution and/or SNR based thresholding to extract principal components (tones) from the full array information. The principal components are used collectively and in a uniform weighting situation. This is particularly important for imaging extended targets. The response matrix for an extended target can have many principal components.[12,16,22–26] The collection of all these principal components contains information about the extended target. On the other hand, instead of using a projection operator as in MUSIC, we use a propagation operator as in the Kirchhoff method which maintains coherent phase information and allows linear superposition of different tones (components) and multiple frequencies. Only at locations with strong scattering are phases of different tones and different frequencies in our imaging function coherent. Like in the Kirchhoff method travel time information is thus utilized in our imaging function. Our approach is based on gaining robustness via using the SVD of the response matrix to extract coherent information and is extremely simple to implement. We remark that other recent approaches like the coherent interferometric (CINT) method[21] aim at extracting information via carefully screened cross correlation computations of the observations. This approach has been shown to work well in a strongly heterogeneous environment but is less direct in its implementation.

III. EXTENDED TARGETS

A. Dirichlet boundary condition

We consider the situation with an extended target. First, let us assume a Dirichlet boundary condition for the target, i.e., a sound-soft target. Let Ω denote the target and Ω^c the exterior of the target. Let $G_\Omega(\boldsymbol{x},\boldsymbol{y})$ be associated Green's function that solves

$$\Delta G_\Omega(\boldsymbol{x}) + k^2 G_\Omega(\boldsymbol{x}) = \delta(\boldsymbol{x}-\boldsymbol{y}), \quad \boldsymbol{x},\boldsymbol{y} \in \Omega^c \subset R^m,$$

$$G_\Omega(\boldsymbol{x},\boldsymbol{y}) = 0, \quad \boldsymbol{x} \in \partial\Omega,$$

and a far field radiation boundary condition. The scattered field at transducer $\boldsymbol{\xi}_j$ corresponding to a point source at $\boldsymbol{\xi}_i$ follows from Greens formula and is

$$P_{ij} = \int_{\partial\Omega} G^0(\boldsymbol{\xi}_i,\boldsymbol{y}) \frac{\partial G_\Omega(\boldsymbol{\xi}_j,\boldsymbol{y})}{\partial \nu} d\boldsymbol{y},$$

where G^0 is free space Green's function. A physical interpretation is that the source of the scattered wave field is a weighted superposition of monopoles at the boundary. The response matrix can be written as

$$P = \int_{\partial\Omega} \mathbf{g}(\boldsymbol{y}) \left[\frac{\partial \mathbf{g}_\Omega(\boldsymbol{y})}{\partial \nu} \right]^T d\boldsymbol{y}, \tag{4}$$

where $\mathbf{g}(\boldsymbol{y})$ is the illumination vector for the homogeneous background, which is known, and $\mathbf{g}_\Omega(\boldsymbol{y})$ is the illumination vector

$$g_\Omega(\boldsymbol{y}) = [G_\Omega(\boldsymbol{\xi}_1,\boldsymbol{y}), \ldots, G_\Omega(\boldsymbol{\xi}_N,\boldsymbol{y})]^T,$$

which is unknown.

Equation (4) gives a factorization of the response matrix that separates the known and unknown components. Thus, the response matrix is superposed from illumination vector $\mathbf{g}(\boldsymbol{y})$, where $\boldsymbol{y}$ belongs to the illuminated parts of the boundary, e.g., where $\partial_{\mathbf{g}\Omega}(\boldsymbol{y})/\partial_\nu$ is not small. Therefore, we apply SVD to the response matrix to extract the singular vectors $\mathbf{u}_m(\mathbf{v}_m)$ and then use the imaging functions (2) for symmetric active array and (3) if the transmitter array and receiver array are different. This imaging function will peak at the well illuminated parts of the boundary. Physically, the peak can be explained by the fact that the boundary acts as a source for the scattered field; thus also, iterated time reversal, i.e., power method for finding singular vectors, will give focusing on the boundary. The thresholding strategy for extended targets introduced in Ref. 11 can be used to determine the thresholding parameter by an optimal cutoff.

We remark that the unknown weight function (illumination strength) for the monopoles at the target boundary, $\partial_{\mathbf{g}\Omega}(\boldsymbol{y})/\partial_\nu$, is not uniform in general due to geometry of the target, such as singularities and concavity of the boundary, and/or the array configuration, such as illumination angles and partial aperture. Locations on the boundary with stronger wave field, i.e. better illuminated by the source, have more weights. These factors will be reflected by the magnitude of singular values for different singular vectors. In our multitone algorithm each principal component will be given an equal weight as long as its corresponding singular value is above the noise threshold. That is why our imaging function gives a fairly uniform intensity on the well illuminated boundary. This is an important aspect of our approach: by taking out the scaling of the tones by the singular values we focus on the geometrical aspects of the extended scatterer and compensate for differences in relative illumination strength. Thus differential parts of the boundary are imaged with a similar fidelity.

B. Neumann boundary condition

For a sound-hard target, with a Neumann boundary condition for the extended target the response matrix has the form

$$P = -\int_{\partial\Omega} \left[\frac{\partial \mathbf{g}(\boldsymbol{y})}{\partial \nu} \right] \mathbf{g}_\Omega^T(\boldsymbol{y}) d\boldsymbol{y}.$$

In other words, the source of the scattered wave field is an (unknown) weighted superposition of dipoles $\partial \mathbf{g}(\boldsymbol{y})/\partial\nu$ at the boundary. Therefore, the normal direction is part of the un-

known in the imaging function. As is done in Ref. 11 we will incorporate a direction search in our imaging function, e.g., among a fixed collection of discretized directions, μ_j, $j=1,2,\ldots$, we maximize the imaging function among these directions at a searching point $\boldsymbol{x}$. Our multitone imaging function is then in the general case:

$$I^M(\boldsymbol{x}) = \max_j \left| \sum_{\omega} \alpha(\omega) \sum_{m=1}^{M^\omega} \left[\frac{\partial \hat{\mathbf{g}}_s^H(\boldsymbol{x};\omega)}{\partial \nu_j} \mathbf{u}_m^\omega \right] \left[\frac{\partial \hat{\mathbf{g}}_r^H(\boldsymbol{x};\omega)}{\partial \nu_j} \bar{\mathbf{v}}_m^\omega \right] \right|.$$

C. Limited or synthetic aperture

For single frequency and full aperture the MUSIC algorithm typically works better than multitone. However, for limited aperture or synthetic aperture with multiple frequency data MUSIC may fail while multitone can work well. We demonstrate below that the multitone algorithm works well also in a case with limited or synthetic aperture.

D. Far field data

In Sec. II the response matrix is defined in terms of near field data, with the sources and receivers in near field. In some applications, the measurement data are far field data, that is, the incident field is essentially a plane wave and the far field pattern of the scattered field is recorded.

We now discuss briefly the case for far field data. For Dirichlet boundary condition, the element of the response matrix P_{ij} corresponds to the far field pattern of the scattered field in the jth direction due to an incident wave coming from the ith direction:

$$P_{ij} = u_\infty(\hat{\boldsymbol{\theta}}_j;\hat{\boldsymbol{\theta}}_j) = \beta \int_{\partial\Omega} \frac{\partial u}{\partial \nu}(\boldsymbol{y};\hat{\boldsymbol{\theta}}_i) e^{-ik\hat{\boldsymbol{\theta}}_j\cdot\boldsymbol{y}} d\boldsymbol{y},$$

where the total field u is due to incident plane wave coming from the direction $\hat{\boldsymbol{\theta}}_i$, where $\beta=-1/4\pi$ for three dimensions and $\beta=-e^{i\pi/4}/\sqrt{8\pi|k|}$ for two dimensions.

In matrix form

$$P = \beta \int_{\partial\Omega} \frac{\partial \vec{u}}{\partial \nu} \hat{\mathbf{g}}^H(\boldsymbol{y}) d\boldsymbol{y}, \tag{5}$$

where

$$\hat{\mathbf{g}}(\boldsymbol{y}) = [e^{ik\hat{\boldsymbol{\theta}}_1\cdot\boldsymbol{y}}, \ldots, e^{ik\hat{\boldsymbol{\theta}}_n\cdot\boldsymbol{y}}]^T,$$

and $\vec{u}$ is the vector of total fields corresponding to the incident plane waves from $\hat{\boldsymbol{\theta}}_1,\ldots,\hat{\boldsymbol{\theta}}_n$. Equation (5) gives a physical factorization of the scattered field into known and unknown parts. The far field pattern is a superposition of the far field patterns of point sources located on the boundary of the target; however, we do not know the weight function which depends on the total field. In other words, the scattering at the target boundary acts as "sources" for the scattered field. In this far field setup, it is natural to use $\hat{\mathbf{g}}(\boldsymbol{y})$ as the illumination vector as discussed in Ref. 12. The signal space of the response matrix should be well approximated by the span of the illumination vectors $\hat{\mathbf{g}}(\boldsymbol{y})$ with $\boldsymbol{y}$ on the well-illuminated part of the boundary of the targets. Hence, we only need to change the form of illumination vectors in the multitone imaging function. Neumann type of boundary conditions can also be dealt in a similar fashion as in the case with near field data. See Ref. 12 for more details.

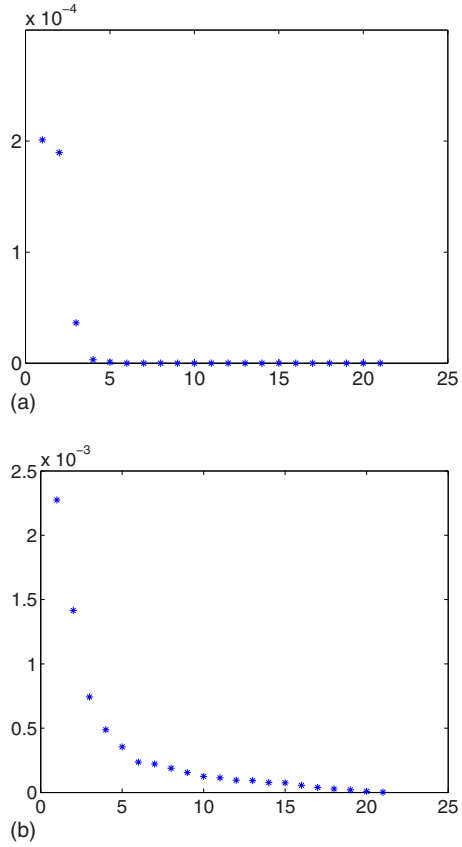


FIG. 1. (Color online) SVD pattern of the response matrix.

IV. NUMERICAL EXPERIMENTS

A. Point targets

First we show a few examples for point targets, targets that are small compared to the resolution of the array. The examples are two dimensional (2D) experiments and simulations.

In the first numerical test, there are three targets with a range of 30–40λ (central wavelength). The linear active array is located at the left side and is composed of 21 transducers that are half wavelength apart, i.e., the aperture is 10λ. The three targets are of size 0.5λ each. We tested this setup in both homogeneous and random media. The weakly heterogeneous medium has a 5% standard deviation and the correlation length is $O(\lambda)$. Finite difference method is used to solve the Helmholtz equation with perfectly matched layer (PML) technique[27] for 21 frequencies that are equally distributed between 0.9λ and 1.1λ with equal weight. The size of grid in numerical scheme is $\lambda/10$ in the 2D rectangle domain. The star shows the true location of targets.

Figure 1 shows the SVD pattern for a fixed frequency (0.9λ) for these two cases. In the homogeneous medium, there are only three dominant singular values. However, the three singular vectors may not have a one to one correspondence to the illumination vector of the three targets due to

multiple scattering among the targets. In particular, the reflected wave from the rear scatterer will be mixed with the reflection of the two front ones.

The numerical data used for imaging targets in heterogeneous medium are the scattered wavefield by the target and the background heterogeneous medium, i.e., the difference of the two wavefields corresponding to the medium with targets and the homogeneous medium, respectively. The goal is to image dominant scatterers/targets without imaging or knowing the details of the background medium, which is very desirable in many practical applications. The situation is also more difficult than using the difference data, i.e., measuring the difference of the two wavefields corresponding to the medium with targets and the same medium without targets, respectively. Figures 2 and 3 show the imaging results using multitone imaging algorithm using different number of frequencies and different number of leading singular vectors. It shows clearly that

- superposition of coherent phases from multiple frequencies improves range resolution;
- using the leading three singular vectors (the best SNR thresholding) produces the best results; however, the imaging result is not very sensitive to thresholding; and
- the partially blocked target has a better visibility compared to the Kirchhoff migration for the reason discussed in Sec. II.

As shown in Fig. 1, the SVD pattern is more complicated in random medium due to multipathing. Figure 4 shows the imaging results using the multitone imaging algorithm, which demonstrates the following:

- The location information of three point targets is not included in the first three singular vectors.
- Involving more singular vectors, even without thresholding, works well since only strong scattering at targets is superposed coherently (in phase) across different frequencies.
- Again the partially blocked target has a better visibility compared to the Kirchhoff migration.

Finally we test our algorithms on real experimental data. The data were kindly provided by Daniel D. Stancil and his group at Carnegie Mellon University. In their experimental setup, transmit array A and receive array B are used, as shown in Fig. 5. The locations of transmitters and receivers are different. The measurements were taken at 201 frequency points ranging from 4 to 6 GHz. An absorbing wall is located behind the test scenario.

Figure 6 shows imaging using data with different numbers of targets. The stars in each figure are the true locations of targets.

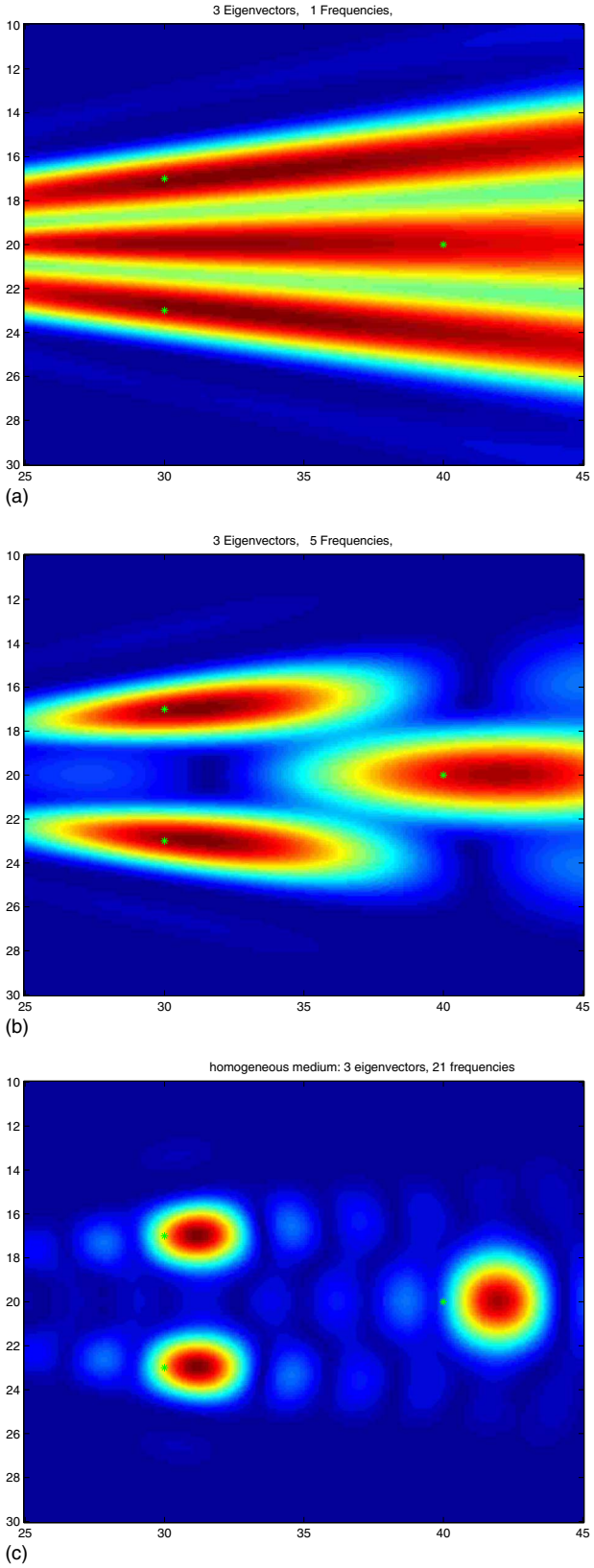


FIG. 2. (Color online) Imaging point targets in homogeneous medium. Multitone algorithm using 3 leading singular vectors and (a) 1 frequency, (b) 5 frequencies, and (c) 21 frequencies.

B. Extended targets

In this section we test our multitone imaging algorithm on extended targets with full aperture, limited aperture, and synthetic aperture using near and far field data. All near field data are simulated by solving the Helmholtz equation using finite difference method with PML (Ref. 27) boundary condition. For inverse problems, the forward solver is not required to be very accurate, we did not use any special treatment at the target boundary, i.e., the standard five point

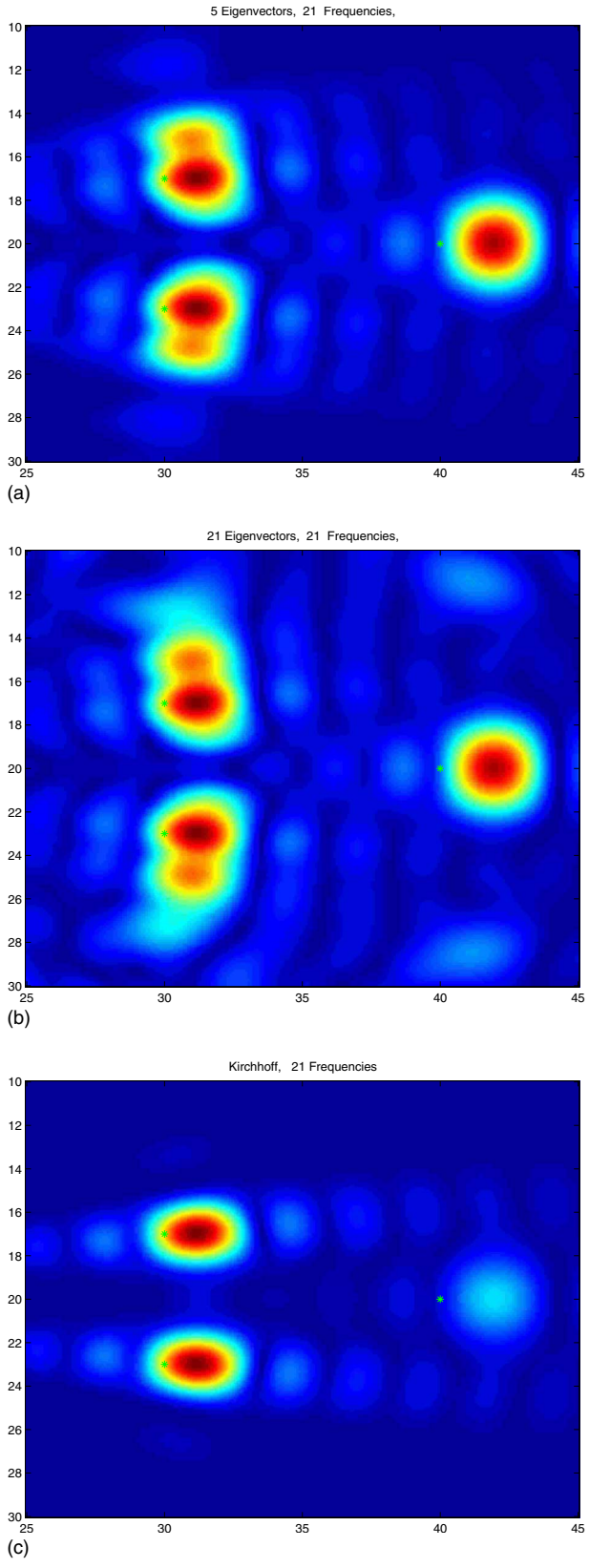


FIG. 3. (Color online) Imaging point targets in homogeneous medium. Multitone algorithm using 21 frequencies and (a) 5 leading singular vectors, (b) 21 singular vectors, and (c) Kirchhoff algorithm using 21 frequencies.

stencil for centered difference is used at every grid node. Far field data are generated using a boundary integral method.[12,28] The simulations are in 2D.

We give two examples with full or limited aperture near field data, two examples with synthetic aperture near field data, one example with full aperture far field data, one example where sources and receivers do not coincide, and finally, one example with limited aperture far field data. For all near field experiments, the transducers are about $200h$ (200 grid cells) away from the target and the forward data are again generated using a finite difference method with the PML technique.[27] The multiplicative noise is modeled by $P_{\text{noisy}}(i,j)=\text{Re}(P(i,j))a+\text{Im}(P(i,j))b$, where a and b are uniformly distributed in $[1-c,1+c]$, where c is 10%. The random medium or *clutter* is modeled as follows: The index of refraction $n(x)$ is a Gaussian with mean one and standard deviation 10% and the correlation length is $10h$, which is comparable but less than the wavelength.

Figure 7 shows imaging of a single extended target in a homogeneous medium. The full circular active array has 80 transducers surrounding the target. When a single frequency is used the corresponding wavelength is $\lambda=16h$. When three frequencies are used, they correspond to wavelengths, $\lambda=16h,24h,32h$. The target is about $200h$ away from the array and its size is about $80-100h$. In this test, no thresholding is used in the multitone imaging function. It is clear that phases across different frequencies are superposed coherently at the boundary only, where strong scattering happens. Figure 8 shows imaging of the same target with limited aperture data. Only half of the circular array from the bottom is used.

Figure 9 shows the imaging of a sound-hard (Neumann boundary condition) target with full aperture circular array (80 transducers). The array is about $200h$ from the center of the target. Six equally spaced frequencies are used, with lowest frequency $\lambda=32h$ and highest frequency $\lambda=16h$.

We next test with synthetic aperture. We use the following implementation of synthetic aperture. Let P be an 80-by-80 response matrix corresponding to an active circular array with full aperture (as above), and $P_n=Q(n:n+19,n:n+19)$, where $n=1,11,21,\ldots,61$, then the P_n's are the 20-by-20 response matrices with limited aperture and a partial overlap. We use P_n at the same six frequencies as above.

Figure 10 shows the multitone imaging function with synthetic aperture data for a sound-soft (Dirichlet boundary condition) object. The top one is the result for homogeneous medium with clean simulated data. For the middle one, 10% multiplicative noise is added to the data. The bottom one shows imaging in a random medium with 10% standard deviation. The correlation length is about a wavelength.

In contrast, the MUSIC algorithm does not provide a good imaging function for limited/synthetic aperture data. Figure 11 shows the result using the MUSIC algorithm with synthetic aperture data. The kite shape is not clear.

Finally we test the multitone imaging algorithm using far field data. The only change made is in the form of the illumination vector, i.e., using the far field pattern of Green's function. Figure 12 shows the multitone imaging function for far field data with clean simulated data (left) and with 100% multiplicative noise added to the simulated data (right).

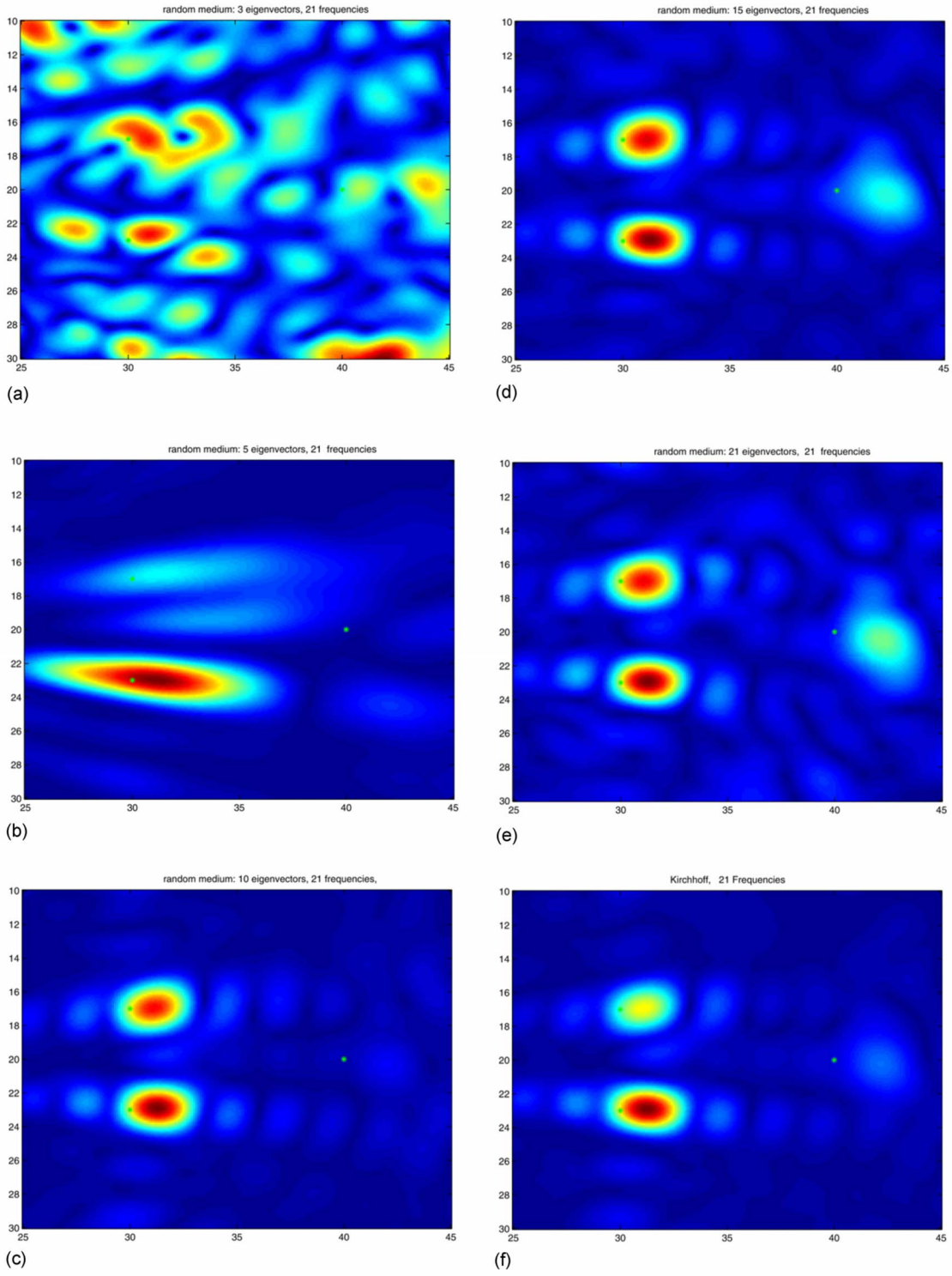


FIG. 4. (Color online) Imaging point targets in random medium. Multitone algorithm using 21 frequencies and (a) 3, (b) 5, (c) 10, (d) 15, and (e) 21 singular vectors, and (f) Kirchhoff algorithm using 21 frequencies.

Three wave numbers are used, $k=5,6,7$, so that the target sizes are on the scale of the wavelength. The forward data are here generated using the boundary integral method. In this case 32 plane incident waves are used and the far field data are collected at the same 32 directions.

We remark that the thresholding strategy discussed in Ref. 11 is used for all the above examples of extended targets except the first one. In principle, with thresholding only the first few dominant singular vectors are used in the multitone imaging function. This is known to be robust as long as the leading singular values are well separated from the remainder. However, in our tests the results are not very sensitive to the thresholding, which means that the multitone imaging function is already quite robust and is easier to use in prac-

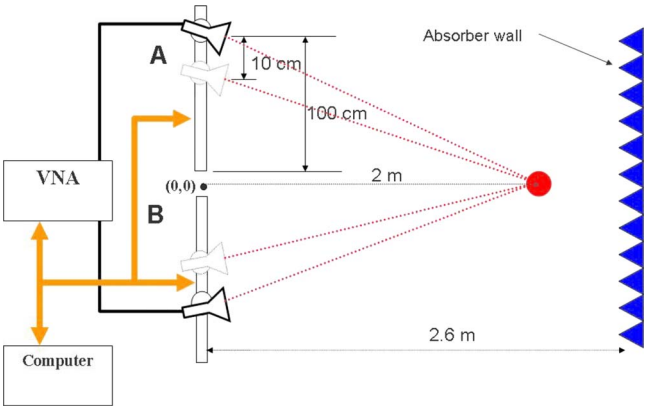


FIG. 5. (Color online) CMU experiment setup.

tice. Thus we also expect that our imaging results are robust with respect to numerical errors and artifacts by our numerical scheme that generates the data.

In the last set of tests, we show imaging with arrays that have transmitters different from receivers, or plane wave incident angles different from far field data angles. Figure 13 shows the multitone imaging function using wave numbers $k=5,6,7$ for far field data with plane wave incident from the right (16 directions) and far field pattern recorded on the left (16 directions). Dirichlet boundary condition is used.

Again, for limited aperture only part of the boundary that is well illuminated is seen in the imaging function. Figure 14 shows the multitone imaging function using wave numbers $k=5,6,7$ for far field data with limited aperture, that is, only plane waves in a 180 deg angle are used (16 directions) and the far field data within the same angle are recorded. Only the part of the kite boundary that is well illuminated by the array can be observed in the imaging function.

V. CONCLUSIONS

We propose a direct imaging algorithm, the multitone method. The algorithm is simple and efficient because no forward solver or iteration is needed. This method provides a framework for balancing spatial diversity via the SVD with

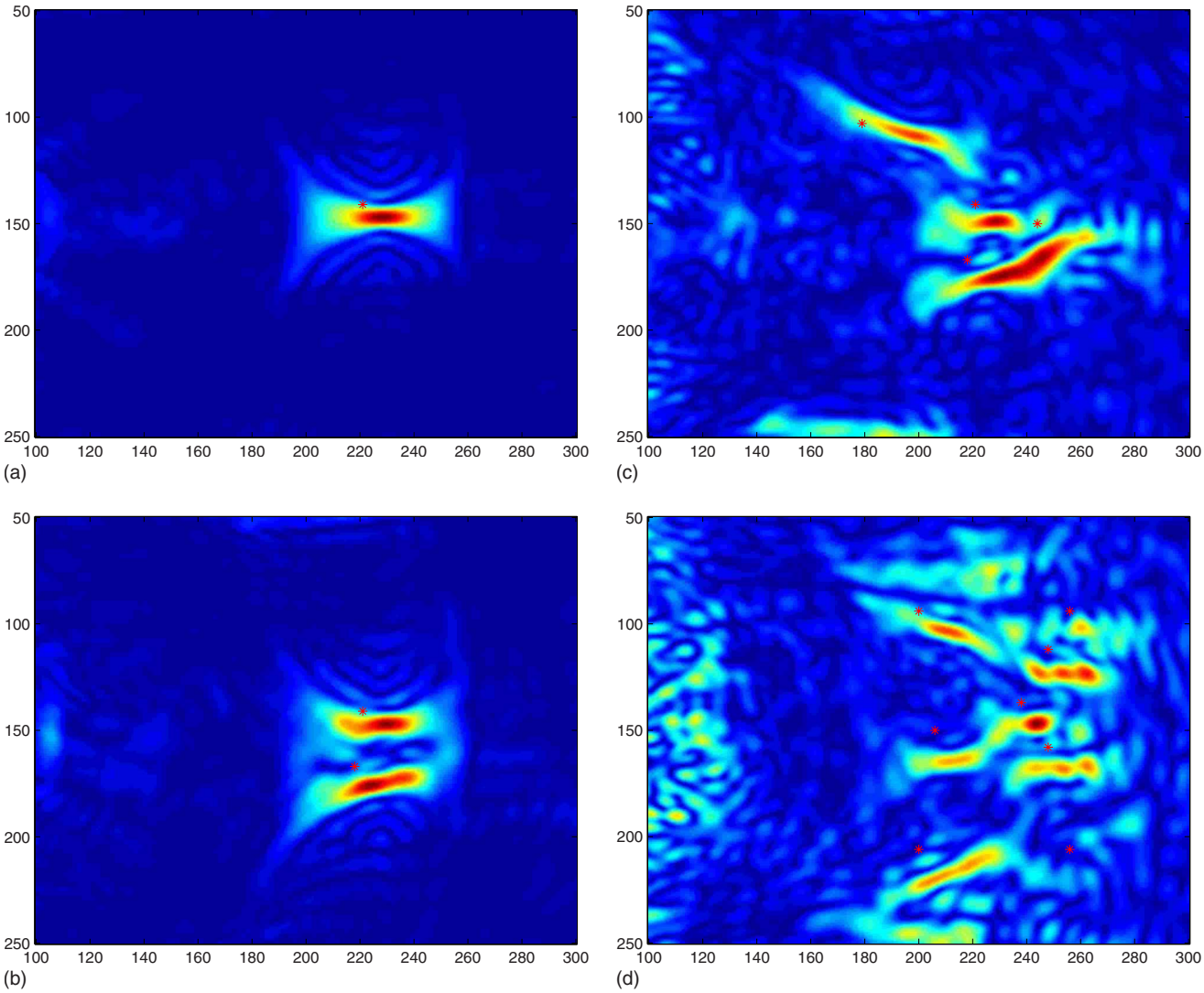


FIG. 6. (Color online) CMU experiment data imaging: (a) one target, (b) two targets, (c) four targets, and (d) eight targets.

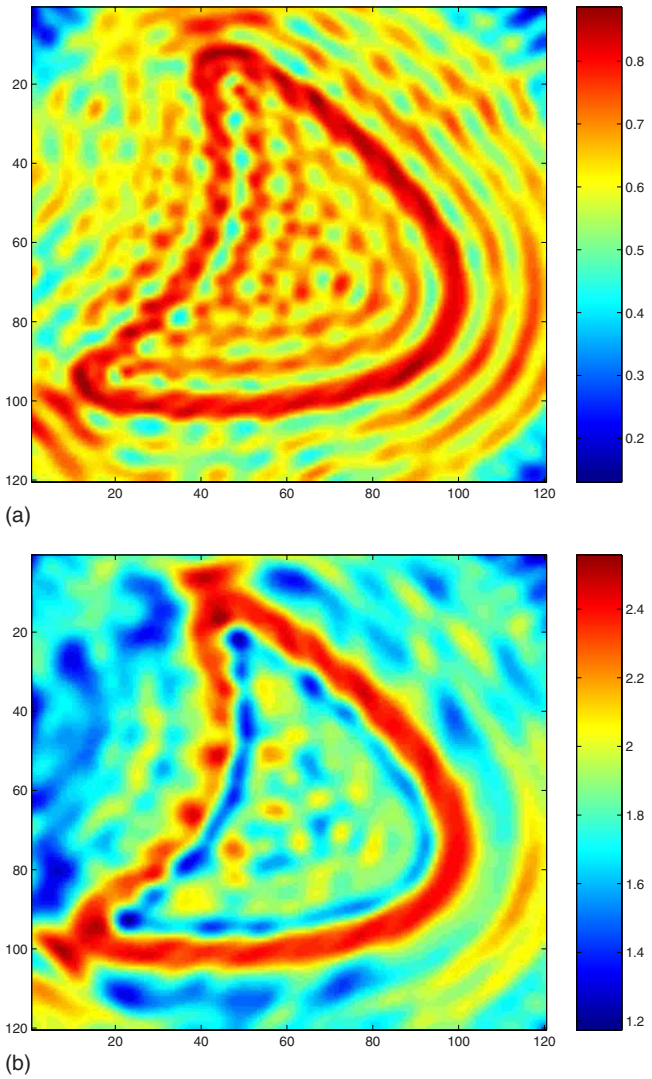


FIG. 7. (Color online) Multitone algorithm using full aperture data with one frequency (left) and three frequencies (right).

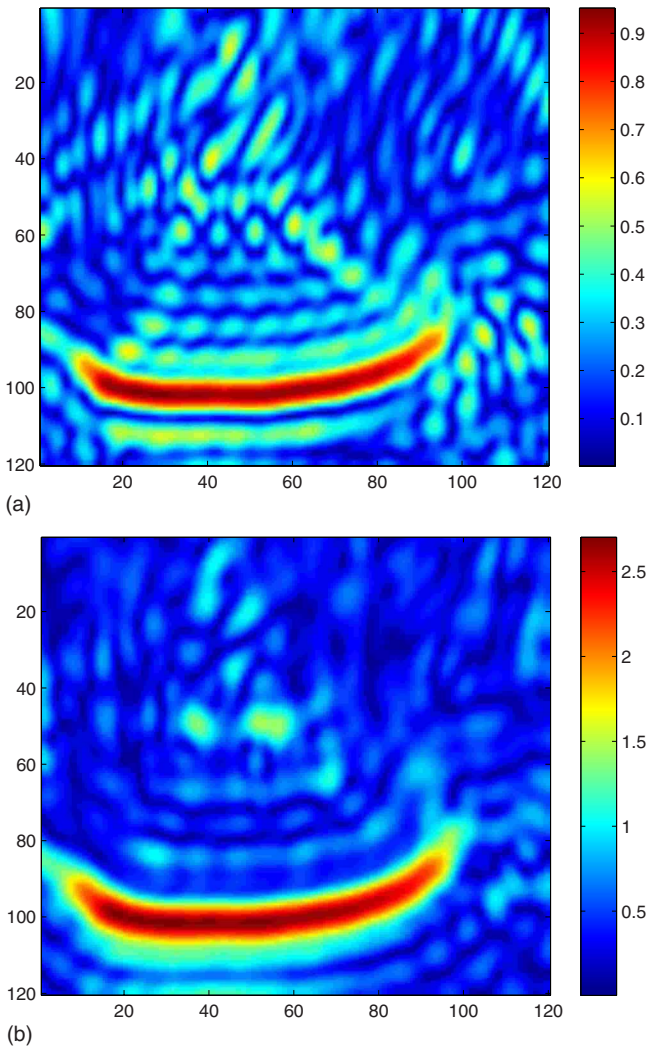


FIG. 8. (Color online) Multitone algorithm using limited aperture data (half of the circular array from the bottom) with one frequency (left) and three frequencies (right).

frequency diversity via superposition of coherent phases. By taking advantage of phase coherence of multiple frequencies, the imaging is enhanced and is robust with respect to noise. The algorithm can deal with limited or synthetic aperture data naturally as well as with different material properties and different types of illuminations and measurements.

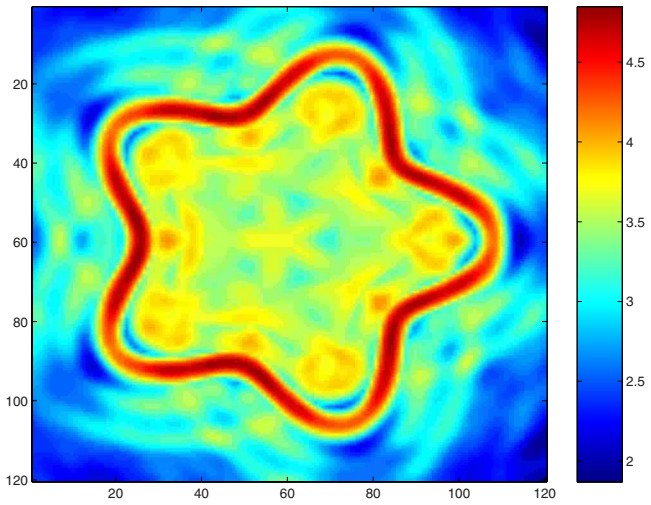
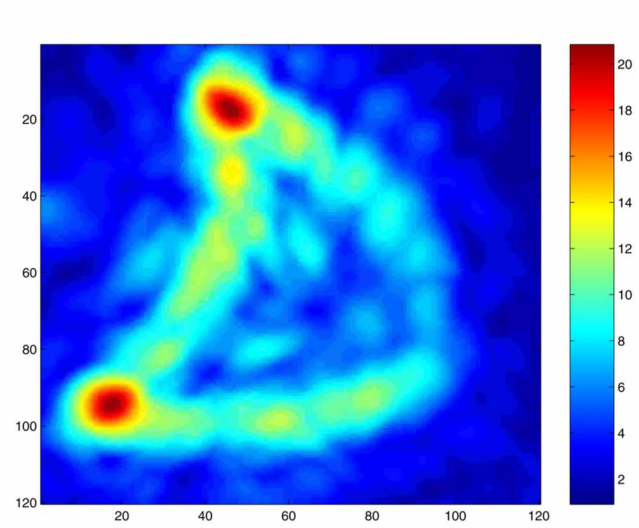
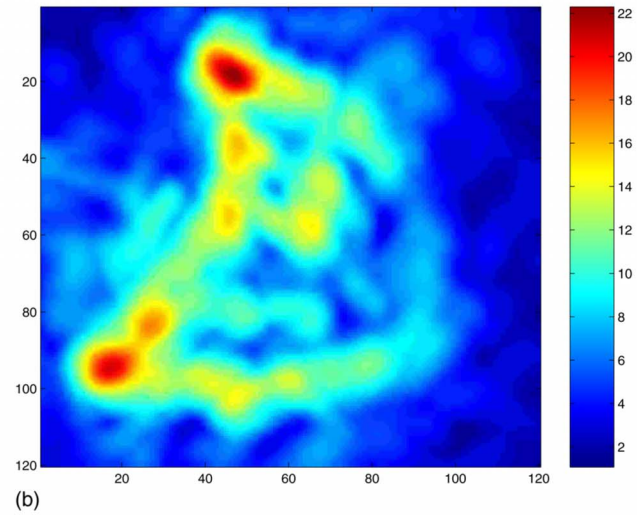


FIG. 9. (Color online) Imaging of extended target with Neumann boundary condition and full aperture using the multitone algorithm with six frequencies.

ACKNOWLEDGMENTS

We would like to thank our collaborators at Carnegie Mellon University, J. M. F. Moura, D. Stancil, and J. Zhu for providing us the experimental data. The research is partially supported by ONR Grant No. N00014-02-1-0090, DARPA Grant No. N00014-02-1-0603, NSF Grant No. 0307011, the Sloan Foundation, and Louisiana Board of Regents RCS Grant No. LEQSF(2008-11)-RD-A-18.





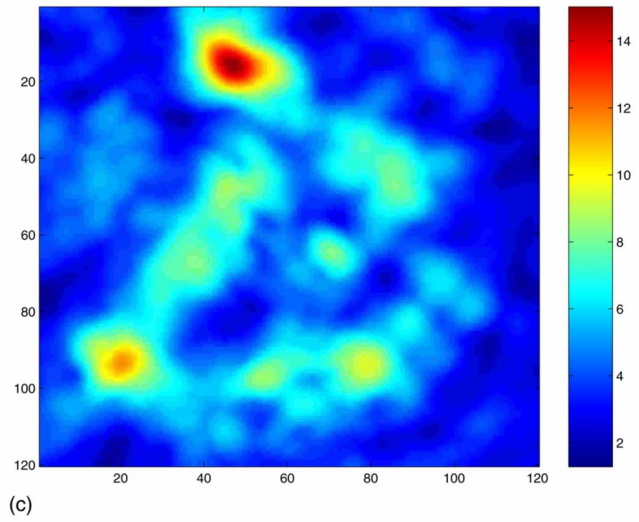


FIG. 10. (Color online) Synthetic aperture multitone imaging for a kite shape with clean data on the top, 10% multiplicative noise in the middle, and 10% random medium fluctuations on the bottom.

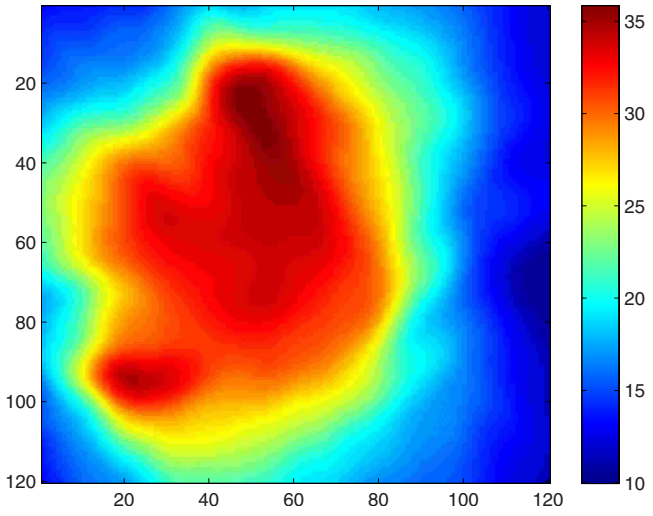
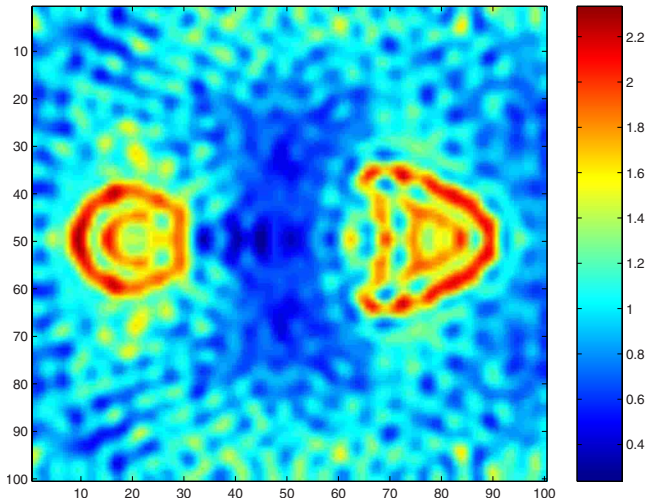


FIG. 11. (Color online) MUSIC imaging function for a kite shape with clean synthetic aperture data. The result is poor.



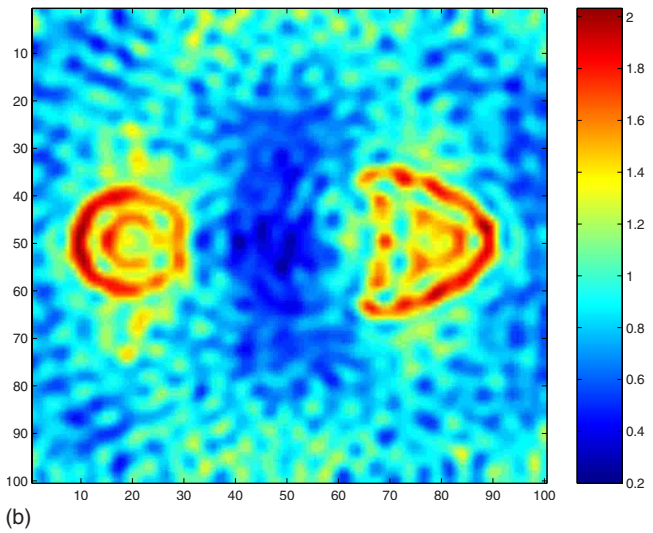


FIG. 12. (Color online) Multitone imaging for a kite shape and a circular shape using far field data with 100% multiplicative noise.

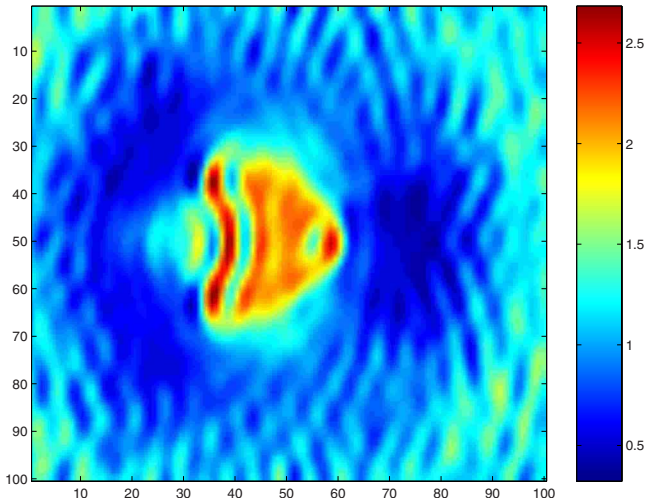


FIG. 13. (Color online) Multitone imaging for a kite shape with incident plane wave directions different from recorded far field data directions.

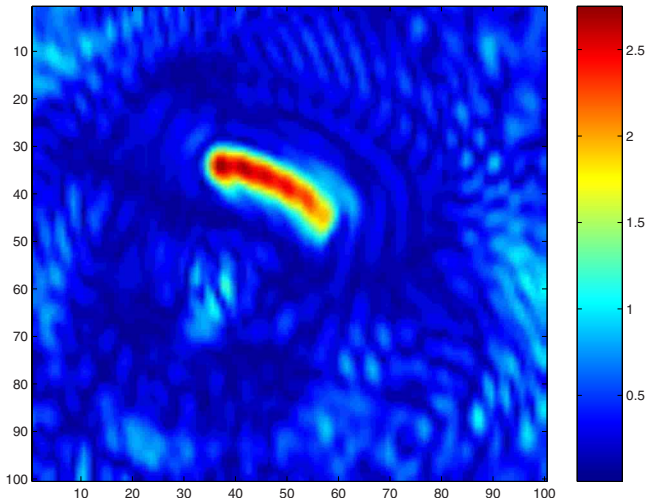


FIG. 14. (Color online) Multitone imaging with limited aperture far-field data.

[1]F. K. Gruber, E. A. Marengo, and A. J. Devaney, "Time-reversal imaging with multiple signal classification considering multiple scattering between the targets," J. Acoust. Soc. Am. **115**, 3042–3047 (2004).
[2]E. Kerbrat, C. Prada, 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).
[3]C. Prada and J.-L. Thomas, "Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix," J. Acoust. Soc. Am. **114**, 235–243 (2003).
[4]R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag. **34**, 276–280 (1986).
[5]E. Kerbrat, C. Prada, D. Cassereau, and M. Fink, "Ultrasonic nondestructive testing of scattering media using the decomposition of the time reversal operator," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1103–1113 (2002).
[6]J.-G. Minonzio, C. Prada, D. Chambers, D. Clorennec, and M. Fink, "Characterization of subwavelength elastic cylinders with the decomposition of the time-reversal operator: Theory and experiment," J. Acoust. Soc. Am. **117**, 789–798 (2005).
[7]G. Montaldo, M. Tanter, and M. Fink, "Real time inverse filter focusing by iterative time reversal," J. Acoust. Soc. Am. **115**, 768–775 (2004).
[8]G. 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).
[9]C. Prada, S. Manneville, D. Spoliansky, and M. Fink, "Decomposition of the time reversal operator: Detection and selective focusing on two scatterers," J. Acoust. Soc. Am. **99**, 2067–2076 (1996).
[10]C. Prada, J.-L. Thomas, and M. Fink, "The iterative time reversal process: Analysis of the convergence," J. Acoust. Soc. Am. **97**, 62–71 (1995).
[11]S. Hou, K. Solna, and H. Zhao, "A direct imaging algorithm for extended targets," Inverse Probl. **22**, 1151–1178 (2006).
[12]S. Hou, K. Solna, and H. Zhao, "A direct imaging method using far field data," Inverse Probl. **23**, 1533–1546 (2007).
[13]E. Marengo, F. Gruber, and F. Simonetti, "Time-reversal music imaging of extended targets," IEEE Trans. Image Process. **16**, 1967–1984 (2007).
[14]E. Marengo, R. Hernandez, and H. Lev-Ari, "Intensity-only signal-subspace-based imaging," J. Opt. Soc. Am. A **24**, 3619–3635 (2007).
[15]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968).
[16]D. Chambers and A. Gautesen, "Time reversal for a single spherical scatter," J. Acoust. Soc. Am. **109**, 2616–2614 (2001).
[17]D. H. Chambers, J. V. Candy, S. K. Lehman, J. S. Kallman, A. J. Poggio, and A. W. Meyer, "Time reversal and the spatio-temporal matched filter (L)," J. Acoust. Soc. Am. **116**, 1348–1350 (2004).
[18]M. Tanter, J.-F. Aubry, J. Gerber, J.-L. Thomas, and M. Fink, "Optimal focusing by spatio-temporal inverse filter. I. Basic principles," J. Acoust. Soc. Am. **110**, 37–47 (2001).
[19]M. Tanter, J.-F. Aubry, J. Gerber, J.-L. Thomas, and M. Fink, "Optimal focusing by spatio-temporal inverse filter. II. Experiments. Application to focusing through absorbing and reverberating media," J. Acoust. Soc. Am. **110**, 48–58 (2001).
[20]M. Tanter, J.-L. Thomas, and M. Fink, "Time reversal and the inverse filter," J. Acoust. Soc. Am. **108**, 223–234 (2000).
[21]L. Borcea, G. Papanicolaou, and C. Tsogka, "Coherent interferometric imaging," Geophysics **71**, S1165–S1175 (2006).
[22]D. H. Chambers and J. G. Berryman, "Analysis of the time-reversal operator for a small spherical scatterer in an electromagnetic field," IEEE Trans. Antennas Propag. **52**, 1729–1738 (2004).
[23]D. H. Chambers and J. G. Berryman, "Time-reversal analysis for scatterer characterization," Phys. Rev. Lett. **92**, 023902 (2004).
[24]D. H. Chambers, "Analysis of the time-reversal operator for scatterers of finite size," J. Acoust. Soc. Am. **111**, 411–419 (2002).
[25]S. Hou, K. Solna, and H. Zhao, "Imaging of location and geometry for extended targets using the response matrix," J. Comput. Phys. **199**, 317–338 (2004).
[26]H. Zhao, "Analysis of the response matrix for an extended target," SIAM J. Appl. Math. **64**, 725–745 (2004).
[27]J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200 (1994).
[28]D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, 2nd ed. (Springer, Berlin, 1998).

Sex differences in distortion-product and transient-evoked otoacoustic emissions compared

Dennis McFadden[a)]
Department of Psychology and Center for Perceptual Systems, University of Texas, 1 University Station A8000, Austin, Texas 78712-0187

Glen K. Martin
VA Loma Linda Healthcare System, Research Service (151), 11201 Benton Street, Loma Linda, California 92357 and Department of Otolaryngology Head and Neck Surgery, Loma Linda University Medical Center, Loma Linda, California 92357

Barden B. Stagner
VA Loma Linda Healthcare System, Research Service (151), 11201 Benton Street, Loma Linda, California 92357

Mindy M. Maloney
Department of Psychology and Center for Perceptual Systems, University of Texas, 1 University Station A8000, Austin, Texas 78712-0187

(Received 14 May 2008; revised 29 October 2008; accepted 6 November 2008)

Although several studies have documented the existence of sex differences in spontaneous otoacoustic emissions (SOAEs) and transient-evoked OAEs (TEOAEs) in humans, less has been published about sex differences in distortion-product OAEs (DPOAEs). Estimates of sex and ear differences were extracted from a data set of OAE measurements previously collected for other purposes. In accord with past findings, the sex differences for TEOAEs were substantial for both narrowband and wideband measures. By contrast, the sex differences for DPOAEs were about half the size of those for TEOAEs. In this sample, the ear differences were small for TEOAEs in both sexes and absent for DPOAEs. One implication is that the cochlear mechanisms underlying DPOAEs appear to be less susceptible to whatever influences are responsible for producing sex differences in TEOAEs and SOAEs in humans. We discuss the possibility that differences in the effective level of the stimuli may contribute to these outcomes.
 [DOI: 10.1121/1.3037231]

PACS number(s): 43.64.Jb, 43.64.Bt, 43.66.Ki [WPS] Pages: 239–246

I. INTRODUCTION

In humans, otoacoustic emissions (OAEs) exhibit marked sex differences. Spontaneous OAEs (SOAEs) are more numerous and stronger in females than in males, and click-evoked OAEs (CEOAEs) are stronger in females than in males (e.g., Bilger *et al.*, 1990; Talmadge *et al.*, 1993; McFadden, 1993b, 1998; McFadden and Loehlin, 1995; McFadden *et al.*, 1996; McFadden and Pasanen, 1998, 1999; McFadden and Shubel, 2003). There appears to be an ear difference overlaid on the sex difference. The number of SOAEs and the strength of CEOAEs are greater in right ears than in left ears (e.g., Bilger *et al.*, 1990; Talmadge *et al.*, 1993; McFadden *et al.*, 1996; McFadden and Pasanen, 1998, 1999; Khalfa *et al.*, 2001), although the magnitudes of these ear differences are smaller than those of the sex differences. Because these sex and ear differences exist in newborns as well as in adults (Strickland *et al.*, 1985; Burns *et al.*, 1992, 1994; Morlet *et al.*, 1995, 1996; Thornton *et al.*, 2003; Berninger, 2007), and because OAEs appear to be reasonably stable through life as long as there is no damage to the cochlea; e.g., Harris *et al.*, 1991; Franklin *et al.*, 1992; Burns *et al.*, 1993, 1994; Engdahl *et al.*, 1994; Marshall and Heller, 1996; McFadden *et al.*, 1996), the implication is that the sex and ear differences are produced by mechanisms operating during prenatal development. Thus, the degree of exposure to androgens is a likely candidate mechanism for these differences (especially the sex differences, see McFadden, 2002, 2008) simply because degree of prenatal exposure to androgens is known to be responsible for so many other sex differences in body, brain, and behavior (e.g., Nelson, 2005).

The evidence available suggests that the sex and ear differences in distortion-product OAEs (DPOAEs) may be smaller than those for SOAEs and CEOAEs (Bonfils *et al.*, 1988; Gaskill and Brown, 1990; Lonsbury-Martin *et al.*, 1991; Moulin *et al.*, 1993; Cacace *et al.*, 1996; Dhar *et al.*, 1998; Bowman *et al.*, 2000; O'Rourke *et al.*, 2002; Dunckley and Dreisbach, 2004; Keefe *et al.*, 2008). If confirmed, these dissimilarities would have implications for theories about the mechanisms underlying OAEs. Shera and Guinan (1999, 2003) have argued that one of the cochlear mechanisms underlying DPOAEs is fundamentally different from that underlying SOAEs and CEOAEs. Namely, SOAEs and

[a)]Author to whom correspondence should be addressed. Electronic mail: mcfadden@psy.utexas.edu

CEOAEs are thought to be primarily the result of a linear, reflection-based mechanism, whereas DPOAEs are thought to be the result of both the linear reflection mechanism operating from a location on the low-frequency side of the primary tones, plus a nonlinear cochlear mechanism operating near the location of the higher primary tone. An absence of, or a diminution in, sex differences in DPOAEs relative to SOAEs and CEOAEs would suggest that the linear, reflection-based mechanism is subject to modulation by some agent(s) such as androgens, but that the nonlinear mechanism is not (or is less) subject to such modulation.

In an attempt to increase the available information about the relative sizes of the sex and ear differences in DPOAEs and other OAEs, we re-examined a set of data acquired by Martin and his colleagues for other purposes. A small subset of those data had been reported previously [Lonsbury-Martin *et al.*, 1991, Fig. 5(b)], but that report did not include all of the subjects eventually tested. Because this data set contained both DPOAEs and TEOAEs obtained on the same subjects, we were able to compare the individual differences in those measures as well as to estimate sex and ear differences.

II. METHODS

Each subject provided informed consent under a human-research protocol approved by the University of Miami School of Medicine's Institutional Review Board. Each subject also received monetary compensation for participation in the study.

A. Subjects

Because one of the interests of the original study was the effects of aging on OAEs, the initial investigators (Lonsbury-Martin *et al.*, 1991) expended considerable effort to obtain subjects with normal hearing (<20 dB hearing level or HL) over the standard set of audiometric test frequencies. However, by necessity, some of the 50 year olds eventually included in their data set did have hearing worse than this target (<30 dB HL), but only at one frequency in one or both ears. Both SOAEs and DPOAEs have been shown to decline in strength with increasing age in a similarly select sample (Lonsbury-Martin *et al.*, 1991; Whitehead *et al.*, 1995), suggesting that even though hearing sensitivity remains nominally normal, there apparently can be some age-related degradation of the cochlear mechanisms responsible for OAEs. (Note that age-related declines in both SOAEs and DPOAEs suggest that both the linear and nonlinear cochlear mechanisms are being affected—see Shera and Guinan, 1999, 2003.) Because here we wished to know about the relative sizes of the sex differences in OAEs prior to age-related declines, we restricted our subject pool to people aged 15–35. A total of 51 females and 57 males satisfied the age criteria. All subjects had normal tympanograms, normal acoustic-reflex thresholds at 1.0 kHz both contralaterally and ipsilaterally, hearing sensitivity of 20 dB HL or better in both ears at the octave frequencies between 0.25 and 8.0 kHz, and no history of hearing disorders.

Whitehead *et al.* (1993) demonstrated that SOAEs are more numerous and stronger in Blacks and Asians than in Whites, but little is known about ethnic or racial effects on the sex difference in DPOAEs. Unfortunately, the question could not be examined here because the number (N) of Blacks and Asians was so small. To be cautious, we excluded the 10 females and the 11 males who identified themselves as Black or Asian, leaving in the data pool 41 females and 46 males who identified themselves either as White or Hispanic (Whitehead *et al.* (1993) excluded Hispanics as well from their study). Also, subjects having OAEs 3.0 or more standard deviations from the mean of their group were excluded from the data analyses for that specific condition, and when ear differences were being assessed, only subjects having acceptable data for both ears were included. Thus, the Ns for individual conditions and comparisons varied but are shown in the figures and tables.

B. General

The general data-collection procedures were described in Lonsbury-Martin *et al.*, 1991 and Whitehead *et al.*, 1995. Two characteristics of the stimuli used to produce DPOAEs were atypical of current practice. The two primary tones were equal in level, and that level (75 dB sound-pressure level or SPL) was higher than typically used in recent years. The possible effects of stimulus level are discussed in Sec. IV. The primary tones used to collect DPOAEs ranged between 0.8 and 8.0 kHz in steps of approximately 0.1 octave. Each frequency step was characterized using the geometric mean of the primary tones. The strength of the DPOAE was measured during presentations of the primary tones that lasted approximately 90 ms. Also, a wideband TEOAE was obtained using the nonlinear procedure and the default settings provided by the Otodynamics ILO88 device. The default click value produced click stimuli averaging about 81 dB peak-equivalent SPL (peSPL) in the ear canals. That wideband response then was filtered to obtain narrowband TEOAE responses of about 0.1 octave in width.[1]

To simplify the analyses, we targeted four frequency regions for analysis: 1.5, 2.0, 3.0, and 4.0 kHz. In order to obtain measures as stable as possible, a mean level of the distortion product was calculated across three conditions for each target frequency for each subject. Those three conditions were the ones having primary tones whose geometric means were closest to each of the four target frequencies (e.g., 1.409, 1.516, and 1.623 kHz for the nominal value of 1.5 kHz; and 3.729, 4.003, and 4.287 kHz for the nominal value of 4.0 kHz). A parallel procedure was used for the TEOAE data. Specifically, means were obtained for each of the four target frequency regions by averaging the measures extracted with the filter set to the three frequencies closest to the four target values. Prior to the calculation of these three-frequency means, all obtained measures that exceeded the mean for an individual frequency condition by 3.0 standard deviations or more were deleted.

Note that the TEOAE frequency regions chosen for examination corresponded to the frequencies of the primary tones used to collect the DPOAEs, not to the frequencies of

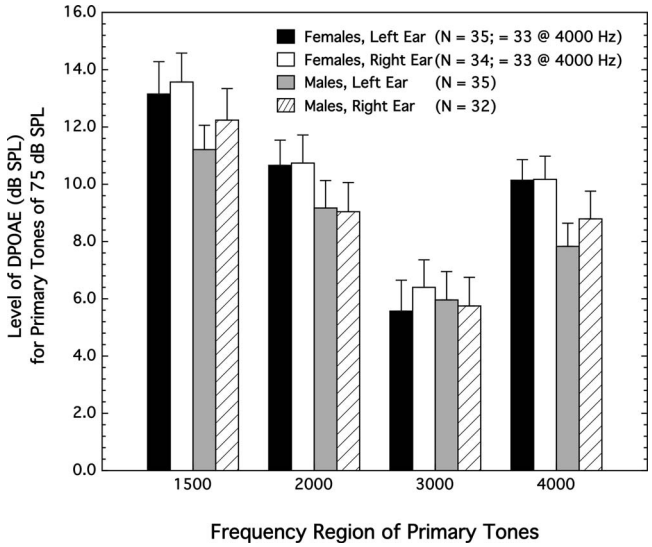


FIG. 1. Level of the DPOAE produced by equal-level primary tones of 75 dB SPL located in the frequency regions indicated. For each designated frequency region, the results for three pairs of primary tones were averaged; see text. Only data from Caucasian and Hispanic subjects included. Flags designate the standard errors of the mean.

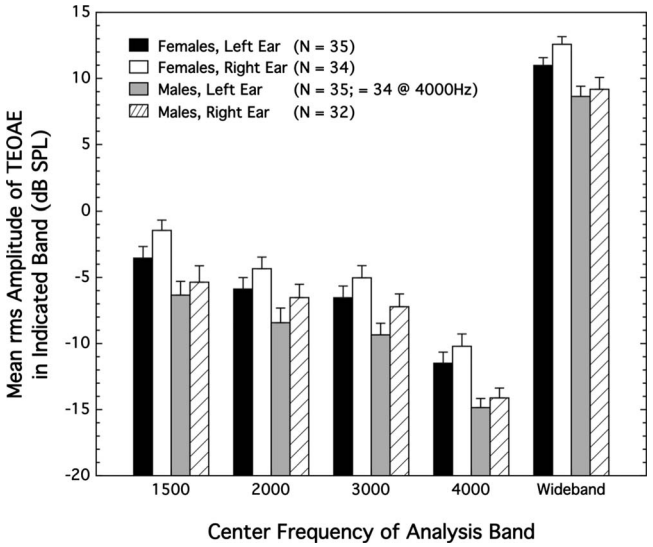


FIG. 2. Strength of the TEOAE responses evoked by a click of about 81 dB SPL. The averaged TEOAE waveform was analyzed in several narrowbands as well as wideband. For the narrowband conditions, the levels shown are for 0.1-octave bands centered approximately on the frequency values shown.

the $2f_1$-f_2 distortion products themselves. The reason for this choice was that the DPOAE data were collected with relatively intense primary levels (75 dB each), meaning that the predominant source of the $2f_1$-f_2 component was the one residing near the f_2 location along the cochlear partition (Shera and Guinan, 1999, 2003; Kalluri and Shera, 2001).

In the majority of cases, data were available for both ears of a subject, but in some cases the data for one ear were excluded. Because only subjects having data for both ears were included in the ear-difference analyses, the Ns available for calculating ear differences (a within-subject difference) were smaller than those available for calculating sex differences (a group difference).

The sex and ear differences will be expressed in terms of effect size, which was calculated here as the difference between the means of the two groups divided by the square root of the weighted mean of their variances. Thus, effect size is a measure like d' (Green and Swets, 1966), which is commonly used in psychoacoustical studies. Cohen (1992) suggested that effect sizes of 0.2, 0.5, and 0.8 could be considered small, medium, and large, respectively. A measure like effect size is preferable to the outcomes of statistical tests when a large number of comparisons are to be made. Here, a two-tailed t-test was calculated whenever an effect size greater than 0.4 was obtained.

III. RESULTS

Both the sex differences and ear differences were smaller for DPOAEs than for TEOAEs. The basic data are shown in Figs. 1 and 2, and Table I shows the effect sizes for the various sex and ear differences. The figures and the effect sizes shown for the sex differences were based on the data for all possible subjects; the effect sizes shown for the ear differences were calculated using only those subjects having data for both ears.

As can be seen by comparing the top and bottom halves of Table I, the effect sizes for sex difference in DPOAEs averaged about 0.25 while those for TEOAEs averaged about 0.56. For DPOAEs, only one of the ten sex differences examined achieved statistical significance, while for TEOAEs seven of ten did. The most significant sex-difference comparison was for the TEOAE in the right ear at 4000 Hz ($p=0.0017$) and when that value was corrected for the 11 t-tests calculated (Darlington, 1990, pp. 249–251), $p=0.019$. The bottommost entries in Table I are for the wideband TEOAE data.

Few other published studies provide the information necessary to calculate effect sizes for comparison, but McFadden and Pasanen (1999) and McFadden and Shubel (2003) reported effect sizes of 0.76 and 0.84, respectively, for the sex difference in a wideband TEOAE (not using a nonlinear procedure like that used here; see footnote [1]). Also, Thornton *et al.* (2003) obtained effect sizes of only about 0.20 for the sex difference in the overall TEOAEs obtained from neonates using the ILO88 system, but the sex differences were substantially larger at high frequencies than at low (their Fig. 5), meaning that calculating the overall TEOAE underestimates the sex differences at higher frequencies. Unfortunately for current purposes, neither McFadden and Pasanen (1999) nor Thornton *et al.* (2003) obtained measures of DPOAEs from their subjects for comparison with the TEOAEs.

Dunckley and Dreisbach (2004) measured DPOAEs over a wide frequency range, and they did provide the information necessary to calculate effect sizes. Unfortunately for comparisons here, however, Dunckley and Dreisbach (2004) used primary tones of unequal level, and they used such strict exclusion criteria that about one-third of females and two-thirds of males presenting with normal hearing were excluded from their study. Such a strong selection bias was bound to have considerable effect on the outcomes. That said, the effect sizes over the frequency range 1.0–5 kHz

TABLE I. Effect sizes for the sex and ear differences shown in Figs. 1 and 2. Positive effect sizes correspond to the mean OAE being stronger in females than males, or stronger in right ears than left ears; Ns for sex differences are shown in Figs. 1 and 2; Ns for ear differences=28/22 for female/male, except=28/21 at 4000 Hz.

	Effect sizes for			
	Sex difference		Ear difference	
Condition	Left ear	Right ear	Female	Male
	Distortion-product OAEs			
1500 Hz	0.326	0.218	0.173	0.190
2000 Hz	0.274	0.296	0.000	0.007
3000 Hz	−0.064	0.117	0.060	−0.041
4000 Hz	0.514[a]	0.271	−0.009	0.204
Mean=	0.263	0.226	0.056	0.090
	Transient-evoked OAEs (nonlinear procedure)			
1500 Hz	0.492[a]	0.668[b]	0.354	0.121
2000 Hz	0.432	0.405	0.358	0.186
3000 Hz	0.535[a]	0.402	0.175	0.277
4000 Hz	0.741[b]	0.804[b]	−0.013	−0.115
Mean=	0.550	0.570	0.218	0.117
Wideband	0.621[a]	0.756[a]	0.413	0.087

[a]Two-tailed t-test: $p<0.05$.
[b]Two-tailed t-test: $p<0.01$.

averaged about 0.11, favoring the females; the effect sizes over the range 6.0–10 kHz averaged about −0.37, favoring the males; and the effect sizes over the range 11.0–15.0 kHz averaged about 0.30, again mostly favoring the females. Because the exclusion criteria focused on the highest-frequency regions, it is those data that were most likely to be affected by the selection bias. All factors considered, the Dunckley and Dreisbach (2004) data appear to confirm the conclusion that sex differences for DPOAEs are small in humans.

The entries at the far right in Table I reveal that the ear differences for TEOAEs were quite small (effect sizes averaging about 0.17) and that the ear differences for DPOAEs were essentially zero (effect sizes averaging about 0.07). None of these comparisons achieved statistical significance, which is generally in accord with what is known about ear differences in OAEs, although very little has been published previously about ear differences in DPOAEs. In the past, ear differences in TEOAEs and SOAEs have yielded effect sizes of about 0.26 and 0.17, respectively (McFadden and Pasanen, 1999). Thornton *et al.* (2003) also reported small effect sizes for the ear difference, especially when the left ear was tested first. Keefe *et al.* (2008) reported significantly stronger TEOAEs in the right ear than the left for newborns, and no ear difference for DPOAEs, but no analyses of sex differences were reported. Note that, for TEOAEs in Table I, the effect sizes for the three lowest target frequencies and the wideband condition were generally larger than those for the 4.0 kHz region. We have no explanation for this pattern.

Partial dissociations between DPOAEs and other types of OAE have been observed following administrations of ototoxic drugs (e.g., Wier *et al.*, 1988; McFadden and Pasanen, 1994; Whitehead *et al.*, 1996), and these dissociations are commonly interpreted as evidence in support of DPOAEs originating from a different mechanism in the cochlea from the one(s) responsible for TEOAEs and SOAEs. [The presence of a seasonal effect for the CEOAEs of rhesus monkeys in the absence of such an effect for DPOAEs (McFadden *et al.*, 2006a) also can be interpreted as a dissociation between types of OAE.] If DPOAEs and TEOAEs do originate from different underlying mechanisms, then one would expect them to correlate less highly than do TEOAEs and SOAEs. To test this inference, correlations were calculated between DPOAEs and TEOAEs. The results are shown in Table II, where the correlations are shown within frequency region, within sex, and within ear. As can be seen, these correlations averaged around 0.5, which is considerably smaller than the correlation of about 0.76 previously reported between TEOAEs and SOAEs (McFadden and Pasanen, 1999). For the Ns involved here, correlations of about 0.32 and 0.42 would be required to be judged significantly different from 0.0 with probability values of 0.05 and 0.01, respectively (two-tailed). Smurzynski and Kim (1992), Moulin *et al.* (1993), and Gorga *et al.* (1993) also reported correlations

TABLE II. Correlations between TEOAEs and DPOAEs, shown separately for the various frequency regions, the two ears, and the two sexes. Ns for females=35/34 for Lt vs Lt and Rt vs Rt, except=33/33 at 4000 Hz; Ns for males=33/32 for Lt vs Lt and Rt vs Rt, except=32/32 at 4000 Hz.

	Females		Males	
Condition	Left vs left	Right vs right	Left vs left	Right vs right
1500 Hz	0.452	0.356	0.483	0.441
2000 Hz	0.527	0.442	0.451	0.465
3000 Hz	0.713	0.552	0.357	0.518
4000 Hz	0.246	0.413	0.404	0.634
Mean=	0.484	0.441	0.424	0.514

TABLE III. Correlations between the two ears for DPOAEs and for TEOAEs measured at various frequencies. Ns for DPOAEs=28/23 for females/males, except=26/23 at 4000 Hz; Ns for TEOAEs=28/22 for females/males, except=28/21 at 4000 Hz.

	Females Right ear		Males Right ear	
Left ear Condition	DPOAEs	TEOAEs	DPOAEs	TEOAEs
		DPOAEs		
1500 Hz	0.812	···	0.771	···
2000 Hz	0.764	···	0.708	···
3000 Hz	0.679	···	0.785	···
4000 Hz	0.680	···	0.776	···
Mean=	0.734	···	0.760	···
		TEOAEs		
1500 Hz	···	0.641	···	0.751
2000 Hz	···	0.308	···	0.683
3000 Hz	···	0.447	···	0.647
4000 Hz	···	0.711	···	0.646
Wideband	···	0.624	···	0.877
Mean=	···	0.546	···	0.721

of about 0.5–0.7 between TEOAEs and DPOAEs. In spotted hyenas, the correlations between DPOAE strength in the 3.5–kHz region and wideband TEOAE strength were about 0.16 for females, about 0.01 for males, and about 0.10 when the data were pooled across the 13 females and 14 males in the various treatment groups (McFadden *et al.*, 2006b). For rhesus monkeys, the previously unpublished correlations between DPOAEs and CEOAEs were about 0.41 and 0.28 for all females combined and all males combined, respectively; these values are for the fall, when the male OAEs were weakest (see McFadden *et al.*, 2006a).

For completeness, we show in Table III the correlations between left and right ears separately for DPOAEs and TEOAEs in these same subjects. The correlations between ears were generally larger than those between DPOAEs and TEOAEs in the same ear. We note that the within-subject correlations were generally similar in magnitude for DPOAEs and TEOAEs. Were this outcome to be confirmed, it would imply that the linear and nonlinear cochlear mechanisms are about equally similar interaurally. For comparison, Thornton *et al.* (2003) reported a correlation of about 0.56 between the TEOAEs measured in the two ears, also obtained using the ILO88 system used here, and Berninger (2007) obtained interaural correlations in the TEOAEs of infants ranging from about 0.3 to 0.7 across his half-octave analysis bands.

Note that the correlations shown in Tables II and III may underestimate the values that would have been obtained had the subjects originally not been so highly selected for hearing sensitivity. That selection surely introduced a restriction of the range of OAE strength.

In passing we note that, for this highly selected group of relatively young subjects, hearing sensitivity was greater in females than in males (compare McFadden, 1993a, 1998; McFadden and Mishra, 1993). Averaged across the four test frequencies of interest here, the effect sizes for the sex difference were 0.45 and 0.59 for the left and right ears, respectively. By comparison, the averaged effect sizes for the ear difference in hearing sensitivity were only 0.05 and 0.14 for females and males, respectively.

IV. DISCUSSION

Although some other investigators have analyzed for sex and/or ear differences in DPOAEs (Bonfils *et al.*, 1988; Gaskill and Brown, 1990; Lonsbury-Martin *et al.*, 1991; Moulin *et al.*, 1993; Cacace *et al.*, 1996; Dhar *et al.*, 1998; Bowman *et al.*, 2000; O'Rourke *et al.*, 2002; Dunckley and Dreisbach, 2004; Keefe *et al.*, 2008), this study is among the first to compare the sizes of the sex and ear differences in DPOAEs with those for TEOAEs in the same subjects. In accordance with previous reports, both the sex and ear differences for DPOAEs were smaller than those for TEOAEs.

When interpreting our various outcomes, the reader needs to remember that the DPOAE data all were collected with both primary tones at 75 dB SPL, and that may not be the optimal level for seeing sex differences, ear differences, etc. Also, the duration of the primary tones here was approximately 90 ms, not the longer presentations used by some investigators.

Two anonymous reviewers of this paper noted that the observed difference in the size of the sex difference for CEOAEs and DPOAEs may be partly or wholly attributable to differences in the effective level of the stimuli used to produce these two types of OAE. Specifically, CEOAEs typically are elicited by click stimuli that range in level from about 80 dB peSPL and below. Because the typical CEOAE click is about 100 μs in duration or shorter, the bandwidth of the click extends to 10.0 kHz or higher, meaning that even with relatively strong clicks, the spectrum level of the typical CEOAE click is only about 40 dB SPL or lower. To the extent that the overall strength of the echo depends on the local magnitude of displacement at various locations along the length of the basilar membrane, then the CEOAEs obtained with clicks of this sort are based on effective stimulation that is weak locally. In contrast, the DPOAEs reported here were measured with primary tones that were strong by current standards: 75 dB SPL each. Thus, if the sex difference in OAE strength declined with increasing stimulus level, then the smaller sex difference for DPOAEs than CEOAEs in humans might be just a natural consequence of the relative difference in effective strength of the stimuli typically used to elicit the two types of OAE. That is, the cochlear amplifier might be contributing comparatively more to the typical CEOAE response than to the typical DPOAE response.

To address this suggestion, we examined some DPOAE data from an ongoing study and some CEOAE data from a previous study. The DPOAE data were collected with a wide range of levels for the primary tones and over three frequency ranges. We obtained estimates of the strength of the DPOAE for each subject in each frequency region for primary tones of both 71 and 50 dB SPL, and then we calculated effect sizes for the sex difference between the 36 males and 13 females in that sample. The effect sizes, averaged

across frequency regions, were 0.50 and 0.58 for the 71 and 50 dB primaries, respectively. That is, the magnitude of the sex difference was larger for the weaker primary tones, although the increment was not large. In a previous study (McFadden and Pasanen, 1998), we collected CEOAE data from 57 males and 57 females using four click levels. For an 18 dB decrease in click level, the effect size for sex difference increased from 0.70 to 0.73. So, for both CEOAEs and DPOAEs, we have evidence that the locally effective level of the stimuli can affect the magnitude of the sex difference, and in the same direction. Thus, the smaller sex differences seen for DPOAEs than CEOAEs might be attributable in part to the stimuli routinely used to elicit them.

While this insight is unquestionably important when it comes to evaluating explanations for the difference between DPOAEs and CEOAEs, it is less important practically. Regardless of how much the difference in stimulus level may contribute to the difference between the sexes in DPOAEs and CEOAEs, DPOAEs will continue to be a poor measure for investigators interested in sex differences in OAEs and hormone effects on OAEs simply because measurements of DPOAEs with truly weak primary tones typically are impractical, and the sex difference is invariably small when moderate and strong primaries are used.

One way of thinking about the smaller sex difference for DPOAEs than for CEOAEs is that the cochlear mechanisms underlying the production of CEOAEs are sensitive to some agent or agents operating differently in the two sexes at some point early in development, and that the mechanisms underlying DPOAEs are less sensitive to those agents. Because DPOAEs are thought to be the result of both the linear, reflection-based mechanism that underlies CEOAE production plus a nonlinear distortion mechanism operating near the location of the f_2 primary tone (Shera and Guinan, 1999, 2003), a second implication of the present findings is that the greater part of the DPOAE response in the ear canal originates from the nonlinear distortion mechanism, at least for the primary levels used here (75 dB SPL each). Otherwise, the sex difference in DPOAEs should have been more similar to that seen for CEOAEs. One possible agent for the production of the sex difference in the reflection-based mechanism is exposure to androgens prenatally (McFadden, 2002, 2008).

When evaluating explanations for the sex and ear differences in OAEs, it is necessary to consider whether differences in the acoustics of the outer or middle ears are playing a major role. Perhaps the most compelling counterargument is that many investigators adjust the levels of their stimuli in the ear canal prior to collecting both TEOAEs and DPOAEs. This procedure ought to greatly reduce, if not eliminate, any group or individual differences in such dimensions as ear-canal volume. Furthermore, Johansson and Arlinger (2003) observed the human-typical sex difference in TEOAEs even though there were no sex differences in middle-ear compliance or middle-ear pressure in those same subjects. Also, Keefe *et al.* (2008) reported that ear differences in vLo (the equivalent middle-ear volume averaged from 0.25 to 1.0 kHz) and rHi (energy reflectance averaged from 2.0 to 8.0 kHz) do exist in infants, but those ear differences did not explain the ear differences observed in the DPOAEs or (nonlinear) TEOAEs of those infants. Margolis *et al.* (1999) reported small, but significant, sex differences in measures of wideband reflectance; however, those differences were primarily at frequencies below 1.0 kHz [their Fig. 5(c)], and the sex differences in OAEs extend to frequencies well above that. Elsewhere, one of us has argued that the sex differences in OAEs may be attributable, at least in part, to differential exposure to androgens during prenatal development (McFadden, 2002, 2008), and that the ear differences may be attributable to asymmetries in the strength of the efferent system (McFadden, 1993a, 1998).

Sex and ear differences have been studied only infrequently in nonhumans, but the existing evidence does suggest that stronger OAEs in females than males may be the basic mammalian pattern. Rhesus monkeys (McFadden *et al.*, 2006a) and sheep (McFadden *et al.*, 2008a, 2008b) show stronger TEOAEs in females than in males, and, just as for humans, the sex difference in rhesus monkeys and sheep was smaller for DPOAEs than for TEOAEs. Lonsbury-Martin and Martin (1988) observed humanlike sex and ear differences in the SOAEs of a colony of pigtail macaques, although the number of males tested and the number of SOAEs recorded both were small. The DPOAE data of Torre and Fowler (2000) also contained a small humanlike sex difference in the youngest group of their rhesus monkeys (all of whom were middle-aged). Valero *et al.* (2008) found substantial sex differences in the DPOAEs of a primitive New World monkey, the marmoset, but little is yet known about TEOAEs in that species. Guimaraes *et al.* (2004) reported stronger DPOAEs in female CBA mice than in males. However, this difference was evident only in middle-aged mice, not in young mice, suggesting that this sex difference actually might be attributable to differential age-related hearing loss rather than a developmental difference in the two sexes. McFadden *et al.* (1999) found no sex differences in the DPOAEs of chinchillas even though an evoked-potential measure of hearing sensitivity suggested that females were more sensitive than males at high frequencies. It is interesting that two amphibian species, leopard frogs and bullfrogs, showed substantially stronger DPOAEs in females than in males (Vassilakis *et al.*, 2004). Establishing whether there is a common mammalian plan for sex and ear differences in OAEs is difficult because TEOAEs and SOAEs generally cannot be found in the small species commonly used for auditory research, and as we have seen, the sex differences exhibited by DPOAEs often are not representative of the sex differences exhibited by TEOAEs or SOAEs in the same ears.

Because the sex differences in TEOAEs and SOAEs exist in newborns as well as in adults (Strickland *et al.*, 1985; Burns *et al.*, 1992, 1994; Morlet *et al.*, 1995, 1996; Thornton *et al.*, 2003; Berninger, 2007), a likely contributing factor to those sex differences is the degree of exposure to androgens during prenatal development (e.g., McFadden, 2002, 2008). The existence of smaller sex differences for DPOAEs than for TEOAEs suggests that the nonlinear cochlear mechanism believed to be primarily responsible for DPOAEs is less susceptible to the effects of prenatal androgen exposure than is

the linear, reflection-based mechanism believed to be primarily responsible for TEOAEs and SOAEs. This knowledge has implications for the ultimate understanding of the molecular mechanisms underlying the sex differences in OAEs.

A. Comparisons across species

Over the years, DPOAEs and CEOAEs have been measured in several mammalian species as well as in humans. One interesting outcome in the present context is that human DPOAEs are considerably weaker than those in all other species tested. For CBA mice, DPOAEs are quite strong, being about 15–40 dB weaker than primary tones that ranged from 55 to 75 dB SPL over the range of about 10–40 kHz (Jimenez *et al.*, 1999). In rabbits, DPOAEs are about 25–30 dB weaker than primary tones ranging from 45 to 75 dB SPL in the vicinity of 7.0–10.0 kHz (e.g., Porter *et al.*, 2006). For rhesus monkeys, spotted hyenas, sheep, and lemurs, the DPOAEs were about 45–55 dB weaker than the primary tones used to produce them (McFadden *et al.*, 2006a, 2006b, 2008b). For marmosets, they were about 50–57 dB weaker (Valero *et al.*, 2008). For the humans in the present study, the DPOAEs were about 65–70 dB weaker than the primaries, in accord with the human studies summarized by Probst *et al.* (1991). Why human cochlear mechanics produce such weak DPOAEs is not known.

Although the absolute strengths of the DPOAEs vary across species, the absence of a large sex difference was common to the DPOAEs of humans, rhesus monkeys, sheep, and spotted hyenas. By comparison, the effect sizes for the sex difference in marmoset DPOAEs were quite large (Valero *et al.*, 2008); the sex difference in CEOAEs is currently being studied in marmosets.

Comparisons with non-mammals are complicated by the marked differences in the mechanics of stimulation; nevertheless, Meenderink and Narins (2007) have reported DPOAEs about 40–45 dB weaker than the primary tones in the leopard frog, and Kettembeil *et al.* (1995) have reported DPOAEs about 65–70 dB weaker than the primary tones in chickens and starlings. So when it comes to DPOAE magnitude, humans are more similar to birds than to other mammals. To our knowledge, no one yet has investigated sex differences in the OAEs of frogs or birds.

ACKNOWLEDGMENTS

Initial data collection was supported by NIDCD grants awarded to author G.K.M. (DC00613) and B. L. Lonsbury-Martin (DC00314). The present reanalyses were supported by NIDCD Grant No. DC00153 awarded to author D.M. The authors thank J. Smurzynski and E. G. Pasanen for their assistance. J. C. Loehlin kindly provided comments on a preliminary version of this report. R. Khare assisted with the reanalyses and K. Hatton helped with the references.

[1]We have used an existing dataset (McFadden and Pasanen, 1998) to compare the linear and nonlinear measures that can be obtained from CEOAE responses. The results originally published were from a simple analysis that included both the linear and nonlinear components. To estimate just the nonlinear component of the CEOAE, the response waveform for one click level was doubled in amplitude and then subtracted from the response waveform for a click 6 dB stronger and obtained from that same ear. When the rms level of that difference waveform (expressed in decibels) was averaged across subjects, the magnitudes of both the sex difference and the difference between heterosexual and nonheterosexual females (calculated as effect sizes—described below) were essentially identical to those published using the linear-plus-nonlinear response waveform (both were filtered between 1.0 and 5.0 kHz). That is, there is reason to believe that the sex differences reported here obtained using the Otodynamics nonlinear procedure would have been very similar to those obtained had a linear procedure been used instead.

Berninger, E. (**2007**). "Characteristics of normal newborn transient-evoked otoacoustic emissions: Ear asymmetries and sex effects," Int. J. Audiol. **46**, 661–669.
Bilger, R., Matthies, M. L., Hammel, D. R., and Demorest, M. E. (**1990**). "Genetic implications of gender differences in the prevalence of spontaneous otoacoustic emissions," J. Speech Lang. Hear. Res. **33**, 418–432.
Bonfils, P., Bertrand, Y., and Uziel, A. (**1988**). "Evoked otoacoustic emissions: Normative data and presbycusis," Audiology **27**, 27–35.
Bowman, D. M., Brown, D. K., and Kimberley, B. P. (**2000**). "An examination of gender differences in DPOAE phase delay measurements in normal-hearing human adults," Hear. Res. **142**, 1–11.
Burns, E. M., Arehart, K. H., and Campbell, S. L. (**1992**). "Prevalence of spontaneous otoacoustic emissions in neonates," J. Acoust. Soc. Am. **91**, 1571–1575.
Burns, E. M., Campbell, S. L., Arehart, K. H., and Keefe, D. H. (**1993**). "Long-term stability of spontaneous otoacoustic emissions," J. Assoc. Res. Otolaryngol. **16**, 98.
Burns, E. M., Campbell, S. L., and Arehart, K. H. (**1994**). "Longitudinal measurements of spontaneous otoacoustic emissions in infants," J. Acoust. Soc. Am. **95**, 385–394.
Cacace, A. T., McClelland, W. A., Weiner, J., and McFarland, D. J. (**1996**). "Individual differences and the reliability of 2F1-F2 distortion-product otoacoustic emissions: Effects of time-of-day, stimulus variables, and gender," J. Speech Hear. Res. **39**, 1138–1148.
Cohen, J. (**1992**). "A power primer," Psychol. Bull. **112**, 155–159.
Darlington, R. B. (**1990**). *Regression and Linear Models* (McGraw-Hill, New York).
Dhar, S., Long, G. R., and Culpepper, N. B. (**1998**). "The dependence of the distortion product 2f1-f2 on primary levels in non-impaired human ears," J. Speech Lang. Hear. Res. **41**, 1307–1318.
Dunckley, K. T., and Dreisbach, L. E. (**2004**). "Gender effects on high frequency distortion product otoacoustic emissions in humans," Ear Hear. **25**, 554–564.
Engdahl, B., Arnesen, A. R., and Mair, I. W. S. (**1994**). "Reproducibility and short-term variability of transient evoked otoacoustic emissions," Scand. Audiol. **23**, 99–104.
Franklin, D. J., McCoy, M. J., Martin, G. K., and Lonsbury-Martin, B. L. (**1992**). "Test/retest reliability of distortion-product and transiently evoked otoacoustic emissions," Ear Hear. **13**, 417–429.
Gaskill, S. M., and Brown, A. M. (**1990**). "The behavior of the acoustic distortion product, 2f1-f2, from the human ear and its relation to auditory sensitivity," J. Acoust. Soc. Am. **88**, 821–839.
Gorga, M. P., Neely, S. T., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., Schulte, L., and Jesteadt, W. (**1993**). "A comparison of transient-evoked and distortion product otoacoustic emissions in normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **94**, 2639–2648.
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Guimaraes, P., Zhu, X., Cannon, T., Kim, S. H., and Frisina, R. D. (**2004**). "Sex differences in distortion product otoacoustic emissions as a function of age in CBA mice," Hear. Res. **192**, 83–89.
Harris, F. P., Probst, R., and Wenger, R. (**1991**). "Repeatability of transiently evoked otoacoustic emissions in normally hearing humans," Audiology **30**, 135–141.
Jimenez, A. M., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1999**). "Age-related loss of distortion product otoacoustic emissions in four mouse strains," Hear. Res. **138**, 91–105.
Johansson, M. S. K., and Arlinger, S. D. (**2003**). "Otoacoustic emissions and tympanometry in a general adult population in Sweden," Int. J. Audiol. **42**, 448–464.
Kalluri, R., and Shera, C. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc.

Am. **109**, 622–637.
Keefe, D. H., Gorga, M. P., Jesteadt, W., and Smith, L. M. (**2008**). "Ear asymmetries in middle-ear, cochlear, and brainstem responses in infants," J. Acoust. Soc. Am. **123**, 1504–1512.
Kettembeil, S., Manley, G. A., and Siegl, E. (**1995**). "Distortion-product otoacoustic emissions and their anesthesia sensitivity in the European Starling and the chicken," Hear. Res. **86**, 47–62.
Khalfa, S., Bruneau, N., Roge, B., Gerogieff, N., Veuillet, E., Adrien, J-L., Barthelemy, C., and Collet, L. (**2001**). "Peripheral auditory asymmetry in infantile autism," Eur. J. Neurosci. **13**, 628–632.
Lonsbury-Martin, B. L., and Martin, G. K. (**1988**). "Incidence of spontaneous otoacoustic emissions in macaque monkeys: A replication," Hear. Res. **34**, 313–318.
Lonsbury-Martin, B. L., Cutler, W. M., and Martin, G. K. (**1991**). "Evidence for the influence of aging on distortion-product otoacoustic emissions in humans," J. Acoust. Soc. Am. **89**, 1749–1759.
Margolis, R. H., Saly, G. L., and Keefe, D. H. (**1999**). "Wideband reflectance tympanometry in normal adults," J. Acoust. Soc. Am. **106**, 265–280.
Marshall, L., and Heller, L. M. (**1996**). "Reliability of transient-evoked otoacoustic emissions," Ear Hear. **17**, 237–254.
McFadden, D. (**1993a**). "A speculation about the parallel ear asymmetries and sex differences in hearing sensitivity and otoacoustic emissions," Hear. Res. **68**, 143–151.
McFadden, D. (**1993b**). "A masculinizing effect on the auditory systems of human females having male co-twins," Proc. Natl. Acad. Sci. U.S.A. **90**, 11900–11904.
McFadden, D. (**1998**). "Sex differences in the auditory system," Dev. Neuropsychol. **14**, 261–298.
McFadden, D. (**2002**). "Masculinization effects in the auditory system," Arch. Sex Behav. **31**, 93–105.
McFadden, D. (**2008**). "What do sex, twins, spotted hyenas, ADHD, and sexual orientation have in common?," Perspect. Psychol. Sci. **3**, 309–323.
McFadden, D., and Loehlin, J. C. (**1995**). "On the heritability of spontaneous otoacoustic emissions: A twins study," Hear. Res. **85**, 181–198.
McFadden, D., and Mishra, R. (**1993**). "On the relation between hearing sensitivity and otoacoustic emissions," Hear. Res. **71**, 208–213.
McFadden, D., and Pasanen, E. G. (**1994**). "Otoacoustic emissions and quinine sulfate," J. Acoust. Soc. Am. **95**, 3460–3474.
McFadden, D., and Pasanen, E. G. (**1998**). "Comparison of the auditory systems of heterosexuals and homosexuals: Click-evoked otoacoustic emissions," Proc. Natl. Acad. Sci. U.S.A. **95**, 2709–2713.
McFadden, D., and Pasanen, E. G. (**1999**). "Spontaneous otoacoustic emissions in heterosexuals, homosexuals, and bisexuals," J. Acoust. Soc. Am. **105**, 2403–2413.
McFadden, D., and Shubel, E. (**2003**). "The relationships between otoacoustic emissions and relative lengths of fingers and toes in humans," Horm. Behav. **43**, 421–429.
McFadden, D., Loehlin, J. C., and Pasanen, E. G. (**1996**). "Additional findings on heritability and prenatal masculinization of cochlear mechanisms: Click-evoked otoacoustic emissions," Hear. Res. **97**, 102–119.
McFadden, D., Pasanen, E. G., Raper, J., Lange, H. S., and Wallen, K. (**2006a**). "Sex differences in otoacoustic emissions measured in rhesus monkeys (Macaca mulatta)," Horm. Behav. **50**, 274–284.
McFadden, D., Pasanen, E. G., Weldele, M. L., Glickman, S. E., and Place, N. J. (**2006b**). "Masculinized otoacoustic emissions in female spotted hyenas (Crocuta crocuta)," Horm. Behav. **50**, 285–292.
McFadden, D., Pasanen, E. G., Valero, M. D., Roberts, E. K., and Lee, T. M. (**2008a**). "Effect of prenatal androgens on click-evoked otoacoustic emissions in male and female sheep (Ovis aries)," Horm. Behav. (in press).
McFadden, D., Pasanen, E. G., Valero, M. D., Roberts, E. K., and Lee, T. M. (**2008b**). "Dissociation between distortion-product and click-evoked otoacoustic emissions in sheep (*Ovis aries*)," J. Acoust. Soc. Am. **124**, 3730–3738.
McFadden, S. L., Henselman, L. W., and Zheng, X. Y. (**1999**). "Sex differences in auditory sensitivity of chinchillas before and after exposure to impulse noise," Ear Hear. **20**, 164–174.
Meenderink, S. W. F., and Narins, P. M. (**2007**). "Suppression of distortion product otoacoustic emissions in the anuran ear," J. Acoust. Soc. Am. **121**, 344–351.
Morlet, T., Lapillone, A., Ferber, C., Duclaux, R., Sann, L., Putet, G., Salle, B., and Collet, L. (**1995**). "Spontaneous otoacoustic emissions in preterm neonates: Prevalence and gender effects," Hear. Res. **90**, 44–54.
Morlet, T., Perrin, E., Durrant, J. D., Lapillone, A., Ferber, C., Duclaux, R., Putet, G., and Collet, L. (**1996**). "Development of cochlear active mechanisms in humans differs between gender," Neurosci. Lett. **220**, 49–52.
Moulin, A., Collet, L., Veuillet, E., and Morgon, A. (**1993**). "Interrelations between transiently evoked otoacoustic emissions, spontaneous otoacoustic emissions and acoustic distortion products in normally hearing subjects," Hear. Res. **65**, 216–233.
Nelson, R. J. (**2005**). *An Introduction to Behavioral Endocrinology*, 3rd ed. (Sinauer, Sunderland, MA).
O'Rourke, C., Driscoll, C., Kei, J., and Smyth, V. (**2002**). "A normative study of distortion-product otoacoustic emissions in 6-year-old schoolchildren," Int. J. Audiol. **41**, 162–169.
Porter, C. M., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2006**). "Distortion-product otoacoustic emission suppression growth in normal and noise-exposed rabbits," J. Acoust. Soc. Am. **120**, 884–900.
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 by 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.
Smurzynski, J., and Kim, D. O. (**1992**). "Distortion-product and click-evoked otoacoustic emissions of normally-hearing adults," Hear. Res. **58**, 227–240.
Strickland, E. A., Burns, E. M., and Tubis, A. (**1985**). "Incidence of spontaneous otoacoustic emissions in children and infants," J. Acoust. Soc. Am. **78**, 931–935.
Talmadge, C. L., Long, G. R., Murphy, W. J., and Tubis, A. (**1993**). "New off-line method for detecting spontaneous otoacoustic emissions in human subjects," Hear. Res. **71**, 170–182.
Thornton, A. R. D., Marotta, N., and Kennedy, C. R. (**2003**). "The order of testing effect in otoacoustic emissions and its consequences for sex and ear differences in neonates," Hear. Res. **184**, 123–130.
Torre, P., III, and Fowler, C. G. (**2000**). "Age-related changes in auditory function of rhesus monkeys (Macaca mulatta)," Hear. Res. **142**, 131–140.
Valero, M. D., Pasanen, E. G., McFadden, D., and Ratnam, R. (**2008**). "Distortion-product otoacoustic emissions in the common marmoset (Callithrix jacchus): Parameter optimization," Hear. Res. **243**, 57–68.
Vassilakis, P. N., Meenderink, S. W. F., and Narins, P. M. (**2004**). "Distortion product otoacoustic emissions provide clues to hearing mechanisms in the frog ear," J. Acoust. Soc. Am. **116**, 3713–3726.
Whitehead, M. L., Jimenez, A. M., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). "Time-windowing to click-evoked otoacoustic emissions to increase signal-to-noise ratio," Ear Hear. **16**, 599–611.
Whitehead, M. L., Lonsbury-Martin, B. L., Martin, G. K., and McCoy, M. J. (**1996**). "Otoacoustic emissions: Animal models and clinical observations," in *Clinical Aspects of Hearing, Springer Handbook of Auditory Research*, Vol. **7**, edited by T. R. van de Water, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 199–257.
Whitehead, M. L., Kamal, N., Lonsbury-Martin, B. L., and Martin, G. (**1993**). "Spontaneous otoacoustic emissions in different racial groups," Scand. Audiol. **22**, 3–10.
Wier, C. C., Pasanen, E. G., and McFadden, D. (**1988**). "Partial dissociation of spontaneous otoacoustic emissions and distortion products during aspirin use in humans," J. Acoust. Soc. Am. **84**, 230–237.

Psychophysical and physiological measures of electrical-field interaction in cochlear implants[a)]

Michelle L. Hughes and Lisa J. Stille
Boys Town National Research Hospital, Lied Learning and Technology Center, 425 North 30th Street, Omaha, Nebraska 68131

(Received 21 April 2008; revised 31 October 2008; accepted 3 November 2008)

The primary purpose of this study was to determine whether the electrically evoked compound action potential (ECAP) can be used to predict psychophysical electrical-field interaction patterns obtained with simultaneous stimulation of intracochlear electrodes. The second goal was to determine whether ECAP patterns are affected by recording location because differences might influence the relation between ECAP and psychophysical measures. The third goal was to investigate whether symmetrical threshold shifts are produced with phase inversion of the interaction stimulus. Nine adults with Advanced Bionics cochlear implants participated. ECAP and psychophysical thresholds were obtained for basal, middle, and apical probe electrodes in the presence of a subthreshold interaction stimulus delivered simultaneously to each of seven to eight interaction electrodes per probe. The results showed highly significant correlations between ECAP and psychophysical threshold shifts for all nine subjects, which suggests that the ECAP can adequately predict psychophysical electrical-field interaction patterns for subthreshold stimuli. ECAP thresholds were significantly higher for recordings from the basal (versus apical) side of the probe, which suggests that recording location may affect relations between ECAP and psychophysical measures. Interaction stimulus phase inversion generally produced symmetrical threshold shifts for psychophysical measures but not for half of ECAP measures. [DOI: 10.1121/1.3035842]

PACS number(s): 43.64.Me, 43.64.Pg, 43.66.Cb, 43.66.Nm [BLM] Pages: 247–260

I. INTRODUCTION

Present cochlear implant (CI) speech-processing strategies use pulsatile stimulation to circumvent the electrical-field interaction problems associated with analog stimulation. With pulsatile stimulation, current pulses are interleaved across electrodes so that they do not overlap temporally. The primary disadvantage of fully sequential stimulation is that the per-channel and overall rates of stimulation are limited, thus reducing the amount of information that can be presented to the auditory neurons. As a compromise, the stimulation rate can be increased by presenting pulsatile stimulation simultaneously to two or more far-spaced electrodes chosen to minimize current field overlap. Overlap of current fields can produce summation or subtraction of the electrical field, depending on the relative phase of each current field. Perceptual results can be either a change in loudness (and consequently threshold) or a shift in pitch (Bierer, 2007; Boëx *et al.*, 2003; Buechner *et al.*, 2008; de Balthasar *et al.*, 2003; Donaldson *et al.*, 2005; Favre and Pelizzone, 1993; Firszt *et al.*, 2007; Shannon, 1983; Shannon, 1985; Stickney *et al.*, 2006; Townshend *et al.*, 1987; Wilson *et al.*, 2003). It is therefore of interest to determine the spatial extent to which electrical fields spread, which may help to determine how closely spaced electrodes can be if they are to be simultaneously activated. Further, it is of interest to determine whether the effects of electrical-field interaction are the same for psychophysical and physiological measures. If both measures show similar effects, then physiological measures would have practical utility in predicting time-consuming behavioral measures of electrical-field interaction, which would be particularly valuable for patients who are unable to provide reliable behavioral information. This latter issue was the focus of the present study.

Several earlier studies have evaluated the psychophysical effects of simultaneous stimulation of two intracochlear electrodes. The results show that the behavioral thresholds on a single electrode are lower when a subthreshold stimulus is presented simultaneously in phase to a second electrode. Conversely, thresholds are higher when the added stimulus is inverted in phase (Boëx *et al.*, 2003; de Balthasar *et al.*, 2003; Favre and Pelizzone, 1993; Stickney *et al.*, 2006). These findings indicate summation or subtraction of the electrical fields prior to neural activation. Favre and Pelizzone (1993) used psychophysical threshold to measure electrical-field interaction patterns with in-phase and inverted-phase monopolar stimuli for two subjects implanted with the Ineraid device and noted asymmetrical threshold shifts between the two phase conditions. Although not specifically reported, it appears from their Fig. 2 that the in-phase condition tended to yield more interaction than the inverted-phase condition. Boëx *et al.* (2003) reported symmetrical

[a)]Portions of this work were presented in Hughes, M. L., and Stille, L. J. (**2007**). "Channel interaction patterns with simultaneous stimulation: psychophysical and physiological measures," in *Auditory Research Bulletin* (Advanced Bionics, Valencia, CA), pp. 70–71 and Hughes, M. L., Stille, L. J., and Neff, D. L. (**2007**). "Physiological and psychophysical channel interaction with simultaneous stimulation," Abstracts of the 2007 Conference on Implantable Auditory Prostheses, Lake Tahoe, CA, 15–20 July, p. 105.

shifts between in-phase and phase-inverted conditions for 12 subjects tested with monopolar stimulation and 2 subjects tested with bipolar stimulation. Statistical analyses were not reported, however, and a visual inspection of their Figs. 5 and 7 shows what appear to be asymmetrical shifts for in-phase and inverted-phase conditions for half of the subjects tested in the monopolar condition and for both subjects tested in the bipolar condition. Specifically, there appears to be larger threshold shifts (i.e., more interaction) for the inverted-phase conditions in most of those cases. In a similar experiment, de Balthasar *et al.* (2003) reported symmetrical threshold shifts between in-phase and inverted-phase conditions for three of the four subjects. The subject with asymmetrical threshold shift showed more interaction with the inverted-phase condition. Other studies that used measures of loudness have also shown examples of asymmetrical shifts; in those cases, it appears that more interaction was measured for the in-phase condition (Shannon, 1983; Shannon, 1985). In summary, it appears that phase inversion can produce asymmetrical shifts with no clear outcome as to which phase combination produces more interaction.

Relatively few studies have evaluated physiological effects of simultaneous electrode stimulation in human CI users. The results obtained using the electrically evoked auditory brainstem response (EABR) have shown larger amplitudes and lower thresholds for simultaneous, in-phase stimulation of two electrodes, compared with smaller amplitudes and higher thresholds when the phase of one stimulus was inverted relative to the other (Abbas and Brown, 1988; Gardi, 1985; White *et al.*, 1984). Cortical potentials from animal preparations also have shown lower thresholds with simultaneous in-phase stimulation (Bierer and Middlebrooks, 2004; Middlebrooks, 2004) compared with phase-inverted stimulation (Bierer and Middlebrooks, 2004). Therefore, physiological and psychophysical studies of threshold shifts with stimulus phase have generated similar results. Very few studies have examined current-field interaction from simultaneous electrode activation using the electrically evoked compound action potential (ECAP) in human CI users (e.g., Abbas *et al.*, 2003).

The majority of psychophysical and physiological studies pertaining to electrical-field interaction have focused on examining the effects of stimulus polarity for fixed electrode pairs (as discussed above). Few studies have evaluated spatial interaction patterns along the length of the cochlea, where one stimulus location is fixed and the location of the other stimulus is systematically varied across many electrodes. A few psychophysical studies have examined the effects of electrical-field interaction as a function of location of the interaction stimulus. Favre and Pelizzone (1993) examined the electrical-field interaction patterns for psychophysical thresholds from two Ineraid subjects using monopolar stimulation. The results showed that interaction effects decreased as the distance between simultaneously stimulated electrodes increased. For both subjects, a subthreshold stimulus fixed on the most apical electrode (E1) had a negligible effect for the largest electrode spacing (E1 and E5; 14.4 mm apart). White *et al.* (1984) reported similar findings for psychophysical thresholds in a single subject implanted with an early version of the 16-electrode Clarion device, where an electrode separation of 14 mm produced very little interaction with monopolar stimulation. More recently, Stickney *et al.* (2006) showed less interaction for greater electrode separations for both monopolar and bipolar stimulation, with less interaction overall for bipolar stimulation. Finally, loudness estimates from single subjects in two different studies (Shannon, 1983; Shannon, 1985) showed less interaction with greater separation between the two stimulated electrodes. In some of those examples, there was little to no interaction for electrode separations of 10–12 mm. In summary, psychophysical results show negligible effects of electrical-field interaction for electrode separations of 10–14 mm with monopolar stimulation.

Studies evaluating physiological spatial interaction patterns with simultaneous stimulation along the length of the cochlea are also sparse. There do not appear to be any published reports that comprehensively evaluate physiological electrical-field interaction as a function of electrode separation. Two studies presented physiological data for limited conditions. Abbas and Brown (1988) reported EABR threshold shifts for several Ineraid patients, where stimuli were presented simultaneously to adjacent monopolar (E1 and E2) or overlapping bipolar (E1–E3 and E2–E4) electrode pairs. Threshold shifts were also measured for a slightly wider separation of monopolar electrodes (E1 and E3) and for adjacent bipolar pairs (E1–E2 and E3–E4). Larger threshold shifts (i.e., more interaction) typically occurred for the adjacent monopolar and overlapping bipolar pairs compared with the other respective pairs spaced slightly farther apart. Cortical measures obtained by Bierer and Middlebrooks (2004) from guinea pigs showed a similar trend; larger threshold shifts were obtained for 1.5 mm separation between bipolar electrode pairs compared with 2.25 mm separations. In summary, the limited physiological data suggest less electrical-field interaction with larger separations between simultaneously stimulated electrode pairs, which is consistent with psychophysical findings. We are unaware of any studies that have directly compared physiological and psychophysical electrical-field interactions within individual subjects.

The primary purpose of this study was to determine whether physiological electrical-field interaction patterns measured with the ECAP are predictive of patterns measured psychophysically. If the measures are strongly correlated, then the ECAP may provide an efficient way to predict behavioral electrical-field interaction patterns for simultaneous stimulation. The secondary goal was to examine the effect of the recording electrode location on ECAP electrical-field interaction patterns, as previous studies have shown that ECAP amplitude varies with position of the intracochlear recording electrode (Abbas *et al.*, 1999; Cohen *et al.*, 2004; Frijns *et al.*, 2002). If recording electrode location affects ECAP electrical-field interaction patterns, then the relation between ECAP and psychophysical measures would also be affected by recording electrode location. The third and final goal was to investigate whether phase inversion of the interaction stimulus produces threshold shifts that are symmetrical to the in-phase condition, as well as to determine whether phase inversion affects the physiological and psychophysical

TABLE I. Demographic information for the subjects participating in this study.

Subject No.	Gender	Internal device	Positioner	Implant ear	Age at implant (yr, mo)	Duration CI use (yr, mo)	Duration deafness (yr, mo)	Etiology
C1	F	CII	Yes	Right	18, 4	4, 10	16, 0	Unknown
C6	M	CII	No	Left	64, 4	4, 0	21, 0	Unknown
C7	F	CII	No	Left	57, 2	4, 2	5, 0	Genetic-unspecified
C8	M	CII	No	Right	55, 7	3, 10	0, 3	Sudden
C10	M	90K	No	Right	51, 10	3, 5	14, 0	Unknown
C11	M	90K	No	Right	58, 1	3, 5	15, 0	Unknown
C13	F	CII	No	Left	77, 1	4, 5	20, 0	Unknown
C15	F	90K	No	Left	39	3, 1	34, 0	Unknown
C16	F	CII	Yes	Left	13, 11	Minimal use 2 yr; nonuse 4.5 yr	13, 11	Connexin 26

F=female, M=male, yr=years, and mo=months

electrical-field interaction patterns in the same way. If phase inversion produces symmetrical threshold shifts that are the same for both physiological and psychophysical measures, then it could be presumed that electrical fields sum and subtract in a simple, linear manner. If phase inversion produces asymmetrical threshold shifts and/or affects physiological and psychophysical measures differently, then it could be presumed that mechanisms other than simple field summation are also actively involved.

II. METHODS

A. Subjects

Nine adult CI recipients participated. Six patients were implanted with the Advanced Bionics CII and three received the HiRes 90K device (Advanced Bionics Corporation, Sylmar, CA). Both devices have the same internal chip and 16 electrodes in the intracochlear array. The primary difference between the two devices is that the CII housing is ceramic and the 90K housing is titanium. Both devices had the HiFocus electrode array. Electrodes are numbered sequentially from apex (E1) to base (E16), with a full insertion length of 21 mm and 1.1 mm between electrodes (center to center). Subject C15 had an open circuit on E16; otherwise all subjects had normal electrode impedance as measured with the clinical programming software (SOUND WAVE). Table I lists the subject number, gender, internal device type, whether subjects had an electrode positioner, the ear implanted, age at implant in years (yr) and months (mo), duration of implant use at the time of participation in the study, duration of deafness prior to implantation, and etiology of deafness for each subject. Subject C16 had only used the implant minimally for the first 2 yr following surgery and had been a nonuser for approximately 4.5 yr at the time of participation in this study.

B. Equipment setup

The Bionic Ear Data Collection System (BEDCS; Advanced Bionics Corporation, Sylmar, CA) research platform was used for both physiological (ECAP) and psychophysical stimulus generation and data collection. BEDCS controlled a Platinum Series Processor (PSP) through a clinical programming interface (CPI II) connected to a laptop computer. The PSP was connected to the subject using a Platinum Headpiece. The subject's own processor and headpiece were not used for any portion of data collection.

C. Psychophysical measures

Prior to data collection, an ascending procedure was used to estimate behavioral threshold and upper comfort levels to determine approximate starting and ending current levels for the ECAP measures and for the psychophysical adaptive procedure. The stimulus was a 340 ms pulse train consisting of cathodic-leading, 50 μs/phase, biphasic current pulses with a 10 μs interphase gap presented at a rate of 30 pps (total of 10 pulses). Monopolar stimulation was used for all measures in this study (return electrodes were always the case ground, called IE1 in the CII and IE2 in the 90K). Because the pulse rate and overall duration of a pulse train affect the threshold for that stimulus, it was important that the stimulus used for the psychophysical measures was as similar as possible to the stimulus used to elicit the ECAP so that threshold shifts for the two measures could be compared more directly. Each pulse train was delivered once before the current level was increased. Stimulus levels were increased using log-based increments with the following formula:

$$b = a10^{n/20},$$

where a is the starting current level, n is the step size factor (typically set between 0.6 and 0.7), and b is the next current level (substituted for a on the following iteration). The subject was instructed to indicate when the sound was first heard and when the sound was loud but not uncomfortable. These judgments corresponded to ratings of 1 and 8, respectively, on a visual scale of 0–10, where 0 was no sound and 10 was too loud.

Psychophysical thresholds were then obtained using an adaptive, three-interval, two-alternative, forced choice task. The initial stimulus level for the probe was chosen on an individual basis to be sufficiently audible but well below the estimated upper-loudness level. Thresholds were first obtained for probe electrodes P5, P9, and P12, which represent the apical, middle, and basal cochlear positions, respectively. For probe-electrode-only conditions, the stimulus was randomly presented to either interval 2 or 3 and the subject indicated which interval contained the sound. Thresholds

TABLE II. Probe and interaction-electrode combinations used in this study.

Probe electrode	Interaction electrodes
5	1, 2, 4, 6, 8, 10, 16
9	1, 4, 6, 8, 10, 12, 14, 16
12	1, 7, 9, 11, 13, 15, 16

were then measured for each probe electrode in the presence of a subthreshold, fixed-level interaction stimulus systematically applied in phase to each interaction electrode. For the probe-plus-interaction-electrode conditions, the interaction electrode was stimulated in all three intervals, with the probe electrode stimulated simultaneously in either interval 2 or 3. Again, the subject's task was to choose which interval (2 or 3) contained the sound. Intervals were represented visually with numbered boxes on a computer screen. Each interval was separated by 700 ms. The subject used the computer keyboard to enter the interval number for each response. Feedback was not provided.

Interaction electrodes were chosen to be one, three, and five electrode positions from the probe electrode in both directions, as well as the two most apical and basal electrodes (E1 and E16, respectively). Deviations from this pattern were necessary for interaction electrodes apical to P5 and basal to P12. Table II lists the specific interaction electrodes tested for each probe electrode. E15 was used instead of E16 for subject C15, who had the open circuit on E16.

The current level of the interaction stimulus used for each probe electrode was the same for both ECAP and psychophysical measures. Because the goal was to measure electrical-field overlap, it was important to use the same current level across all interaction electrodes for a given probe electrode. The interaction stimulus level was fixed at one-half (50%) of the behavioral threshold (in microamperes) for each probe electrode obtained with the adaptive procedure.[1] This same current level was also used for the respective ECAP interaction stimulus. In all but one case, the interaction stimulus level was subthreshold across all interaction electrodes. The exception was C6, whose probe thresholds for P5, P9, and P12 were 65.6, 130.1, and 136.6 μA, respectively. The interaction levels for P9 and P12 were therefore 65 and 68 μA, respectively, which was just at or slightly above threshold for E5. In this case, the subject was instructed to choose which interval contained the sound that was different.

Psychophysical thresholds for the interaction condition were determined using a three-down, one-up adaptive procedure, which estimates 79.4% correct (Levitt, 1971). Each block consisted of nine reversals. The initial step size was 2 dB for the first three reversals, followed by 1 dB for the next two reversals, and finally 0.5 dB for the remaining four reversals. Threshold for each block was computed as the mean of the last four reversals. Final threshold was an average of three to five blocks.

D. ECAP measures

ECAP amplitude growth functions were obtained using ascending stimulus levels beginning at a current level near the value that the subject indicated as behavioral threshold (rating of 1) and ending at the level indicated as a rating of 8, with the step size as described in Sec. II C. The stimulus used to evoke the ECAP consisted of 50 μs/phase biphasic current pulses with a 10 μs interphase gap presented in monopolar mode (relative to the extracochlear case electrode), repeated at a rate of approximately 30 pps. Each recorded waveform consisted of 120 averages with a gain of 300. Alternating polarity was used to reduce stimulus artifact.

ECAP growth functions were obtained for P5, P9, and P12. A subthreshold, fixed-level, in-phase interaction stimulus was then delivered simultaneously to another electrode in the array and the ECAP growth function was repeated for each probe electrode. The current level and electrode location for the interaction stimulus were the same as those for the psychophysical measures. For each stimulated electrode (or electrode pair, in the case of simultaneous stimulation), two ECAP growth functions were recorded: one from an apical recording electrode and one from a basal recording electrode, each typically located two positions away from the probe electrode. In two cases, excessive stimulus artifact necessitated changing the recording site farther away from the probe. These exceptions were subject C10, P9 (recorded from E4 instead of E7 for interaction E6) and subject C16, P9 (recorded from E5 instead of E7 for all interaction electrodes). The recording reference electrode was the case electrode (IE1) for CII subjects and the ring electrode (IE1) for 90K subjects.

ECAP waveforms were read into a custom analysis program written in MATLAB (The MathWorks, Inc., Natick, MA). ECAP amplitudes were calculated as the difference between the first negative peak (N1) and the following positive peak or plateau (P2), which were manually marked by the investigators. ECAP threshold was visually determined as the lowest current level that produced a nonzero amplitude that was above the noise floor. The noise floor yielded ECAP thresholds that were approximately 20–40 μV in most cases. Final ECAP threshold was the average of the thresholds for basal and apical recording conditions.

III. RESULTS

A. Effect of recording electrode on ECAP thresholds

Because previous research has shown that relative position of the intracochlear recording electrode can affect ECAP amplitudes (Abbas *et al.*, 1999; Cohen *et al.*, 2004; Frijns *et al.*, 2002), it is reasonable to assume that ECAP thresholds could also potentially be affected. Therefore, it was important to first examine the effect of the recording electrode location on ECAP thresholds obtained with simultaneous stimulation, as this would affect the relation between ECAP and psychophysical measures.

Figure 1 shows four individual examples of ECAP thresholds for a fixed probe electrode (noted in each panel) as a function of interaction-electrode location. In each panel, the symbols represent recordings made from an electrode two positions apical to the probe (filled circles) and two positions basal to the probe (open triangles). The solid gray line represents the average of the two recording positions. In Fig.

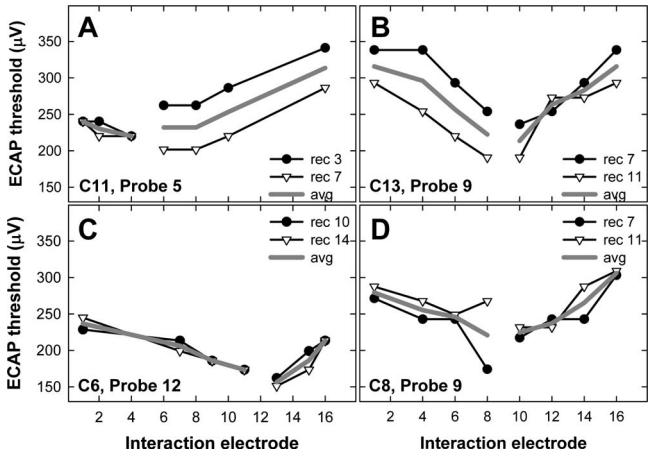


FIG. 1. Individual examples of recording electrode effects on ECAP electrical-field interaction patterns. Subject number and probe electrode are indicated on each graph. The symbols represent probe thresholds as a function of interaction electrode. The filled circles and the open triangles represent the apical and basal recording electrodes, respectively. Specific recording electrode numbers are listed in each figure legend. The solid gray lines represent the average of both recording sites.

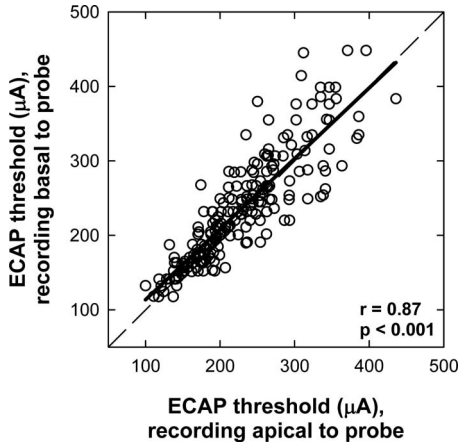


FIG. 2. ECAP thresholds for basal (ordinate) vs apical (abscissa) recording electrode locations (in regard to probe electrode) for all stimulating-electrode conditions in all subjects. The dashed diagonal represents unity. The bold solid line represents linear regression results.

1(a), similar ECAP thresholds were obtained for both recording positions when the interaction stimulus was applied to the apical side of the probe (i.e., low-numbered electrodes) but not for interaction electrodes basal to the probe. In Fig. 1(b), the opposite pattern is seen: similar thresholds are obtained for both recording positions when the interaction stimulus was applied basal to the probe but not for interaction electrodes apical to the probe. In Fig. 1(c), similar thresholds were obtained for basal and apical recording sites across all interaction electrodes. In Fig. 1(d), similar ECAP thresholds were obtained for all interaction electrodes except for E8. A lower ECAP threshold was obtained when the recording electrode (E7) was adjacent to the interaction electrode (E8). This may be due to stimulus artifact adding to the neural response, making it appear larger. As the examples in Fig. 1 illustrate, there was no systematic pattern across subjects or electrodes for ECAP threshold differences between the two recording sites.

Figure 2 shows the ECAP thresholds for basal (ordinate) versus apical (abscissa) recording electrode locations for all stimulating-electrode conditions in all subjects. There was a strong correlation between the two recording sites ($r=0.87, p<0.001$).[2] The slope of the linear regression line was 0.95. Because the data sets were not normally distributed, a Wilcoxon Signed Rank test was used to determine whether ECAP thresholds differed significantly for basal versus apical recording sites. A significant difference was found between the basal and apical recordings ($p=0.003$), with higher thresholds on average for recordings from the basal side of the probe electrode. Figure 2 shows greater differences in ECAP thresholds between the two recording sites when thresholds were higher overall. This likely reflects the larger step sizes used at higher levels due to the log step scale, while the data in Fig. 2 are plotted on a linear scale.

B. Psychophysical versus physiological electrical-field interaction patterns

The primary goal of this study was to compare psychophysical and physiological electrical-field interaction patterns to determine whether ECAP measures can be used as an effective way to predict behavioral electrical-field interaction patterns. For each probe electrode, psychophysical and ECAP thresholds were measured with and without in-phase simultaneous stimulation of a second (interaction) electrode. The interaction stimulus was identical to the stimulus delivered to the probe. Thresholds obtained in the presence of the interaction electrode were subtracted from the probe-only condition to yield the amount of threshold shift. It was hypothesized that the ECAP threshold shift would not be significantly different from the psychophysical threshold shift as a function of interaction-electrode location.

Figure 3(a) shows an example of ECAP (open squares) and psychophysical (filled circles) thresholds for P9 from subject C13. The small symbols with a horizontal dotted line represent the respective probe-only thresholds. Thresholds obtained in the presence of the interaction stimulus are plotted with larger symbols and solid lines as a function of interaction-electrode position. The error bars represent one standard error around the mean (± 1 SEM). For both ECAP and psychophysical measures, probe thresholds were lowest when the interaction stimulus was adjacent to the probe electrode, representing the greatest contribution from the interaction electrode and thus the greatest amount of current-field summation. Probe thresholds increased with greater separation between interaction and probe electrodes, consistent with less overlap and thus less contribution from the interaction electrode. If the probe threshold in the presence of the interaction stimulus is equal to the probe-alone threshold, it is presumed that there is no overlap of current fields from the two simultaneously activated electrodes. Conversely, if the probe threshold in the presence of the interaction stimulus is equal to the difference between probe-alone threshold and interaction stimulus level, it is presumed that there is a complete overlap of current fields from the two simultaneously activated electrodes.

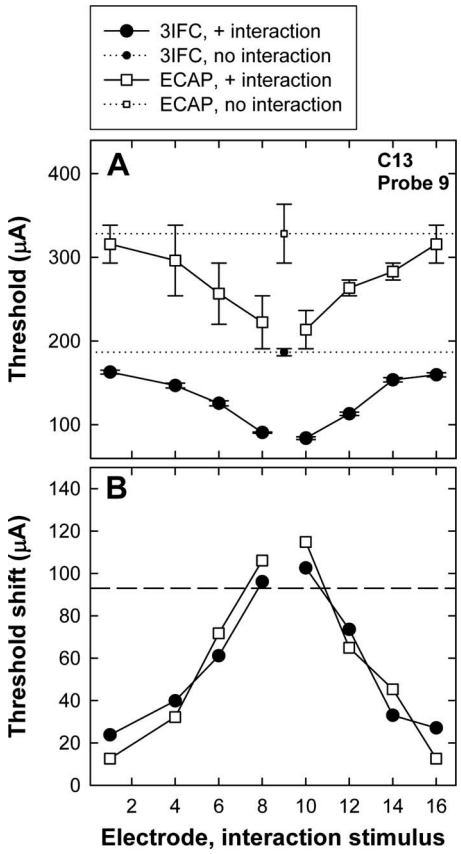


FIG. 3. (a) Example of ECAP (open squares) and psychophysical three-interval forced choice (3IFC; filled circles) thresholds for P9 from subject C13. The small symbols with a horizontal dotted line represent the respective probe-only conditions. The larger symbols with the solid connecting lines represent the thresholds obtained in the presence of the interaction stimulus, plotted as a function of interaction-electrode position. The error bars represent ±1 SEM. (b) Threshold shifts for ECAP (open squares) and psychophysical measures (filled circles) as a function of interaction-electrode location. Threshold shifts were calculated as probe-alone threshold minus probe threshold in the presence of the interaction stimulus. The horizontal line represents the current level of the interaction stimulus.

ECAP and psychophysical threshold shifts were calculated by subtracting the mean threshold in the interaction condition from the mean threshold in the probe-only condition. Figure 3(b) shows the mean threshold shifts for the data presented in Fig. 3(a). The horizontal dashed line indicates the current level of the interaction stimulus. Ideally, the threshold shifts should not be greater than the level of the interaction stimulus. (This situation is described further in Sec. IV.) In this example, smaller threshold shifts occurred with greater separation between probe and interaction electrodes, as expected. The threshold shifts were not significantly different between ECAP and psychophysical measures in this example (paired t test).

Figures 4–6 show the individual threshold-shift patterns for all subjects. Data are plotted as in Fig. 3(b). Data for the apical P5 are shown in Fig. 4, the middle P9 in Fig. 5, and the basal P12 in Fig. 6. An asterisk next to the subject number in each panel indicates a significant correlation (Pearson's r, $p<0.05$) between the ECAP and psychophysical threshold shifts. Correlation coefficients and p values for each subject are listed in Table III. Across Figs. 4–6, 23 of the 27 probe-electrode patterns (85%) exhibited a significant correlation between ECAP and psychophysical data.

Figure 4 (apical probe) shows a significant positive correlation between ECAP and psychophysical threshold shifts for eight of the nine subjects. C6 showed measurable threshold shifts with the ECAP, but virtually no threshold shifts psychophysically, resulting in a lack of correlation between the two measures. This subject's psychophysical probe-alone threshold for P5 was 65.6 μA, which was significantly lower than all other electrodes in all subjects (range of 128.7–196.1 μA). Because the interaction stimulus level was calculated as one-half of the probe-alone threshold, the interaction level in this case (33 μA) may have been too small to have a measurable effect. It is possible that the threshold shifts for C6 obtained with ECAP measures were due to the large amount of variability generally associated with ECAP measures [as illustrated in Fig. 3(a) for C13 by large standard-error bars].

For the middle probe-electrode patterns, Fig. 5 shows a significant positive correlation for eight of the nine subjects. The correlation for C7 was not significant, which was most likely due to the large difference between data points for the interaction stimulus on E4. For the basal probe-electrode patterns in Fig. 6, there were significant positive correlations for seven of the nine subjects. For subject C7, the two patterns were somewhat similar, but the magnitude of the threshold shift was much larger psychophysically than physiologically for the apical interaction electrodes. For subject C15, the patterns appeared quite similar but were not statistically significant. This subject also had one less interaction condition than the other subjects due to the open circuit on E16.

Figure 7 shows the psychophysical threshold shifts plotted relative to ECAP threshold shifts for all electrode conditions within each subject. The subject number is indicated in each panel along with correlation coefficients and p values. The bold solid lines indicate linear regression results and the dashed diagonal lines represent unity. Correlations ranged from 0.70 to 0.91 across subjects. For all subjects, there was a highly significant correlation between ECAP and psychophysical threshold shifts ($p<0.001$). Paired t test results, shown in Table IV, indicate no significant difference between ECAP and psychophysical threshold shifts as a function of interaction-electrode location for five of the nine subjects (C7, C8, C10, C11, and C13), consistent with the hypothesis that similar patterns would be obtained with both measures. For the remaining four subjects, C6 had significantly larger threshold shifts for the ECAP measures, whereas C1, C15, and C16 exhibited greater threshold shifts psychophysically. When data were pooled across all electrodes and all subjects, a strong correlation was observed between ECAP and psychophysical threshold shifts ($r=0.82$, $p<0.0001$). A paired t test for the group data indicated a statistically significant difference between ECAP and psychophysical threshold shifts, with greater shifts occurring for psychophysical thresholds ($t=4.93$, $p<0.001$, df=196).

C. Effect of interaction stimulus phase

The third goal of the present study was to investigate whether phase inversion of the interaction stimulus produced

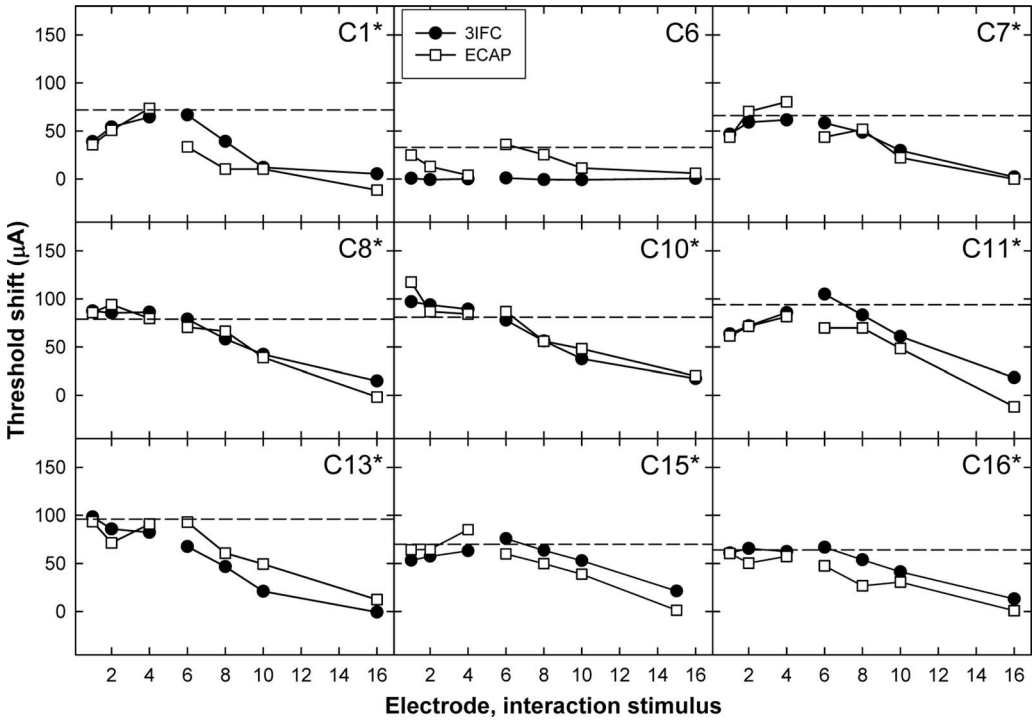


FIG. 4. Psychophysical (3IFC; filled circles) and ECAP (open squares) threshold shifts plotted as a function of interaction-electrode location for apical P5. Each graph represents data from a different subject. The horizontal dashed line represents the level of the interaction stimulus. The asterisks next to the subject numbers indicate a statistically significant correlation between ECAP and psychophysical threshold shifts (see Table III).

a pattern of threshold shifts that was symmetrical to that obtained with the in-phase interaction stimulus. It was also of interest to determine whether phase inversion affected physiological and psychophysical electrical-field interaction patterns in the same way. If phase inversion produces symmetrical threshold shifts that are the same for both physiological and psychophysical measures, then it could be presumed that electrical fields sum and subtract in a simple, linear manner. If phase inversion produces asymmetrical threshold shifts and/or affects physiological and psychophysical measures differently, then it could be presumed that mechanisms other than simple field summation are also actively involved.

For all but subject C11, ECAP and psychophysical

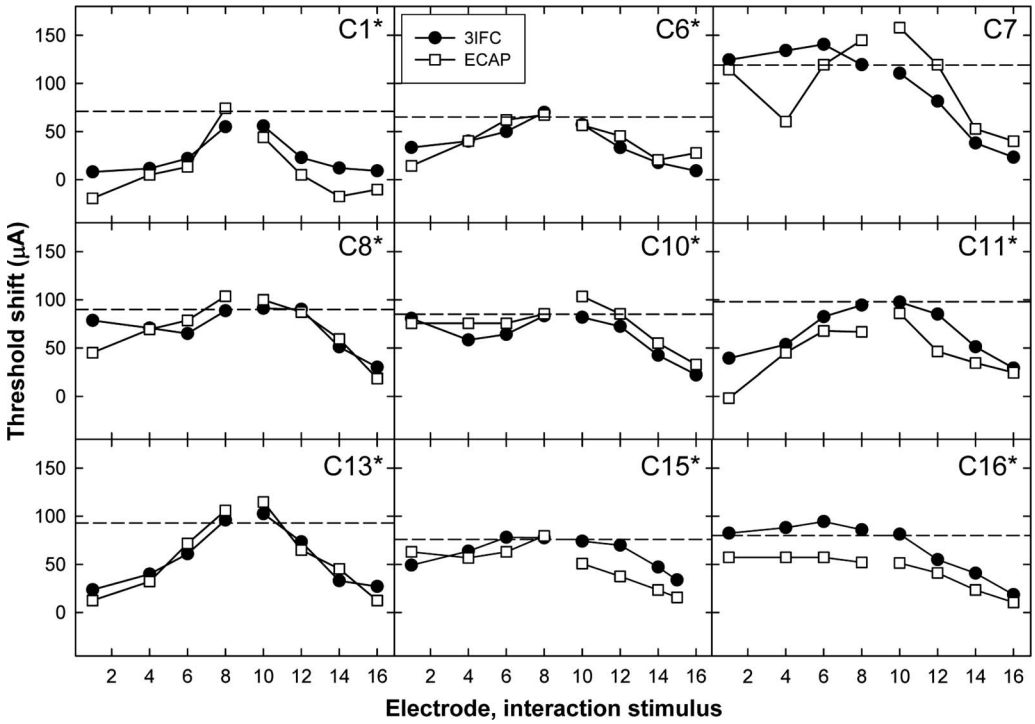


FIG. 5. Psychophysical (3IFC; filled circles) and ECAP (open squares) threshold shifts plotted as a function of interaction-electrode location for middle P9. Data are plotted as in Fig. 4.

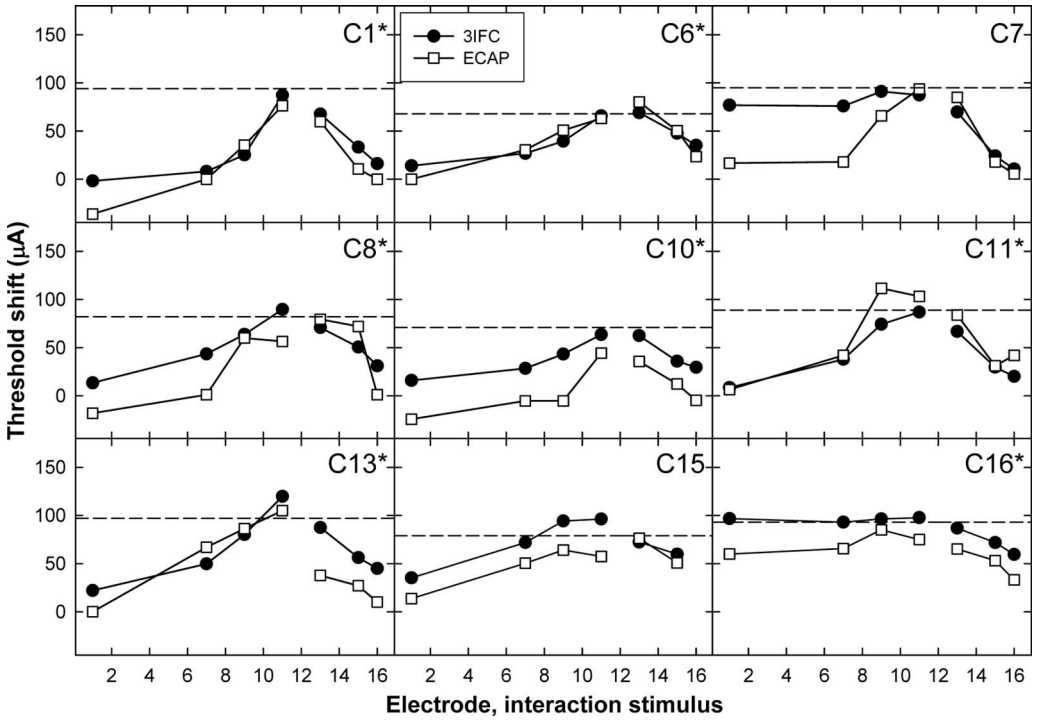


FIG. 6. Psychophysical (3IFC; filled circles) and ECAP (open squares) threshold shifts plotted as a function of interaction-electrode location for basal P12. Data are plotted as in Figs. 4 and 5.

thresholds were measured for P9 in the presence of an inverted-phase interaction stimulus applied to the same electrodes as indicated in Table II. The current level of the inverted-phase interaction stimulus was the same as that used for the in-phase portion of the study. In the inverted-phase condition, pulses delivered to the probe were cathodic leading, while pulses delivered to the interaction electrode were anodic leading. Because ECAP measures used alternating polarity to eliminate stimulus artifact, the stimulus polarity for the interaction electrode was always opposite of that presented to the probe for the inverted-phase condition. All other aspects of data collection were the same as described for the in-phase condition.

Figure 8 shows the threshold shifts for the psychophysical task (filled circles) and for the ECAP (open squares) as a function of interaction electrode for each subject. The solid lines connecting the symbols represent data from the in-phase condition (taken from Fig. 5), and the dashed lines represent the data from the inverted-phase condition. The horizontal solid gray lines represent the current level of the interaction stimulus, where the positive values are in phase with the probe and the negative values are phase-inverted relative to the probe. Ideally, all symbols should fall between the solid gray lines and should be symmetric about $y=0$. In general, most of the data followed these expected trends. However, differences in threshold shift between in-phase and inverted-phase conditions can be seen in several of the individual graphs. For example, ECAP threshold shifts were larger than the magnitude of the interaction stimulus for subjects C6 and C15 in the inverted-phase condition but not for the in-phase condition.

TABLE III. Correlation coefficients (r values) and significance (p values) for comparison of threshold shifts between psychophysical and physiological measures (data presented in Figs. 4–6). The asterisks denote statistical significance ($p<0.05$).

	Probe 5		Probe 9		Probe 12	
Subject	r	p	r	p	r	p
C1	0.84	0.02*	0.94	<0.001*	0.94	0.002*
C6	0.32	0.48	0.83	0.01*	0.95	0.001*
C7	0.92	0.003*	0.61	0.1	0.63	0.1
C8	0.97	<0.001*	0.83	0.01*	0.81	0.03*
C10	0.95	<0.001*	0.92	0.001*	0.94	0.002*
C11	0.90	0.006*	0.87	0.005*	0.96	<0.001*
C13	0.91	0.004*	0.97	<0.001*	0.81	0.03*
C15	0.80	0.03*	0.72	0.046*	0.76	0.078
C16	0.91	0.005*	0.98	<0.001*	0.88	0.009*

To summarize the data in Fig. 8, the mean threshold shift across all interaction electrodes was calculated for both phase conditions of the ECAP and psychophysical data. Figure 9 shows the mean threshold shifts for each subject. The black bars represent the in-phase condition and the gray bars represent the inverted-phase condition. Figure 9(a) shows the psychophysical data and Fig. 9(b) shows the ECAP data. Because threshold shifts were calculated as the no-interaction condition minus the interaction condition, the inverted-phase condition typically resulted in negative values whereas the in-phase condition typically yielded positive values. Therefore, the inverted-phase threshold shifts were multiplied by −1 prior to calculating the average across interaction electrodes so that threshold shifts could be compared for the two phase conditions.[3]

For each subject, a two-way analysis of variance with measure and phase as factors was used to evaluate whether threshold shifts across interaction electrodes (data from Fig.

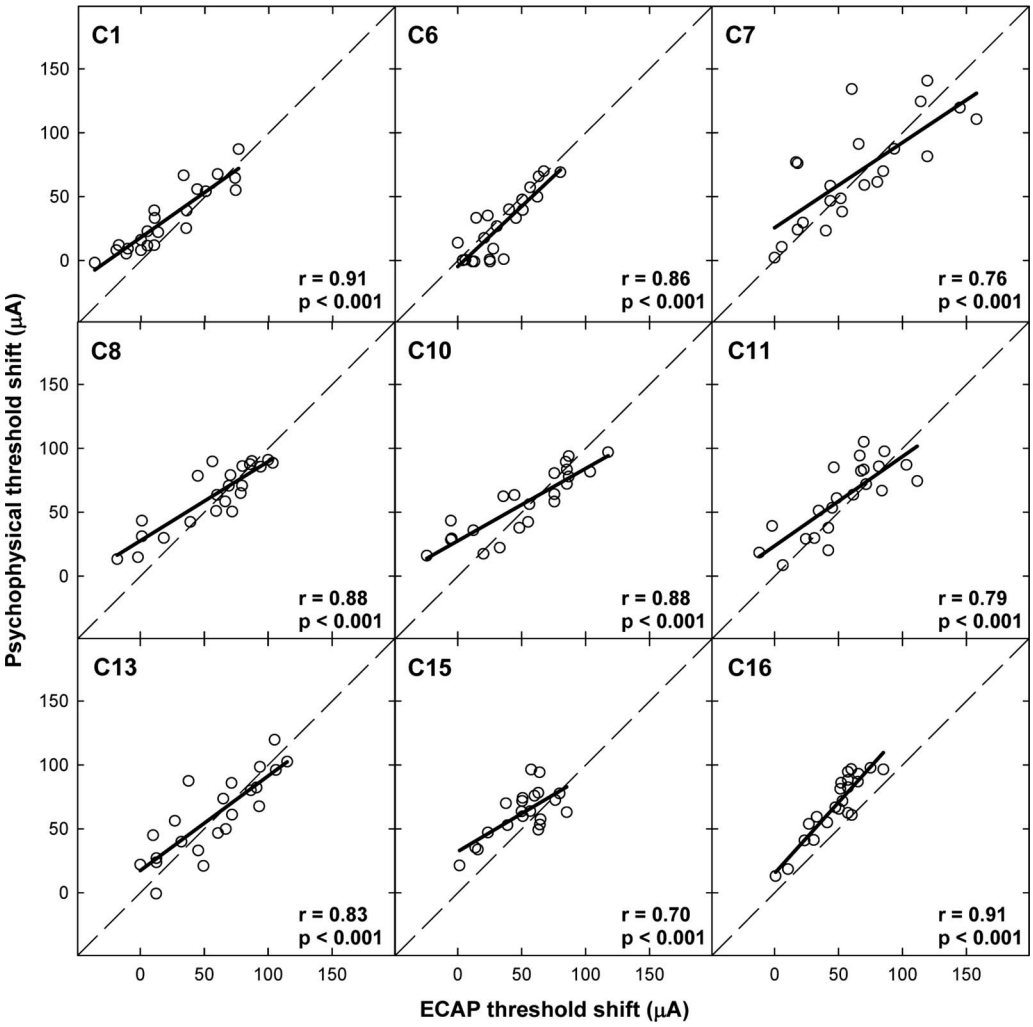


FIG. 7. Individual scatter plots comparing psychophysical and ECAP threshold shifts across all electrode conditions within a subject. Subject number, correlation coefficient (r value), and significance level (p value) are indicated on each graph. The diagonal dashed lines represent unity and the bold solid lines represent the linear regression results.

8) were the same for the in-phase versus inverted-phase conditions. The asterisks above the bars in Fig. 9 indicate statistically significant differences between the in-phase and inverted-phase conditions ($p<0.05$). For psychophysical data, only subject C8 demonstrated significantly more interaction (i.e., greater threshold shift) for the in-phase stimulus compared with the inverted-phase stimulus. For ECAP data, subjects C1, C6, and C16 exhibited significantly larger mean threshold shifts for the inverted-phase condition and subject C10 showed a significantly larger mean threshold shift in the in-phase condition. For five subjects (C1, C6, C8, C10, and C16), phase inversion did not affect ECAP threshold shifts and psychophysical threshold shifts in the same way. Specifically, C1, C6, C10, and C16 showed asymmetrical threshold shifts as a function of interaction stimulus phase for ECAP measures but not psychophysically. C8 showed an asymmetrical threshold shift for psychophysical measures but not with the ECAP. The remaining three subjects (C7, C13, and C15) demonstrated no significant difference in threshold shifts for the two phase conditions with either measure.

TABLE IV. Statistical results from paired t tests for psychophysical vs ECAP threshold shifts for data collapsed across the three probe electrodes within each subject (see Fig. 7). The asterisks denote statistical significance ($p<0.05$). df=degrees of freedom.

Subject	t	p	df
C1	4.05	<0.001*	21
C6	−2.38	0.03*	21
C7	0.64	0.53	21
C8	1.57	0.13	21
C10	1.13	0.27	21
C11	1.77	0.09	21
C13	0.57	0.58	21
C15	3.30	0.004*	20
C16	9.40	<0.001*	21

IV. DISCUSSION

The primary goal of this study was to evaluate the relation between physiological and psychophysical measures of electrical-field interaction in CI recipients. First, to determine whether recording site might affect this relation, we examined the effect of recording electrode location on electrical-field interaction patterns measured with the ECAP. Because recording electrode location was found to significantly affect

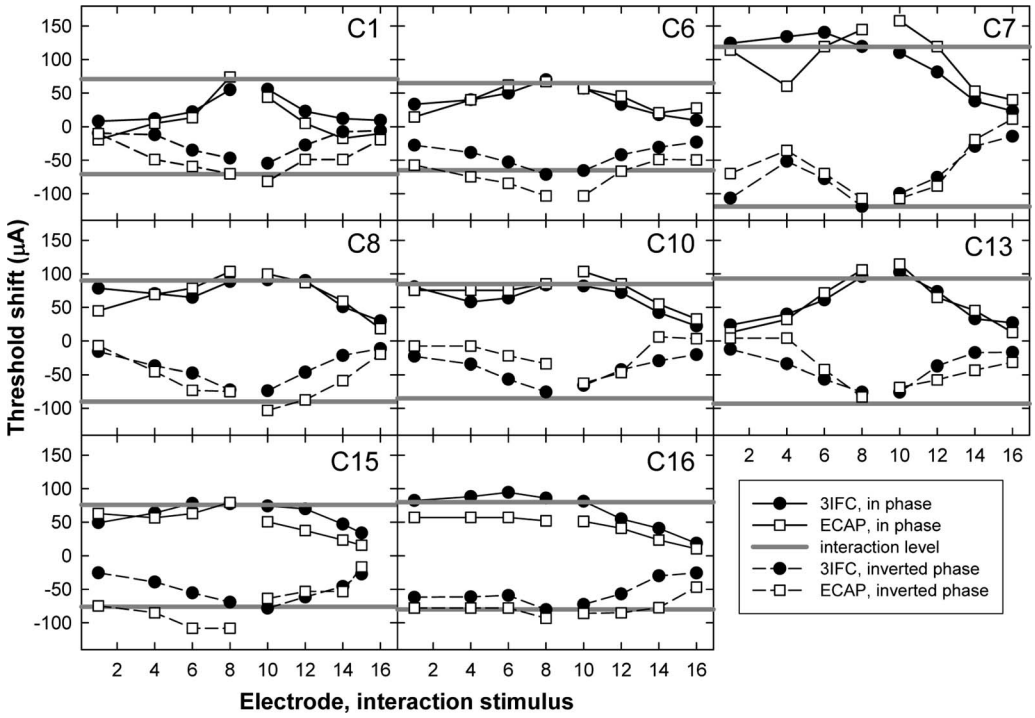


FIG. 8. Individual interaction patterns for P9; each panel represents data from a different subject. ECAP (open squares) and psychophysical (3IFC; filled circles) threshold shifts are plotted as a function of interaction-electrode location. The symbols connected with the solid lines represent the threshold shifts with the addition of an in-phase interaction stimulus and the dashed lines represent the inverted-phase condition. The solid gray lines represent the interaction stimulus level, with negative values representing the inverted-phase condition.

the ECAP results, ECAP thresholds were averaged for two recording sites for comparison with psychophysical thresholds. The results showed a strong correlation between the ECAP and psychophysical spatial interaction patterns, which suggests that the ECAP has practical utility in predicting psychophysical measures of electrical-field interaction. Finally, the effect of interaction stimulus phase was examined for both ECAP and psychophysical measures. The results showed different effects of phase between the two measures. These results are discussed further in Secs. IV A through IV C.

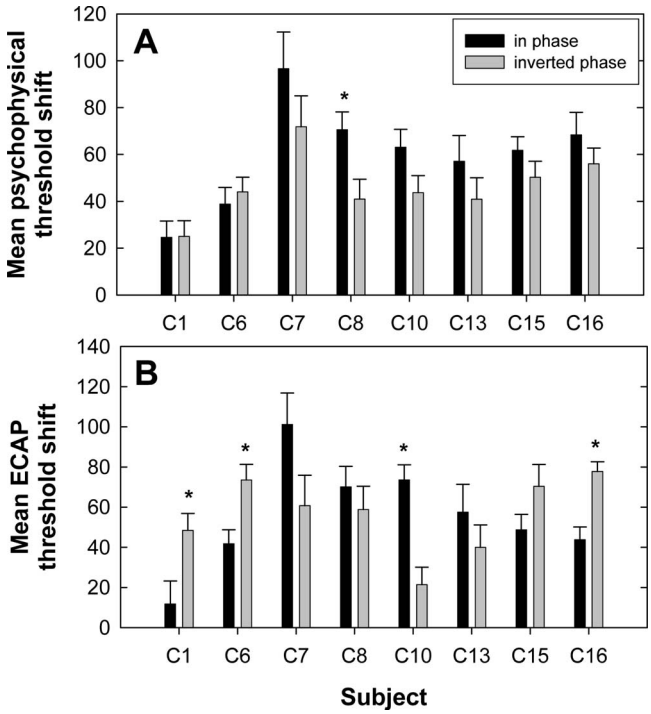


FIG. 9. Mean threshold shifts (+1 SEM) across P9 interaction electrodes for in-phase (black bars) and inverted-phase (gray bars) conditions for each subject. Data were calculated from Fig. 8: (a) psychophysical threshold shifts; (b) ECAP threshold shifts. The asterisks indicate a statistically significant difference between phase conditions ($p<0.05$).

A. Effect of recording electrode on ECAP thresholds

As illustrated by the examples in Fig. 1, the recording electrode location can strongly affect the ECAP threshold measures. As Fig. 2 shows, greater discrepancies between apical and basal recordings were seen for higher thresholds, likely due to the larger step sizes used at higher stimulus levels. Several previous studies have reported that ECAP amplitude varies as the position of the intracochlear recording electrode changes relative to the stimulating electrode (Abbas *et al.*, 1999; Cohen *et al.*, 2004; Frijns *et al.*, 2002). Specifically, ECAP amplitude tends to decrease as the recording site is located farther away from the stimulating site. The recording electrode measures the voltage change associated with neural discharge. Because the cochlea is filled with fluid, that voltage is conducted along the length of the cochlea allowing relatively large voltages to be measured from distal recording sites. In some cases, however, the measured voltage can approach zero for the largest separation between the stimulating and recording electrodes (e.g., Frijns *et al.*, 2002). While these studies have reported measurable effects of recording electrode position on suprathreshold ECAP amplitudes, no studies have assessed the effect of recording

electrode position on ECAP thresholds. However, if ECAP amplitudes are smaller at farther recording sites, it is probable that thresholds will be higher at farther recording sites.

The literature is inconsistent in regard to the symmetry of ECAP amplitudes obtained from recording sites that are apical versus basal to the stimulated electrode. Abbas *et al.* (1999) compared the difference in ECAP amplitudes between the recordings made at the basal and apical sites that were equidistant from the stimulated electrode and found virtually no difference in amplitude between the two recording positions. In contrast, Frijns *et al.* (2002) reported a tendency for amplitudes to be larger when recorded from a location apical to the stimulated electrode than from an equidistant basal location. They attributed those findings to the tapered anatomy of the cochlea from base to apex, which likely puts electrodes in closer proximity to neural elements at the apical end. ECAP data from the present study showed that higher thresholds occurred more often for recordings from the basal side of the stimulating electrode, which is consistent with the amplitude data reported by Frijns *et al.* (2002). The variability in ECAP thresholds and amplitudes across recording electrode locations suggests that ECAP recordings should be averaged across more than one recording location for a more robust measurement, particularly if ECAP measures are to be compared with psychophysical measures within subjects.

As shown in Fig. 3(a), mean ECAP thresholds typically exhibited a much larger standard error than psychophysical thresholds. This may be due to several measurement-related issues. First, only two ECAP threshold measures were obtained for the average (each from a different recording electrode), whereas psychophysical results were an average of 3–5 threshold estimates. Second, a log step size was used for both ECAP and psychophysical measures, and since ECAP thresholds were always higher than psychophysical thresholds, the step size was therefore larger for ECAP measures. The third reason is that the psychophysical measures used an adaptive procedure where step size was systematically reduced several times near threshold. ECAP thresholds were obtained using an ascending (rather than adaptive) procedure with a fixed log-based step size. It is possible that a smaller SEM would have resulted from a repeated reduction in step size near threshold, as in the psychophysical procedure. The relatively large SEM for ECAP measures suggests that alternative methods for determining final ECAP thresholds should be considered when relating these measures to psychophysical results. Specifically, it may be worthwhile to investigate the efficacy of using an adaptive procedure to obtain physiological thresholds.

B. Psychophysical versus physiological electrical-field interaction patterns

The present results represent a comprehensive first report of electrical-field interaction patterns obtained with the ECAP in human CI recipients. In general, the results from the present study showed smaller threshold shifts (i.e., less interaction) with greater separation between simultaneously activated electrodes for both measures. These trends are consistent with previous psychophysical threshold studies (Favre and Pelizzone, 1993; Stickney *et al.*, 2006; White *et al.*, 1984) and with studies that measured spatial interaction physiologically using either EABR (Abbas and Brown, 1988) or cortical responses (Bierer and Middlebrooks, 2004).

It was hypothesized that ECAP threshold shifts would not be significantly different from psychophysical threshold shifts as a function of interaction-electrode location. Paired *t* tests revealed no significant difference between ECAP and psychophysical threshold shifts for five of the nine subjects. However, for four individual subjects (see Table IV) and for group data, there was a significant difference in the amount of threshold shift between the two measures. Specifically, larger shifts occurred for psychophysical thresholds than for ECAP thresholds for three of those four subjects and for the group data as a whole. It is interesting to note that those three subjects (C1, C15, and C16) were the only subjects in this study who had either been born deaf or became deaf in early childhood. Two of those subjects (C1 and C16) were the only two subjects with a positioner. It seems unlikely, however, that the positioner would affect one measure and not the other.

For one subject (C1), ECAP thresholds in the presence of the in-phase interaction stimulus were higher than the probe-alone threshold for 5 of the 22 electrode pairs. This trend is opposite of what is expected based on previous physiological findings. We would expect thresholds to be lower than the probe-alone condition due to the summation of current fields resulting in less current needed from the probe electrode to elicit threshold (e.g., Abbas and Brown, 1988; Bierer and Middlebrooks, 2004; Gardi, 1985; Middlebrooks, 2004; White *et al.*, 1984). Because threshold shifts were calculated as the probe-alone threshold minus the threshold obtained in the presence of the interaction stimulus, this resulted in negative threshold shifts for several of the ECAP measures. As a result, ECAP measures yielded smaller threshold shifts compared with psychophysical measures. It is possible that the higher thresholds obtained in the interaction condition may be attributed to the larger SEM observed with ECAP measures, as shown in Fig. 3(a) and discussed in Sec. IV A.

For the remaining two subjects (C15 and C16), psychophysical threshold shifts were not only larger than ECAP threshold shifts but were also larger than the magnitude of the interaction stimulus for 5 of the 21 and 10 of the 22 electrode pairs, respectively (see Figs. 4–6). These subjects' psychophysical responses were highly repeatable across blocks, so measurement variability was not an issue. In theory, a probe threshold shift should not be larger than the magnitude of the interaction stimulus. If it is, this suggests that a mechanism other than simple current summation is contributing to detection of the combined stimuli. In other words, the total amount of current delivered to the probe and interaction electrodes together is less than the amount of current needed to elicit threshold on the probe electrode alone. As the interaction stimulus is moved farther from the probe, the amount of electrical-field overlap is reduced. However, the subthreshold interaction stimulus may recruit enough individual nerve fibers that are adjacent to the region stimulated by the probe, potentially resulting in a broader excita-

tion pattern. It is possible that subthreshold neural activity in the vicinity of the interaction electrode produces more aggregate neural activity, leading to a lower perceptual detection threshold. This theory can be likened to the difference in neural recruitment patterns between monopolar and bipolar stimulation configurations, where lower current levels are needed for broader stimulation patterns due to an increase in the total number of neurons recruited. Perhaps psychophysical threshold shifts that are larger than the magnitude of the interaction stimulus indicate areas of greater neural survival or possibly a larger number of low-threshold, high-spontaneous-rate fibers in the region of the interaction electrode. Alternatively, the subthreshold interaction stimulus may change the "center of gravity" of the summed electric field, which shifts the area of excitation toward the interaction electrode. This is the concept used to achieve intermediate or virtual channels using current steering (e.g., Donaldson *et al.*, 2005; Firszt *et al.*, 2007; Townshend *et al.*, 1987; Wilson *et al.*, 2003). It is possible that the threshold of the interaction electrode was lower than the threshold of the probe electrode, and detection was largely based on percepts in the vicinity of the interaction electrode. This was likely the case for C16, whose behavioral thresholds were lowest for the apical electrodes. Recall that the stimulus level for the interaction electrode was fixed at one-half the psychophysical threshold of the probe electrode. If the interaction electrode had a lower threshold than the probe (but still higher than the level of the interaction stimulus), it would take less current from the probe for stimulus detection in the region of the interaction electrode if electrical fields from both electrodes overlapped sufficiently. However, only one subject (C6) had a large enough difference in thresholds across the electrode array such that an interaction stimulus presented at one-half the threshold of one electrode would have been audible on another (interaction) electrode, and that subject did not exhibit psychophysical threshold shifts that were larger than the magnitude of the interaction stimulus on the apical electrodes (see Figs. 4–6).

In a previous study, we compared ECAP and psychophysical spatial interaction patterns obtained with nonsimultaneous stimulation using a forward-masking paradigm (Hughes and Stille, 2008). Results showed that while there was a significant correlation between the two measures ($r=0.55$, $p<0.0001$), the correlation was not strong enough to suggest using ECAP measures alone to predict psychophysical electrical-field interaction patterns. In contrast, the present results showed a highly significant correlation ($p<0.001$) between ECAP and psychophysical threshold shifts for all individual subjects (see Fig. 7) and for group data ($r=0.82$, $p<0.0001$). Two methodological differences between Hughes and Stille (2008) and the present study likely contributed to the higher correlation found in the present study. First, the previous study used nonsimultaneous stimulation (forward masking) to measure spatial interaction patterns. Spatial interaction patterns measured with forward masking reflect the overlap of neural populations recruited by the masker and probe electrodes, as well as relative temporal aspects of the stimulated neurons (i.e., neural refractory-recovery mechanisms). The use of simultaneous stimulation in the present study allowed more direct assessment of spatial interaction, because it is presumed that electrical fields sum or subtract prior to neural activation. Therefore, temporal aspects related to neural refractory periods do not confound the relationship between ECAP and psychophysical threshold measures as they do with forward masking. Second, Hughes and Stille (2008) used fast-rate pulse trains for psychophysical measures and single pulses repeated at a relatively slow rate for ECAP measures. This resulted in the use of overall higher current levels for the ECAP measures due to less temporal integration. In the present study, care was taken to use stimuli that were as similar as possible (i.e., slow rate) for both ECAP and psychophysical measures to reduce confounding effects of stimulus current level associated with temporal integration. However, it should be emphasized that even if stimuli are similar for both ECAP and psychophysical threshold measures, there are differences in the respective underlying mechanisms. These differences likely contribute to the variance in the relationship between the two measures. The ECAP is a synchronized response of auditory nerve fibers to a single current pulse, which is repeated at a relatively slow rate to avoid neural refractory and adaptation effects in the averaged response. In contrast, psychophysical thresholds involve temporal integration across the duration of the pulse train. The findings of the present study are consistent with other studies that have shown stronger correlations between physiological and psychophysical measures when the same stimulus was used for both measures (e.g., Brown *et al.*, 1994; Brown *et al.*, 1996).

Another difference in results between the present study and Hughes and Stille (2008) is that the forward-masking study showed more masking overall for ECAP measures than for psychophysical measures; the opposite result was found for electrical-field interaction in the present study. In Hughes and Stille (2008), the greater amount of masking for ECAP patterns was attributed to the use of higher current levels for ECAP than for psychophysical measures, which may have led to more current spread and thus more interaction (e.g., Eisen and Franck, 2005). In the present study, more interaction was measured psychophysically; however, the difference between ECAP and psychophysical measures was not as large as in the forward-masking study (Hughes and Stille, 2008). In general, the present results for simultaneous stimulation show a strong correlation between ECAP and psychophysical measures, which suggests that ECAP measures can be used as an effective alternative to time-consuming psychophysical measures of electrical-field interaction.

Although it was not the goal of this study to explicitly measure an optimal electrode separation that results in no interaction, it can be seen in Figs. 4–6 that electrode interactions were, in many cases, minimal at the largest electrode spacing. These results are consistent with those reported in several other studies (Favre and Pelizzone, 1993; Shannon, 1983; Shannon, 1985; White *et al.*, 1984). The partially simultaneous high-resolution sound processing (HiRes-P) used in current Advanced Bionics devices effectively doubles the overall stimulation rate by activating two electrodes at the same time that are located half the array apart. When HiRes programs are created, behavioral thresholds (T-levels) are

typically set to zero or some small percentage (e.g., 10%) of the most-comfortable levels (M-levels). Thus, the two simultaneously activated electrodes could potentially be activated at subthreshold levels, as was the case in the present study, or a combination of sub- and suprathreshold levels. Results from the present study can provide insight about channel interactions that may occur with a partially simultaneous strategy. As can be seen in Fig. 5, some subjects had virtually no interaction for electrodes spaced half the array apart (e.g., C1, C11, and C13), whereas others had significant interactions (e.g., C10, C15, and C16). These differences may account for differences in performance between the partially simultaneous (HiRes-P) and fully sequential (HiRes-S) versions of HiRes within and across subjects (e.g., Buechner *et al.*, 2005). Further research is needed to determine whether ECAP electrical-field interaction patterns such as those used here have potential clinical utility for determining an optimal processing strategy (e.g., HiRes-P versus HiRes-S) on an individual basis. It is also worth investigating whether these ECAP measures have clinical utility in estimating channel independence in regard to speech perception. That is, can these measures be used to estimate the number of channels that result in optimal speech-perception performance?

Finally, the presence of an electrode positioner did not seem to affect electrical-field interaction patterns for either ECAP or psychophysical measures. Only two subjects (C1 and C16) had an electrode positioner, so there is insufficient power for a statistical analysis. However, interaction patterns in Figs. 4–6 for those two subjects do not appear narrower than the other subjects who did not have a positioner. These results are consistent with those reported by Boëx *et al.* (2003) and Stickney *et al.* (2006), who compared electrical-field interactions for various electrode array types, including those with a positioner.

C. Effect of interaction stimulus phase

Several studies have reported results that support the notion of direct vector summation of current for in-phase and phase-inverted simultaneous stimulations of two electrodes; however, examples are seen in many of those studies in which the amount of interaction (or threshold shift) was not equal for in-phase versus phase-inverted conditions (e.g., Boëx *et al.*, 2003; de Balthasar *et al.*, 2003; Favre and Pelizzone, 1993; present study). Asymmetries may be due to differences in how the shape of the electrical field changes with level. The inverted-phase condition essentially produces a smaller resultant electrical field when fields of opposite polarity sum. Therefore, more current must be delivered to the probe electrode to compensate for the cancellation of current from the inverted-phase stimulus on the other electrode in order to achieve threshold. The shapes of the current fields produced by each electrode are likely to be slightly different due to differences in current level, electrode impedance, and geometry and impedance of the tissue in the vicinity of each electrode (e.g., Finley *et al.*, 1990; Frijns *et al.*, 1995; Frijns *et al.*, 1996; Kral *et al.*, 1998; Ruddy and Loeb, 1995). Thus, the shape of the summed field may differ depending on the relative current levels of the two contributing electrodes. Differences in anatomical geometry, impedance, etc., across subjects may explain why some individuals showed asymmetry between phase conditions while others did not.

Results from Fig. 9 show that phase inversion did not affect psychophysical and ECAP measures in the same way for more than half of the subjects. Specifically, C8 demonstrated asymmetrical shifts for psychophysical thresholds but not for the ECAP, and four other subjects (C1, C6, C10, and C16) demonstrated asymmetrical threshold shifts for the ECAP but not psychophysically. Recall that psychophysical measures were obtained with cathodic-leading pulse trains presented to the probe electrode, whereas ECAP measures were obtained with pulses of alternating polarity for artifact reduction. Therefore, half of the ECAPs in each averaged response were obtained with anodic-leading pulses. A recent study by Macherey *et al.* (2008) showed that anodic-leading pulses resulted in more effective electrical stimulation in human CI users than cathodic-leading pulses. It is not clear whether electrical fields from two electrodes stimulated with anodic-leading pulses will sum in the same way as electrical fields from two electrodes stimulated with cathodic-leading pulses. However, given the difference in physiological responses between cathodic and anodic stimulus polarities as reported by Macherey *et al.* (2008), it is possible that the alternating polarity paradigm used for artifact reduction in the ECAP measures contributed to the differences seen in Fig. 9 between ECAP and psychophysical measures.

V. CONCLUSIONS

The primary purpose of this study was to determine whether physiological electrical-field interaction patterns measured with the ECAP are predictive of patterns measured psychophysically. Results showed a highly significant correlation ($r=0.82$, $p<0.0001$) between ECAP and psychophysical threshold shifts obtained with simultaneous stimulation of two electrodes. There was no significant difference between ECAP and psychophysical threshold shifts for five of the nine subjects, which was consistent with the hypothesis. However, group data showed statistically significantly larger shifts for psychophysical thresholds than for the ECAP. In general, ECAP data predicted psychophysical data with enough confidence to suggest that ECAP measures can adequately predict psychophysical electrical-field interaction patterns for subthreshold stimuli. ECAP data from the present study also showed that higher thresholds occurred more often for recordings made from the basal side of the stimulating electrode compared with the apical side. The variability in ECAP thresholds and amplitudes across recording electrode locations suggests that ECAP measures should be averaged across more than one recording electrode location, particularly if ECAP measures are to be compared with psychophysical measures within subjects. Finally, phase inversion of the interaction stimulus resulted in asymmetrical threshold shifts for some subjects, particularly for ECAP measures. This finding may be attributed to differences in the relative current levels of the two contributing electrodes and how their respective electrical fields change shape with level.

Phase inversion did not affect psychophysical and ECAP measures in the same way for more than half of the subjects. This result may be influenced by the alternating polarity paradigm used to elicit ECAP measures.

ACKNOWLEDGMENTS

This research was funded by the NIH/NIDCD, Grant No. R03 DC007017. Human-subjects recruitment was supported by the NIH/NIDCD, Grant No. P30 DC04662. The authors thank Hongyang Tan of Boys Town National Research Hospital and Leonid Litvak of Advanced Bionics Corporation for technical assistance with programs used for data analysis and collection, respectively. We also thank Donna Neff, Walt Jesteadt, and two anonymous reviewers for helpful comments on earlier drafts of this manuscript, and the nine subjects who participated in this study.

[1]Exceptions were P5, P9, and P12 for subject C1 and P9 in C7. For those initial two subjects, the interaction stimulus level was 20% of the difference between behavioral threshold and loudness comfort. This corresponded to 41% of probe threshold for P5 and P9 and 58% for P12 for subject C1, and 68% of P9 threshold for C7. This protocol was subsequently changed to one-half of the probe threshold to ensure the interaction stimulus level was sufficiently below threshold.

[2]All statistical calculations reported in this study were made with SIGMA STAT 3.0 (SPSS Inc., Chicago, IL). An alpha level of 0.05 was used to determine statistical significance.

[3]This method was preferred over taking the absolute value because it captured instances where the threshold shift went in the opposite direction from what was expected (e.g., inverted-phase interaction stimulus produced thresholds that were lower than the no-interaction condition).

Abbas, P. J., and Brown, C. J. (**1988**). "Electrically evoked brainstem potentials in cochlear implant patients with multi-electrode stimulation," Hear. Res. **36**, 153–162.

Abbas, P. J., Brown, C. J., Hughes, M. L., Etler, C. P., Behrens, A., and Dunn, S. M. (**2003**). "The electrically evoked compound action potential: Channel interaction measures," Abstracts of the 2003 Conference on Implantable Auditory Prostheses, Pacific Grove, CA, p. 41.

Abbas, P. J., Brown, C. J., Shallop, J. K., Firszt, J. B., Hughes, M. L., Hong, S. H., and Staller, S. J. (**1999**). "Summary of results using the nucleus CI24M implant to record the electrically evoked compound action potential," Ear Hear. **20**, 45–49.

Bierer, J. A. (**2007**). "Threshold and channel interaction in cochlear implant users: Evaluation of the tripolar electrode configuration," J. Acoust. Soc. Am. **121**, 1642–1653.

Bierer J. A., and Middlebrooks,J. C. (**2004**). "Cortical responses to cochlear implant stimulation: Channel interaction," J. Assoc. Res. Otolaryngol. **5**, 32–48.

Boëx, C., de Balthasar, C., Kós, M.-I., and Pelizzone, M. (**2003**). "Electrical field interactions in different cochlear implant systems," J. Acoust. Soc. Am. **114**, 2049–2057.

Brown, C. J., Abbas,P. J., Borland, J., and Bertschy, M. R. (**1996**). "Electrically evoked whole nerve action potentials in Ineraid cochlear implant users: Responses to different stimulating electrode configurations and comparison to psychophysical responses," J. Speech Hear. Res. **39**, 453–467.

Brown, C. J., Abbas, P. J., Fryauf-Bertschy, H., Kelsay, D., and Gantz, B. J. (**1994**). "Intraoperative and postoperative electrically evoked auditory brain stem responses in nucleus cochlear implant users: Implications for the fitting process," Ear Hear. **15**, 168–176.

Buechner, A., Brendel, M., Krüeger, B., Frohne-Büchner, C., Nogueira, W., Edler, B., and Lenarz, T. (**2008**). "Current steering and results from novel speech coding strategies," Otol. Neurotol. **29**, 203–207.

Buechner, A., Frohne-Büchner, C., Stoever, T., Gaertner, L., Battmer, R.-D., and Lenarz, T. (**2005**). "Comparison of a paired or sequential stimulation paradigm with Advanced Bionics' high-resolution mode," Otol. Neurotol. **26**, 941–947.

Cohen, L. T., Saunders, E., and Richardson, L. M. (**2004**). "Spatial spread of neural excitation: Comparison of compound action potential and forward-masking data in cochlear implant recipients," Int. J. Audiol. **43**, 346–355.

de Balthasar, C., Boëx, C., Cosendai, G., Valentini, A., Sigrist, A., and Pelizzone, M. (**2003**). "Channel interactions with high-rate biphasic electrical stimulation in cochlear implant subjects," Hear. Res. **182**, 77–87.

Donaldson, G. S., Kreft, H. A., and Litvak, L. (**2005**). "Place-pitch discrimination of single-versus dual-electrode stimuli by cochlear implant users (L)," J. Acoust. Soc. Am. **118**, 623–626.

Eisen M. D., and Franck, K. H. (**2005**). "Electrode interaction in pediatric cochlear implant subjects," J. Assoc. Res. Otolaryngol. **6**, 160–170.

Favre, E., and Pelizzone, M. (**1993**). "Channel interactions in patients using the Ineraid multichannel cochlear implant," Hear. Res. **66**, 150–156.

Finley, C. C., Wilson, B. S., and White, M. W. (**1990**). in *Cochlear Implants: Models of the Electrically Stimulated Ear*, edited by J. M. Miller and F. A. Spelman (Springer-Verlag, New York), pp. 55–93.

Firszt, J., Burton Koch, D., Downing, M., and Litvak, L. (**2007**). "Current steering creates additional pitch percepts in adult cochlear implant recipients," Otol. Neurotol. **28**, 629–636.

Frijns, J. H. M., Briaire, J. J., de Laat, J. A. P. M., and Grote, J. J. (**2002**). "Initial evaluation of the Clarion CII cochlear implant: Speech perception and neural response imaging," Ear Hear. **23**, 184–197.

Frijns, J. H. M., de Snoo, S. L., and Schoonhoven, R. (**1995**). "Potential distributions and neural excitation patterns in a rotationally symmetric model of the electrically stimulated cochlea," Hear. Res. **87**, 170–186.

Frijns, J. H. M., de Snoo, S. L., and ten Kate, J. H., (**1996**). "Spatial selectivity in a rotationally symmetric model of the electrically stimulated cochlea," Hear. Res. **95**, 33–48.

Gardi, J. N. (**1985**) in *Cochlear Implants*, edited by R. A. Schindler and M. M. Merzenich (Raven, New York), pp. 351–363.

Hughes M. L., and Stille, L. J. (**2008**). "Psychophysical versus physiological spatial forward masking and the relation to speech perception in cochlear implants," Ear Hear. **29**, 435–452.

Kral, A., Hartmann, R., Mortazavi, D., and Klinke, R. (**1998**). "Spatial resolution of cochlear implants: The electrical field and excitation of auditory afferents," Hear. Res. **121**, 11–28.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Macherey, O., Carlyon, R. P., Van Wieringen, A., Deeks, J. M., and Wouters, J., (**2008**). "Higher sensitivity of human auditory nerve fibers to positive electrical currents," J. Assoc. Res. Otolaryngol. **9**, 241–251.

Middlebrooks, J. C. (**2004**). "Effects of cochlear-implant pulse rate and inter-channel timing on channel interactions and thresholds," J. Acoust. Soc. Am. **116**, 452–568.

Ruddy H. A., and Loeb, G. E. (**1995**). "Influence of materials and geometry on fields produced by cochlear electrode arrays," Med. Biol. Eng. Comput. **33**, 793–801.

Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," Hear. Res. **12**, 1–16.

Shannon, R. V. (**1985**). in *Cochlear Implants*, edited by R. A. Schindler and M. M. Merzenich (Raven, New York), pp. 323–334.

Stickney, G. S., Loizou, P. C., Mishra, L. N., Assmann, P. F., Shannon, R. V., and Opie, J. M. (**2006**). "Effects of electrode design and configuration on channel interactions," Hear. Res. **211**, 33–45.

Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**). "Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–115.

White, M. W., Merzenich, M. M., and Gardi, J. N. (**1984**). "Multichannel cochlear implants: Channel interactions and processor design," Arch. Otolaryngol. **110**, 493–501.

Wilson, B. S., Wolford, R., Schatzer, R., Sun, X., and Lawson, D. (**2003**). "Speech processors for auditory prostheses," Seventh Quarterly Progress Report, NIH Contract No. N01-DC-2-1002, Neural Prosthesis Program, National Institutes of Health, Bethesda, MD, October–December.

Spectral and temporal integration of brief tones

Evelyn M. Hoglund and Lawrence L. Feth
Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210

(Received 22 February 2008; revised 17 September 2008; accepted 19 September 2008)

Both spectral and temporal integration of tones have been explored in detail, but integration of tones varying across both dimensions has received little attention. This study explores temporal integration of tone pulses that vary over a range of frequencies. Baseline thresholds were obtained for both spectral and temporal integration with the same signals and compared with prior research. The signals were then varied on both dimensions in several ways: with equivalent spectral and temporal step sizes, different spectral and temporal step sizes, and a random pattern of frequency presentation. The data were also analyzed by spectral step size, temporal step size, frequency range, direction and slope of frequency change, and predictability. The spectral and temporal integration conditions showed that the current procedures and signals yielded the same improvement in detection thresholds as prior studies. The spectrotemporal integration conditions showed the improvement for overall detection of the signals to be limited by spectral integration, with improvement related primarily to the number of tones, regardless of timing and frequency. Surprisingly, trial-by-trial random presentation of signal frequencies did not negatively influence detection. These results support the multiple looks hypothesis [Viemeister, N. F. and Wakefield, G. H. (1991). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865] as applied to spectrotemporal integration. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.2999338]

PACS number(s): 43.66.Ba, 43.66.Cb [RLF] Pages: 261–269

I. INTRODUCTION

The attempt to understand the nature of human auditory perception has provided a wide variety of directions for study. Important aspects of sound that have been of great interest include the temporal processing of sounds, and processing within the frequency, or spectral, domain. For both of these areas of auditory perception, there is the dilemma related to specificity versus generality. In other words, how do we achieve both fine acuity (resolution), of the "small picture," and also wide integration (summation) of the "big picture?"

The dilemma occurs because fine temporal acuity requires very brief time windows, and good frequency selectivity demands narrow auditory filters. Integration over a wide frequency range, or long signal duration, on the other hand, dictates that the processing window should be extended. Most of the research has been conducted on either temporal or spectral processing while keeping the other dimension constant. However, by holding one dimension constant, the results obtained may not reflect perception outside the laboratory since time and frequency are not independent of one another.

Early work on temporal or spectral summation of tones applied the energy detector model (Green, 1958). If the auditory system is simply summing up signal energy across N tones, the detectability of two or more tones set to be equally detectable with $d_i' = (2E/N_o)^{1/2}$ (where energy in the signal tone is E, and N_o is the spectrum level of the background noise) should improve with increasing number of tones according to $d_N' = [\Sigma_i^N (d_i')^2]^{1/2}$. For a fixed percentage of correct responses in a forced-choice experiment, this amounts to a detection threshold improvement of $-10\log(N)$, or 3 dB of improvement for each doubling of N. Human listeners do not achieve this "ideal" level performance in either temporal or spectral summation tasks, but improvements characterized by $-k\log(N)$, with $k < 10$, are common (van den Brink and Houtgast, 1990a, 1990b; Hicks and Buus, 2000).

A. Temporal processing

In their work on temporal integration, Viemeister and Wakefield (1991) introduced a "multiple looks" model for temporal integration that offers a good explanation for both temporal summation and resolution. Prior to this model, integration was studied in terms of either a long integration "window" (hundreds of milliseconds), which provided a good explanation for the summation, or a short integration window (3–5 ms), which provided a good explanation for the resolution. Unfortunately, neither type of model could offer a satisfactory explanation for the opposite extreme: long windows could only explain resolution via "leaky" integration (allowing some integration while summing over the window), and short windows offered poor prediction of summation.

The multiple looks model proposes that the auditory system actually uses short time constant windows or "looks" at the acoustic input. Thus, one look could detect a short duration signal, consistent with thresholds measured on resolution tasks. As the duration of the signal increases, the auditory system uses an increasing number of these looks consecutively, and the information from these windows is accumulated for detection, accounting for improved thresholds for longer duration signals. By measuring detection thresholds for one versus two pulse signals, Viemeister and Wakefield (1991) were able to show that listeners are able to utilize "intelligent" sampling to detect sounds. In quiet, detection thresholds for 200 μs pulse pairs separated by 1 ms

exhibited a 4 dB improvement relative to single pulses. The threshold increased until the separation was 5 ms, with no further change for longer separations. This suggests integration in a single window at 1 ms, partial integration up to 5 ms, then independent processing with longer separations. They then introduced a noise into the middle 50 ms of a 100 ms separation. When this intervening noise was introduced, the detection thresholds for two pulses averaged 2.5 dB lower than for single pulses, regardless of the level of the noise. The consistent improvement in threshold supports the use of intelligent processing of sounds, rather than summation of the entire duration of the signal, as would be assumed for a long integration window.

Further research by Buus (1999) considered the weighting of the pulses based on their temporal location in the signal, which revealed that detection of the pulses is approximately independent of the other pulses in the train, regardless of the noise masker used. Additionally, Buus (1999) found that the improvement in detection grows less quickly for pulse trains masked by single-band (50 Hz wide) and incoherent maskers (masking band with six flanking bands with unrelated envelopes) than for coherent maskers (masking band with six flanking bands with identical envelopes). His results support the predictions of the multiple looks theory for the single-band and incoherent masker conditions. The greater improvement in detection observed with the coherent masker condition, however, is inconsistent with the theory.

In their work with multiple looks in informational maskers, Kidd *et al.* (2003) suggested the influence of auditory streaming on thresholds for increasing number of tones. They argued that the consistency in the signal allowed streaming to facilitate detection.

B. Spectral processing

Early work involving spectral processing (Fletcher, 1940; French and Steinberg 1947; Schafer and Gales, 1949; Zwicker *et al.*, 1957; Greenwood, 1961; among others) established the critical band as the limit or look for spectral integration. In 1979, Spiegel (1979) studied the critical band and spectral integration to determine both the maximum limit of integration and the critical bandwidth in the same listeners. He found that thresholds for his noise signal increased at a rate consistent with the energy detector model, which predicts that for signals beyond the critical band, several bands will be combined to process the signal. The threshold increases, rather than decreases, as a result of increased noise through the bands along with the signal, at a rate proportional to $10\log(N)$. Based on his use of a masker noise with a bandwidth of 100–3000 Hz, Spiegel (1979) concluded that spectral integration could potentially occur through the entire range of audibility. Additionally, for detection of single tones, Green (1961) found that when the frequency was uncertain, detectability only decreased by 3 dB in the most extreme cases (frequency varying between 500 and 4000 Hz). Thus, detection was only minimally affected by the inability to predict at what frequency the tone would occur.

van den Brink and Houtgast (1990a, 1990b) reported a series of studies in which they measured spectral integration and temporal integration for long signals (100 ms) and brief signals (4.7 cycles). They found that brief tones were integrated more efficiently, at a rate of $8\log(N)$, than the longer tones, which were integrated at a rate of $5\log(N)$. These improvements contrast with the $10\log(N)$ improvement predicted by the energy detector model. Increased efficiency was also reported for temporal integration when the bandwidth of the signals was less than one critical band. Using the same signals, Hicks and Buus (2000) measured psychometric functions for long versus brief tones and tone complexes. They found steeper psychometric functions for brief tones, with integration functions corresponding to $8\log(N)$ for brief tones and $5\log(N)$ for long tones.

Work by Grose and Hall (1997) found that thresholds for tone complexes improved with increasing number of components, with thresholds improving at a rate of $10\log(N)$, consistent with predictions based on the multiple looks model. They found a parallel improvement [still $10\log(N)$], for detection of increments in narrow band noise, but with thresholds 2 dB higher than the pure tones. They then found less than $10\log(N)$ improvement for detecting a decrement in noise. In a final experiment, they carried the principle of decremented signals to its extreme, performing a gap detection experiment with the same narrow band noise complexes, which revealed improvement in detection greater than the predicted $10\log(N)$. A notable difference between these two decrement experiments relates to possible temporal integration differences; the decrement study used a 200 ms signal, but the gap detection study involved detection of gaps between 20 and 100 ms. Bacon *et al.* (2002) found consistent improvement in threshold for three-tone complexes compared with individual tones in an unmodulated noise masker, regardless of spectral distance between components. In modulated noise conditions, however, the amount of integration appeared to decrease with increased spacing between the components, possibly exhibiting a limit in spectral integration. Additionally, they combined pure tones at equally detectable levels for their threshold measures, an improvement over previous research that used equal physical levels.

C. Spectrotemporal processing

Since there is evidence to support the idea of multiple looks in both the time and the frequency dimensions of sound integration, the question remains about how these two processes may relate to each other. There has been little work published relating to both dimensions simultaneously. Spiegel (1979) made reference to the flexibility of the auditory system for both resolving short duration signals and summing longer duration signals. His data supported the idea that the spectral processing of sounds must be similarly flexible. In their study of spectral integration, Grose and Hall (1997) noted that the differences between their experiments on level decrement detection and gap detection may have been due in part to the differences in the temporal integration of the signals. However, no quantitative analysis of this suggestion is available.

Dai and Green (1993) compared thresholds for 3- and 21-tone complexes with signal durations ranging from 10 to 1000 ms. They reported that the threshold for the 3-component complex with closer frequency spacing was 10 dB lower than the 21-component complex at 10 ms, but the 21-component complex showed more rapid improvement in detectability with increasing duration than the 3-tone complex. They then measured the thresholds for a 3-component complex with the same frequency spacing as the 21-component complex. The results indicated that the improvement in thresholds was related to the frequency spacing of the tones rather than the total number of components. They invoked the multiple looks model for explanation of the results for the different duration conditions. However, this model would only predict the improved thresholds observed in the data for durations under 100 ms. Beyond this duration, no evidence of further integration can be seen. As a modification to the multiple looks model, Dai and Green (1993) suggested a "single look" model, which assumes that the filters are initially wide, and become increasingly narrow, with maximum tuning reached by 100 ms. This duration is the point at which their three-tone complex with the narrowest frequency spacing reached the steady portion of the threshold curve.

The purpose of the current study was to measure quiet thresholds for tone complexes that differed in both frequency and duration, in order to quantify the integration across both dimensions, and to examine the potential expansion of the multiple looks hypothesis into spectrotemporal integration. If the integration of increasing numbers of tones differing in both temporal and spectral location is similar to that of each domain individually, support may be evident for the multiple looks model with spectrotemporally varying signals. This study was designed to systematically measure quiet thresholds for tone complexes that vary either in spectral or temporal composition, or along both dimensions, in order to consider the potential application of multiple looks to the integration for both dimensions in the same subjects. The results should provide information about the relative salience of perception in the two dimensions and how discrete changes in both dimensions of the signals are integrated.

II. METHODS AND RESULTS

A. Subjects

Six normal hearing young adult listeners were included. Normal hearing was defined as air conduction thresholds ≤20 dB hearing level (HL) at the standard audiometric frequencies (250–8000 Hz) with normal otoscopic findings. Five female subjects and one male subject were included, ranging in age from 20 to 37 years old. No prior experience with psychoacoustics research was required for participation; however, introductory training was provided to assure familiarity with the procedure and signals and to assure stable quiet threshold measurements. The same subjects participated in all experiments and completed all experiments in four to five 2 h sessions per week for four to five weeks.

B. Procedures

All experiments were conducted in a sound attenuated room. Signals were generated digitally with MATLAB (version 2006b), processed through Digital Audio Labs CardDeluxe sound cards, attenuated with a TDT PA4 programmable attenuator, and presented monaurally over Sennheiser HD580 headphones. All subjects used the right ear, with the exception of subject No. 6, who used the left ear due to a slight increase in pure tone threshold in the right ear at one frequency. The experiments were conducted using an adaptive tracking [one up three down two interval forced choice (2IFC)] procedure to target a threshold at the 79% level (Levitt, 1971). The step size for signal level was initially set at 5 dB and was reduced to 2 dB after the first reversal, then reduced again to 1 dB after the next reversal. The starting level for each run was at 10 dB of attenuation, with signal generation levels set to target a pure tone threshold of 35 dB of attenuation at each frequency. Each run consisted of 50 trials and included between 5 and 10 reversals, with the first 4 reversals discarded in the calculation of the threshold estimate. Thresholds were calculated using three separate runs for each condition, with each threshold estimate based on a total of 150 trials. Occasionally, subjects ran more than three runs for a condition, if one or more run yielded an inconsistent result for the threshold estimate. These inconsistencies were generally a result of errors early in the run, causing the step size to decrease too quickly and not allowing the run to stabilize, or variation in levels for the retained reversals, causing a wide variance around the eventual calculated threshold. Results from these runs were discarded.

C. Stimuli

Stimuli consisted of one to eight 10 ms sinusoidal tone bursts. Tone frequencies were 356, 494, 663, 870, 1125, 1442, 1838, and 2338 Hz. These frequencies were selected based on the work by Grose and Hall (1997), with tones centered in alternating equivalent rectangular bands (ERB_ns) (Moore and Glasberg, 1983). Each 10 ms burst was gated on and off with quarter sine wave 5 ms rise and decay times, with no steady state component (Viemeister and Wakefield, 1991). Tones were separated by 10 ms silent intervals in order to minimize interaction between tones. The temporal separation was chosen to separate them into alternating equivalent rectangular durations (ERDs). ERDs were reported by Moore *et al.* (1988) to range between 8 and 10.8 ms for brief tones, depending on the conditions. While the two dimensions cannot be equated in a quantitative measure, the spacing in both dimensions was made as equivalent as possible. Overall duration of signals and silent intervals were specified to ensure that all temporal integration be contained within a 150 ms maximum window. The nominal stimulus duration represented in the stimulus matrix (Fig. 1) was 160 ms for all the experimental conditions, but the final 10 ms interval was always silent. A 300 ms interstimulus interval was used. All signals were presented in quiet. Figure 1 represents the matrix of time and frequency cells used in the experimental signals. Rows in the figure represent the temporal domain, while columns represent the spectral domain.

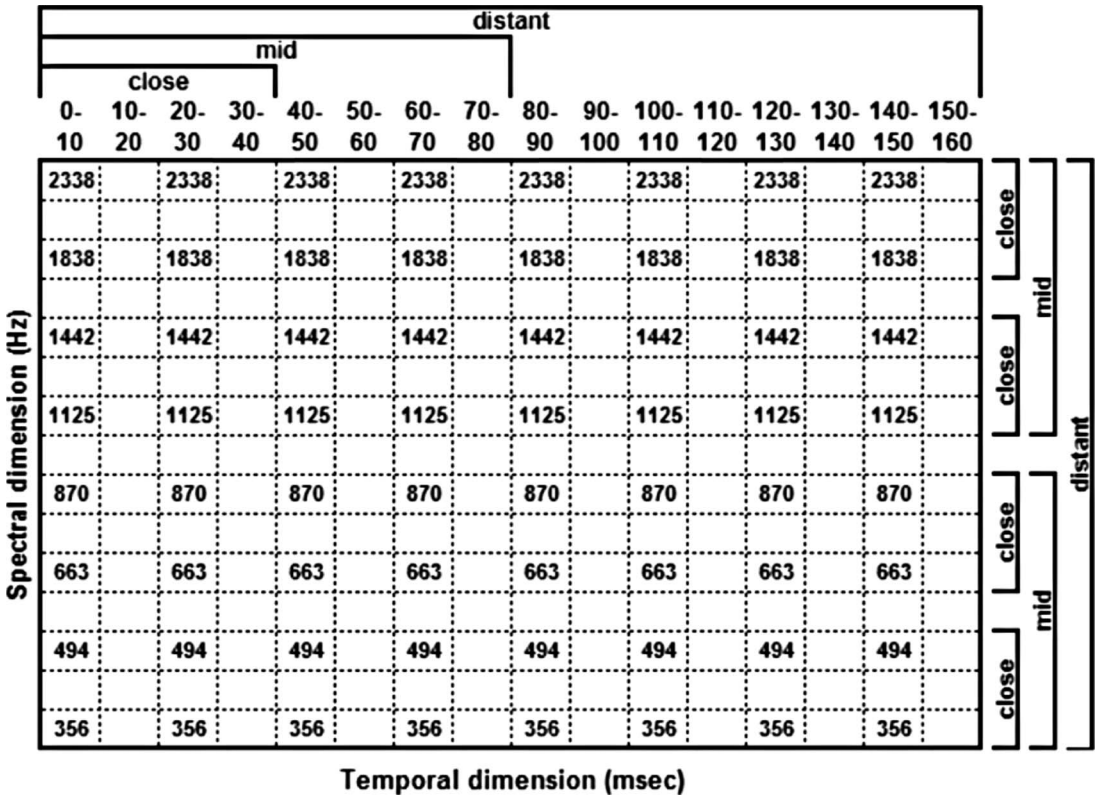


FIG. 1. TFR of the stimuli used in the study. The time axis is divided into 10 ms windows. The frequency axis is divided into one ERB_n filter band. Signals were separated in time and frequency so that empty squares indicate combinations for which signals were not presented. Brackets along the top and right side reflect the spacing for two-tone signals. Four-tone signals were created using consecutive or alternating tones and time intervals for close or mid conditions. Spectrotemporal signals are represented along diagonal lines.

Thus the time frequency representation (TFR) for each spectrotemporal condition, as represented by a line linking the relevant cells across the matrix, would have a slope of ± 1, $< \pm 1$, or $> \pm 1$. The diagonal to be used was determined by the conditions within the context of each particular experiment.

D. Baseline

1. Procedure

A baseline experiment was conducted to determine the appropriate presentation levels for the individual sine wave tones for each subject. Initial thresholds were obtained, then the amplitude of each tone was adjusted to equalize the detectability of the tones across the frequencies. This level adjustment ensured that the detection of the overall signal complex in the remaining experiments was not based solely on the signal(s) with the lowest individual threshold(s). Thresholds obtained for each frequency at equal presentation levels ranged from 27.1 to 50.0 dB of attenuation, with the lowest thresholds measured for the higher frequencies, as would be expected based on the shape of the normal audibility curve. After these thresholds were obtained, the presentation levels were adjusted until all the measured thresholds were within a range of less than 2.5 dB, and the overall mean was between 34.75 and 35.25 dB of attenuation. With the range of initial thresholds, it was expected that without adjustment, the subjects would have likely been able to detect signals on the basis of the higher frequency components.

2. Results

As seen in Fig. 2, the mean threshold for all frequencies ranged from 34.56 to 35.64 dB, with an overall mean of 35.14 dB. A repeated measures analysis of variance (ANOVA) indicated no significant difference between frequency thresholds, $F(1,7)=0.891$, $p>0.05$. The thresholds for individual frequencies were averaged across subjects, and

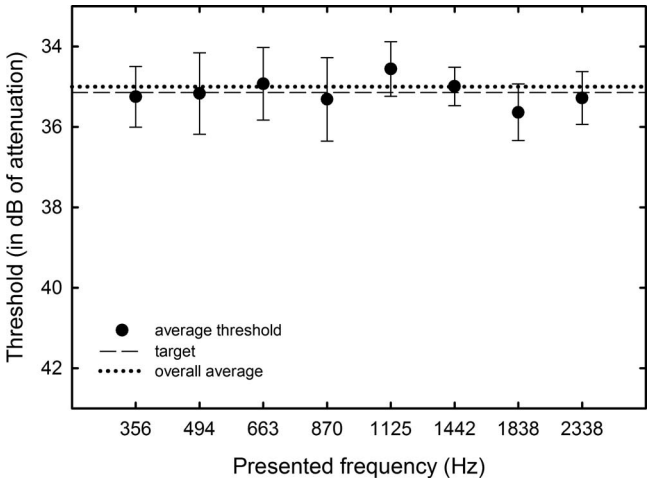


FIG. 2. Thresholds for baseline frequencies averaged across subjects. The abscissa is the frequency for which the threshold was obtained, and the ordinate is the decibel of attenuation for that threshold. The dotted line is set at 35 dB of attenuation (the target threshold), and the dashed line indicates the obtained average for all eight frequencies, 35.14. Error bars represent one standard deviation around the mean.

then averaged across frequency. Variance for these averaged thresholds was 1.08 dB. Just noticeable differences (JNDs) in intensity for signals with a duration of 15 ms are between 3 and 4 dB (Oxenham and Buus, 2000) when presented at a level of 65 dB sound pressure level (SPL). It was assumed in the current study that the JNDs at absolute threshold would not be substantially different than those at higher presentation levels. For this reason, the variance found within individual subjects' thresholds was believed to be acceptable for the assumption of equal detectability. To check the accuracy of this assumption, the calculated improvement in thresholds based on the obtained data was compared with the improvement predicted by the energy detection model. The model predicts an improvement of 9.03 dB [i.e., 10 log(8)] over all eight of the frequencies in an ideal listener. When the average thresholds were submitted to the same calculation, the improvement in threshold was 8.9 dB, with the most disparate overall anticipated threshold being 8.76 dB of improvement (for subject No. 4).

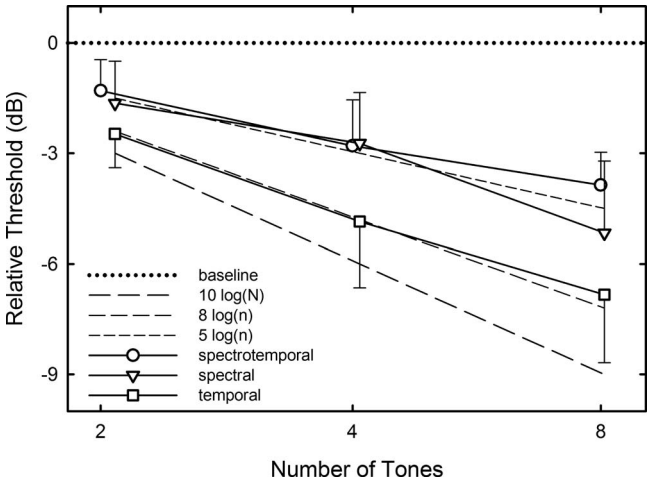


FIG. 3. Integration of multicomponent tones for experiments 1 and 2. Open circles are the thresholds for spectrotemporal signals with the same step sizes. The open triangles are the thresholds for the spectral dimension alone, and the open squares are the thresholds for temporal integration alone. Data points represent two, four, and eight tone signals collapsed across all other conditions. The dotted line is the average threshold for single tones. The long dashed line illustrates 10 log(N), the medium dashed line represents 8 log(N), and the short dashed line represents 5 log(N).

E. Experiment 1 (single dimension)

1. Procedure

The first experiment was conducted to establish a reference for spectral and temporal integration within the context of the frequencies and time window used in this study. Using the adjusted presentation levels for the individual frequencies obtained in the baseline experiment for each listener, thresholds were determined for spectral integration for combinations of two, four, and eight tones. The configuration of these signals can be seen in the TFR shown in Fig. 1. The terms "high," "mid," and "low" were used to describe the location within the frequency range used for the two-tone signal conditions. For example, the high condition included the frequencies at the upper end of the spectrum used in the study (e.g., 1838 and 2338 Hz). Additionally, the terms "close," mid, and "distant" were used to describe the step size between components. Two-tone combinations were presented at 1838 and 2338 Hz for the close-high condition, at 870 and 1125 Hz for the close-mid condition, and at 356 and 494 Hz for the close-low condition. These conditions use tones next to each other within the experimental matrix and were separated by one intervening ERB_n. The tones used for the mid step size condition were 663 and 1442 Hz and were separated by five ERB_ns. The distant condition consisted of 356 and 2338 Hz, the most disparate frequencies in the matrix, and separated by 13 ERB_ns. The four-tone combinations consisted of 663, 870, 1125, and 1442 Hz for the close condition, again including adjacent tones and having only one intervening ERB_n between components, and 356, 663, 1125, and 1838 Hz for the mid condition, with three ERB_ns between components. The eight-tone combination included all the identified frequencies. All temporal and spectral integration conditions were presented randomly in order to eliminate order effects.

Temporal integration was measured for two, four, and eight signals, again using the adjusted presentation levels obtained in the baseline experiment. The signals were presented at 356, 1125, or 2338 Hz. These frequencies were selected to represent the high, middle, and low frequency ranges in the study. Again, the configuration of the signals can be seen in Fig. 1. Two-tone signals were presented with an intervening silent interval of 10 ms for the close condition, 50 ms for the mid condition, and 130 ms for the distant condition. Four-tone signals were presented with a 10 ms intertone interval in the close condition and a 30 ms intertone interval for the mid condition. The temporal integration for eight tones was measured with 10 ms intervals between signals.

2. Results

Spectral integration. The spectral integration measures show an improvement in threshold between approximately 1 and 2.5 dB for each doubling of the number of tones. The overall improvement in threshold from one to eight tones was 5.17 dB. In Fig. 3, this can be seen with the open triangles representing the average of the thresholds of all conditions plotted against the number of tones. The data were collapsed across all conditions since all comparisons based on frequency range and spectral spacing were found to be statistically nonsignificant. Linear regression revealed that the best fit to this function was the 5 log(N) predicted curve (R^2=0.688, slope=1.1). A repeated measures ANOVA indicated a significant difference for increasing numbers of tones, $F(1,3)=187.40$, $p<0.05$. Differences seen in thresholds on the basis of frequency range and spectral spacing were not significant.

Temporal integration. The temporal integration measures show an improvement in threshold between approximately 1.98 and 2.47 dB for each doubling of the number of tones. The overall improvement in threshold from one to eight tones was 6.84 dB. In Fig. 3, this can be seen with the open squares representing the average of the thresholds of all conditions plotted against the number of tones. Linear re-

gression revealed that the best fit to this function was the 8 log(N) curve for these data (R^2=0.805, slope=0.95). Repeated measures ANOVA revealed that increasing number of tones was significant, $F(1,3)=287.42$, $p<0.05$.

The temporal integration conditions were run at three frequencies to sample potential differences on the basis of the spectral range. The change in detection threshold on the basis of the frequency was not significant, $F(2,4)=0.119$, $p>0.05$. This result confirms that the remaining conditions can readily be centered around any of the included frequencies without concern about the influence of any spectral differences confounding the perception of these signals.

F. Experiment 2 (equivalent spectral-temporal spacing)

1. Procedure

The second experiment measured thresholds for signals differing in both the temporal and spectral domains, with the signals configured in equally spaced "steps" in both the spectral dimension and the temporal dimension. The resulting signals followed the major and minor diagonals in the TFR. Signals included complexes of two, four, and eight sine wave tones varying in either an "up" or "down" frequency direction. For example, the two-tone stimuli were presented in conditions defined as close-close (adjacent frequencies and adjacent time intervals), mid-mid (steps that spanned approximately one-half the total possible range both in frequencies and temporal intervals), and distant-distant (endpoint values for frequency and time). The close-close conditions were presented in both the high frequency range, that is, with 1838 and 2338 Hz tones, and the low frequency range, with 356 and 494 Hz tones. The mid-mid conditions also used the high frequency range (1125 and 2338 Hz) and the low frequency range (356 and 870 Hz). The close conditions presented tones in the 1st and 3rd time windows, with starting times of 0 and 20 ms, the mid conditions presented the tones in the 1st and 4th time windows, with starting times of 0 and 60 ms, and the distant conditions presented the tones in the 1st and 15th time windows, with starting times of 0 and 140 ms. In similar fashion, four-tone stimuli were presented with close conditions in consecutive frequency and time windows and presented in the high and low frequency ranges. The high frequency conditions included 1125, 1442, 1838, and 2338 Hz tones presented in time windows with starting times of 0, 20, 40, and 60 ms. The four-tone mid signals included alternate frequencies starting with an endpoint. For example, the up condition included the 356, 663, 1125, and 1838 Hz tones presented in the time windows with starting times of 0, 40, 80, and 120 ms. The eight-tone stimuli were all presented in close intervals since this was the only possibility within the spectral and temporal constraints of the experiment. A total of 10 two-tone conditions, 6 four-tone conditions, and 2 eight-tone conditions was used.

2. Results

For the spectrotemporal signals, the overall improvement in threshold from one to eight tones was 3.86 dB. Linear regression revealed that the best fit to this function was the 5 log(N) curve (R^2=0.765, slope=0.872). Repeated measures ANOVA revealed that the thresholds were significantly different than for temporal integration alone, $F(1,5)=43.184$, $p<0.05$. The differences between spectrotemporal and spectral integration were not significant, $F(1,5)=1.667$, $p>0.05$. The open circles in Fig. 3 show the overall thresholds for these equal step size signals plotted by the number of tones.

The data were also considered on the basis of spectral and temporal spacing. Because of the design of this experiment, the step size in both dimensions was equivalent in each condition. The improvement from one to two tones was nearly identical in all spacing conditions. Across all comparisons, the differences in thresholds were not statistically significant on the basis of spectral and temporal spacing or direction of frequency change.

G. Experiment 3 (different spectral-temporal spacing)

1. Procedure

The third experiment measured the thresholds for signals with different step sizes in the spectral and temporal dimensions. For example, the steps in the spectral domain may be distant, while the temporal steps may be mid or close. The alternative difference was also presented, with steps in the spectral domain being smaller than those in the temporal domain. In this way, the perceptual slope for the signals may be either steeper than 1 (as in the first example) or shallower than 1 (as in the second example), depending on the specific condition. These conditions were included in order to allow for the possibility of determining the relative importance of the two dimensions (spectral and temporal) for threshold. If one of these dimensions has a greater influence on the overall integration than the other, then the thresholds in this configuration should show an influence of slope. Because the signals required a difference in step size between the two dimensions, no eight-tone signals could be used, as the constraints of the experimental design required that these signals were always at a close step size for both dimensions. As for equivalent spacing, the signals were presented in both upward and downward slopes. A total of eight two-tone conditions and four four-tone conditions was used. These conditions were centered around the middle frequencies in the matrix.

2. Results

The magnitude of integration for these signals considered all together is less than the comparable conditions with equivalent spacing. The threshold for signals with different step sizes showed an overall improvement from one to four tones of 1.74 dB. This is different from the improvement in threshold for the same step size signals, which showed an improvement of 2.80 dB for the same change from one to four tones. Linear regression for these data showed a poor fit to any of the predicted threshold improvements, with the best fit to the 5 log(N) curve (R^2=0.575, slope=0.577). Regression would suggest a possible fit to a slope somewhat less than 5 log(N). These thresholds are represented in Fig. 4 with open squares and compared with the same step size thresh-

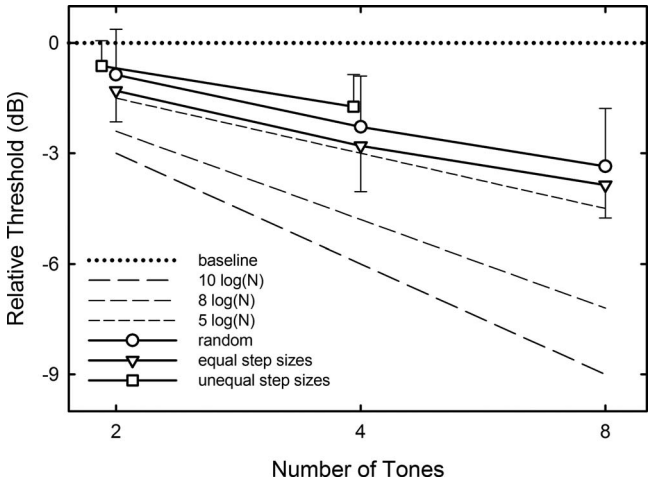


FIG. 4. Spectrotemporal integration across experiments 2–4. Open triangles are the signals with the same step sizes, open squares are signals with different step sizes, and open circles are random signals. Data points represent two, four, and eight tone signals collapsed across all other conditions. The dotted line is the average threshold for integration of single tones and the long dashed line is $10\log(N)$, the medium dashed line represents $8\log(N)$, and the short dashed line represents $5\log(N)$.

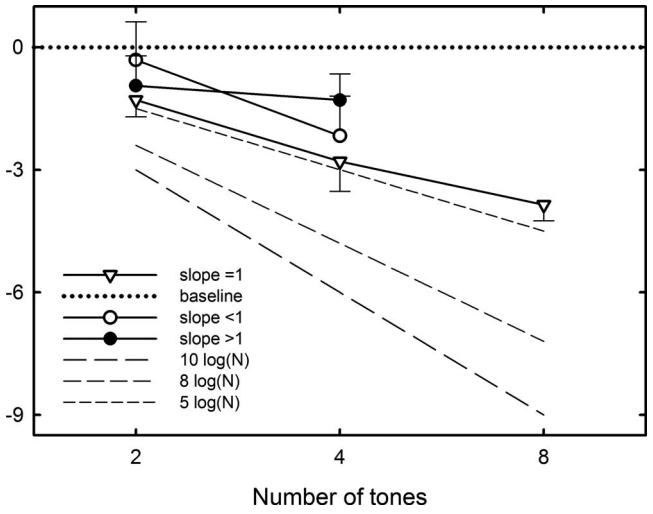


FIG. 5. Integration of signals with different slopes. Filled circles are the thresholds for signals with greater spectral distance (slope >1), open circles are thresholds for signals with greater temporal distance (slope <1), and open triangles are threshold for signals with equivalent step sizes (slope $=1$). The dotted line is the average threshold for integration of single tones and the long dashed line illustrates $10\log(N)$, the medium dashed line represents $8\log(N)$, and the short dashed line represents $5\log(N)$.

olds (open triangles). Recall that there were no conditions that included eight tones since the constraints of the TFR would not allow a difference in the step size for the temporal and spectral domains, while still remaining within the matrix.

As in experiment 2, the data were analyzed for differences in spectral and temporal spacing and direction, and these were not significant. Additionally, the data were compared with those of equivalent spacing on the basis of the slope of the TFR of the signals. The data for signals with different spacings were all represented by slopes of greater than or less than 1, while all the signals with equivalent spacing from experiment 2 had a slope equal to 1. The thresholds for signals with a slope of <1 are plotted with open circles and those with a slope of >1 are filled circles in Fig. 5. While the thresholds for signals with different slopes appeared to be different, the difference was not statistically significant.

H. Experiment 4 (random presentation order)

1. Procedure

The fourth experiment was conducted to measure the thresholds for the complex signals when the component tones were randomly selected. The signals were presented in random order without replacement for each trial. Every tone complex presented was different. The temporal and spectral step size constraints were limited to equal step sizes between the dimensions and defined as in experiment 2. Thus, the two-tone close condition consisted of two neighboring frequencies presented in consecutive temporal windows, but the two tones were randomly selected from among the eight possible. Similarly, the two-tone mid condition consisted of two frequencies spaced approximately half the distance of the frequency range, and in the first and middle temporal windows, with the specific frequencies randomly selected. The conditions used in this experiment included three two-tone conditions, two four-tone conditions, and one eight-tone condition. All different spectral and temporal distances (close, mid, and distant) were represented in the conditions.

2. Results

The overall improvement in threshold for signals with random frequencies showed a very close match to the thresholds for the signals with equivalent spacing. This can be seen in Fig. 4 by comparing the open circles with the open triangles. Thus, there appears to be little difference in the spectrotemporal integration for signals when there is no ability to predict the frequency range in which the tones will occur. Linear regression showed the best fit to the $5\log(N)$ predicted curve ($R^2=0.568$, slope$=0.765$). Repeated measures ANOVA revealed that the thresholds for signals with random frequency presentation did not differ significantly from those of the signals with equivalent spacing, $F(1,5)=1.667$, $p>0.05$, nor from those with different spacing, $F(1,5)=2.486$, $p>0.05$.

When the random frequency signal conditions were analyzed on the basis of the spectral and temporal spacings, the only significant difference appeared between the two-tone conditions with close and mid spacings. All other comparisons were nonsignificant.

I. General results

A significant difference can be seen between integration for the spectral domain and the temporal domain individually. The slope of spectral integration closely approximated $5\log(N)$, while the slope of temporal integration was closer to $8\log(N)$. Spectrotemporal integration was not significantly different than the spectral dimension alone. The signals throughout the study were analyzed on the basis of spectral distance between tones, temporal distance between tones, direction of slope representing the signal change, and steep-

ness of that slope. The signals were also randomized for the equal spectral and temporal step size conditions. None of these analyses were statistically significant.

III. DISCUSSION

A. Spectral integration

The spectral integration conditions of the first experiment were conducted to replicate the work by Grose and Hall (1997), with minor adjustments to the method to allow for comparison with the remainder of the experiments within this study. The current study showed improvements of 1.02–2.44 dB per doubling of tones, comparable to their improvement of 1.5–2 dB per doubling of tones. These improvements were a good match for their results, despite different durations, as well as the adjustment to the signal levels to make them equally detectable. The wider variation in the threshold improvements may be related to the brief duration of the current tones, as little data are available related to the stability of threshold measures for these brief tones. These threshold improvements are inconsistent with the results of van den Brink and Houtgast (1990a, 1990b) for brief tones, which showed thresholds for brief duration tones to improve more rapidly than longer duration tones. While Grose and Hall (1997) used only signals centered around the 1125 Hz signal, this study also included two-tone signals in the high and low frequencies in order to consider the potential for differential integration on the basis of frequency range. The differences found in these data were not significant.

To consider integration across the spectrum, the concept of multiple looks could be considered in the frequency domain. The auditory system may be combining the signals across independent filters in a spectral multiple look in much the same way that Viemeister and Wakefield (1991) proposed for the temporal domain and selectively responding to the bands that include energy. In the current study, the distance between component tones did not affect spectral integration, which was consistent with the unmodulated noise conditions reported by Bacon *et al.* (2002). The multiple looks hypothesis would suggest that the auditory system can "intelligently" select the time intervals containing information for integration. It appears with the current results that this may be the case in the spectral domain.

B. Temporal integration

The temporal integration conditions of the first experiment were also conducted to replicate earlier research. In their work, Viemeister and Wakefield (1991) demonstrated an improvement in threshold of 2.5 dB from one tone to two tones, while the current study showed an improvement of 2.47 dB for the same condition. The extension beyond their work yielded improvements of 2.39 dB from two to four tones and 1.98 dB from four to eight tones. No significant difference can be seen for different temporal spacings, from 10 ms of silence between tones to 130 ms of silence. All of these distances are greater than the 5 ms gap required for independence of the observations. Thus the integration appears to be the same throughout the overall time window used in all the signals, supporting the multiple looks hypothesis for temporal integration.

The similarity in the detection of the signals presented at the three frequencies (356, 1125, and 2338 Hz) additionally confirms that the temporal integration task does not vary with the spectral range being used. As a result, further study can reliably be pursued in any frequencies in the range included here, with good assurance that the results will be the same.

C. Spectrotemporal integration

The combination of the two domains yields additional information about the function of the auditory system that has previously not been explored. The results of this study indicate that the limits of spectrotemporal integration are due to the limits in the ability to integrate spectral information, with very similar detection thresholds between the two sets of integration conditions. The spectrotemporal signals did not show significant differences on the basis of spectral and temporal step sizes or direction of frequency change. This, again, supports the multiple looks hypothesis. The auditory system appears to be capable of monitoring the filters throughout the entire range of frequencies included here and detecting signals that occur in any of the time windows. The information contained in those tones is accumulated over the course of the longer, overall, time window for use in detection.

While the same step size signals were processed in a way that was very similar to those varying in only the spectral dimension, the processing of the different step size signals appeared to be dissimilar. The thresholds for these signals were consistently poorer than for the same step size signals. This came as a surprise after the data collection was initiated. Prior expectation was that there should be no difference between the two types of spectrotemporal signals. One would expect that this prediction would be further supported based on the nonsignificant differences related to step sizes in experiment 2. Separate analysis of slope showed that the thresholds for signals that changed over greater spectral than temporal range showed essentially no improvement with the increase from two tones to four. While these differences were not statistically significant, they suggest the possibility that the auditory system may be near the limit for tracking changes in signals at this rate, as these signals consisted of a total duration of 30 ms, with spectral changes from 7 to 15 ERB_ns. Further study with different signal durations and different spectral ranges may provide additional insight into this uncertainty.

D. Random presentation order

If a phenomenon such as auditory streaming is involved in the detection of these signals, then trial-by-trial random order of presentation should have a negative effect on the integration of the tones. Listeners should be unable to use streaming as a cue to detect the signal in the random presentation order conditions, which should result in higher thresholds for these signals. Alternatively, if selective monitoring

of critical bands is important for this detection, then again, the random order of presentation of frequencies should have a negative effect on integration. The results showed that the detection of the randomly generated signals was closer to that of the same step size signals than was detection of the different step size signals. This also supports the multiple looks hypothesis, in that clearly the listeners are not simply establishing which critical bands contain the component tones and then monitoring those in order to detect the total signals. Rather, they are monitoring all the critical bands that may be included and detecting the tones wherever they may be present. In fact, auditory streaming is not supported as a factor in detection for these signals, as the poorest overall detection in the random frequency signals was for the closely spaced tones. Even in the predictable conditions, the thresholds showed no difference on the basis of spectral spacing. However, since the differences in the thresholds are quite small overall, and the signals are brief, the analysis of the potential influence of streaming in spectrotemporal integration cannot be fully addressed here.

IV. SUMMARY AND CONCLUSIONS

In all experiments in this study, the one variable that resulted in improved thresholds was the number of tones. Regardless of spectral or temporal spacing, the thresholds showed improvement with every increase, in spectral integration alone, temporal integration alone, and spectrotemporal integration. The same improvement occurred when signals were presented with random frequency selection. Thus, detection appears to be based on the use of multiple looks for signals even when changing in both dimensions trial to trial. The auditory system appears to be able to monitor all the critical bands and select the windows that include the information in order to detect the signals.

All of the results presented here are consistent with the multiple looks hypothesis, with some limits in the spectral domain. It is unclear why spectral integration is poorer than temporal integration, but the analysis of the predicted threshold improvement calculated with baseline measures indicates that it is not due to differences in the detectability of each frequency. The results are also consistent with prior separate studies of these dimensions. Kidd *et al.* (2003) suggested an auditory stream coherence explanation for the results in their temporal integration study. This cannot be accepted directly on the basis of the current results since the random presentation order conditions were not significantly different than the predictable conditions. In fact, the thresholds for random frequency signals were more similar to those for equal step size signals than were the unequal step size thresholds. While it seems reasonable that auditory stream coherence may play a role in the detection of these signals, further study is required to determine what that role may be, particularly in light of the potential lower limit on temporal detection seen here, and similarity in thresholds for all temporal and spectral spacing.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institutes of Health, NIH/NIDCD R01-DC006879. This article is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at The Ohio State University, Columbus, OH. Thanks also to Walt Jesteadt and an anonymous reviewer for helpful comments on an earlier draft. Additional thanks to Eric Healy and Robert Fox for their input.

Bacon, S. P., Grimault, N., and Lee, J. (**2002**). "Spectral integration in bands of modulated or unmodulated noise," J. Acoust. Soc. Am. **12**, 219–226.
Buus, S. (**1999**). "Temporal integration and multiple looks, revisited: Weights as a function of time," J. Acoust. Soc. Am. **105**, 2466–2475.
Dai, H., and Green, D. M. (**1993**). "Discrimination of spectral shape as a function of stimulus duration," J. Acoust. Soc. Am. **93**, 957–965.
Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.
French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **9**, 90–119.
Green, D. M. (**1958**). "Detection of multiple component signals in noise," J. Acoust. Soc. Am. **30**, 904–911.
Green, D. M. (**1961**). "Detection of auditory sinusoids of uncertain frequency," J. Acoust. Soc. Am. **33**, 897–903.
Greenwood, D. D. (**1961**). "Auditory masking and the critical band," J. Acoust. Soc. Am. **33**, 484–502.
Grose, J. H., and Hall, J. W. (**1997**). "Multiband detection of energy fluctuations," J. Acoust. Soc. Am. **102**, 1088–1096.
Hicks, M. L., and Buus, S. (**2000**). "Efficient across-frequency integration: Evidence from psychometric functions," J. Acoust. Soc. Am. **107**, 3333–3342.
Kidd, G., Mason, C. R., and Richards, V. M. (**2003**). "Multiple bursts, multiple looks, and stream coherence in the release from informational masking," J. Acoust. Soc. Am. **114**, 2835–2845.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
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., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). "The shape of the ear's temporal window," J. Acoust. Soc. Am. **83**, 1102–1116.
Oxenham, A. J., and Buus, S. (**2000**). "Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions," J. Acoust. Soc. Am. **107**, 1605–1614.
Schafer, T. H., and Gales, R. S. (**1949**). "Auditory masking of multiple tones by random noise," J. Acoust. Soc. Am. **21**, 392–397.
Spiegel, M. F. (**1979**). "The range of spectral integration," J. Acoust. Soc. Am. **66**, 1356–1363.
van den Brink, W. A., and Houtgast, T. (**1990**). "Efficient across-frequency integration in short-signal detection," J. Acoust. Soc. Am. **87**, 284–291.
van den Brink, W. A., and Houtgast, T. (**1990**). "Spectro-temporal integration in signal detection," J. Acoust. Soc. Am. **88**, 1703–1711.
Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). "Critical bandwidth in loudness summation," J. Acoust. Soc. Am. **29**, 548–557.

Pitfalls in behavioral estimates of basilar-membrane compression in humans[a)]

Magdalena Wojtczak[b)] and Andrew J. Oxenham
Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

(Received 12 May 2008; revised 17 October 2008; accepted 20 October 2008)

Psychoacoustic estimates of basilar-membrane compression often compare on- and off-frequency forward masking. Such estimates involve assuming that the recovery from forward masking for a given signal frequency is independent of masker frequency. To test this assumption, thresholds for a brief 4-kHz signal were measured as a function of masker-signal delay. Comparisons were made between on-frequency (4 kHz) and off-frequency (either 2.4 or 4.4 kHz) maskers, adjusted in level to produce the same amount of masking at a 0-ms delay between masker offset and signal onset. Consistent with the assumption, forward-masking recovery from a moderate-level (83 dB SPL) 2.4-kHz masker and a high-level (92 dB SPL) 4.4-kHz masker was the same as from the equivalent on-frequency maskers. In contrast, recovery from a high-level (92 dB SPL) 2.4-kHz forward masker was slower than from the equivalent on-frequency masker. The results were used to simulate temporal masking curves, taking into account the differences in on- and off-frequency masking recoveries at high levels. The predictions suggest that compression estimates assuming frequency-independent masking recovery may overestimate compression by as much as a factor of 2. The results suggest caution in interpreting forward-masking data in terms of basilar-membrane compression, particularly when high-level maskers are involved.
 [DOI: 10.1121/1.3023063]

PACS number(s): 43.66.Dc, 43.66.Ba [BCM] Pages: 270–281

I. INTRODUCTION

Psychophysical methods for estimating gain and compression in the healthy human cochlea have received considerable attention over the past decade (e.g., Oxenham and Plack, 1997; Nelson *et al.*, 2001; Lopez-Poveda *et al.*, 2003; Plack *et al.*, 2004; Rosengard *et al.*, 2005). These methods are attractive because they provide a potential window into the characteristics of the basilar-membrane (BM) response in humans, which cannot be measured directly. A number of different psychophysical paradigms have been used. In one, termed growth of masking (GOM), the level of a forward masker required to just mask a brief tonal signal is measured as a function of signal level, with the gap between masker offset and signal onset held constant (Oxenham and Plack, 1997). In another paradigm, termed the temporal masking curve (TMC) method, the level of a forward masker required to just mask a brief tonal signal is measured as a function of the gap between the masker offset and the signal onset, with the level of the signal held at a low constant value (Nelson *et al.*, 2001; Lopez-Poveda *et al.*, 2003; Plack *et al.*, 2004; Rosengard *et al.*, 2005). In both paradigms, on-frequency masking functions, where the masker and signal are at the same frequency, are compared with off-frequency masking functions, where the masker frequency is well below that of the signal.

Because these measures are indirect, they rely on a number of assumptions. First, all the techniques assume that the signal is detected when it evokes excitation at or near a place along the BM with a characteristic frequency (CF) equal to the signal frequency (CF_s) that is sufficient to produce a certain criterion increase in the decision variable over the value representing the response to a masker alone. This assumption allows conclusions to be drawn about the response at the CF_s, based on the signal's masked threshold. The conditions necessary for this assumption to hold are maintained by presenting the signal in the presence of a spectrally shaped noise designed to limit the audible spread of the signal's excitation (e.g., Oxenham and Plack, 1997) or by limiting the signal to a very low level, so that its spread of excitation along the BM remains minimal (e.g., Nelson *et al.*, 2001). A second assumption, which is also necessary for both the GOM and TMC paradigms, is that the response of the BM at CF_s is linear when presented with tones well below the signal frequency. This assumption is based on physiological findings in other mammals, showing that the BM response to tones more than half an octave below the CF of the place of measurement is linear, at least in the base of the cochlea (e.g., Rhode and Robles, 1974; Murugasu and Russell, 1995; Ruggero *et al.*, 1997). A third assumption is that the effectiveness of a forward masker after a given delay from its offset is determined by cochlear filtering at the signal place, and that once filtering has occurred, all maskers with the same temporal envelope characteristics have an equivalent masking effect, regardless of their spectral com-

[a)]Portions of these data were presented at the 153rd Meeting of the Acoustical Society of America and at the 31st ARO Midwinter Meeting, Phoenix, Arizona.
[b)]Electronic mail: wojtc001@umn.edu

position. In other words, the masker excitation at CF_s determines the signal threshold, regardless of the masker spectrum.

This third assumption has no direct physiological support, although for simultaneous masking it forms the basis of the well-known power spectrum model of masking (e.g., Fletcher, 1940). From a physiological standpoint, given the complex interactions that occur between inhibitory and excitatory inputs as early as the cochlear nucleus (e.g., Wickesberg and Oertel, 1988; Ryugo and Parks, 2003), it seems at least possible that such an assumption may not hold. Nevertheless, early studies using these assumptions have derived estimates of BM input-output functions that are in good agreement with direct physiological measures in other mammals (e.g., Oxenham and Plack, 1997; Nelson *et al.*, 2001). Furthermore, the effect of cochlear hearing loss on these estimates (e.g., Oxenham and Plack, 1997; Plack *et al.*, 2004) seems to be in line with physiological measures in other species following cochlear damage or dysfunction (Ruggero *et al.*, 1997).

Incorporated within the third assumption is the prediction that the recovery from forward masking for a given signal frequency is independent of masker frequency, once masker effectiveness has been equated. For instance, if two forward maskers of the same duration but different frequencies (or with different spectral content) produce the same amount of masking at one masker-signal delay, then they should also produce the same amount of masking at any other delay. This assumption is necessary for the TMC method, in which on- and off-frequency masker levels at threshold are compared over a large range of masker-signal delays. With this method, the comparison between on- and off-frequency masker levels can be used to derive BM input-output curves only if it is assumed that the "internal" recovery from forward masking (i.e., after the effects of peripheral filtering and compression) is the same for both on- and off-frequency maskers.

Unlike the first two assumptions, the third—of masker-independent recovery from forward masking—can be tested psychophysically. There are some studies that have compared forward masking at different signal frequencies, using both on-frequency (Jesteadt *et al.*, 1982) and off-frequency (Stainsby and Moore, 2006) maskers. However, these do not address the question of whether the rate of recovery from forward masking changes with masker frequency when the signal frequency is kept constant. We are aware of only one psychophysical study that has addressed this question directly. Nelson and Pavlov (1989) used a 1-kHz signal that was fixed at a level just above the quiet threshold. They measured the TMC, i.e., the level of a forward masker needed to just mask the signal as a function of the masker-signal delay, for masker frequencies of 900, 1000, and 1100 Hz. Although the TMCs appeared to have different slopes for the different masker frequencies, after incorporating a simple normalization based on the presumed attenuation of the masker level by the auditory filter centered on the signal frequency, the data for all three masker frequencies could be fitted very well by one line. Nelson and Pavlov (1989) concluded that the apparent differences between the original slopes were due to the difference between masker levels that were needed to produce the same output of the filter centered on the signal frequency. However, because all the masker frequencies were within 10% of the signal frequency, none of the conditions provides a very strong test of masker-frequency independence for forward masking. In particular, to render this assumption valid for behavioral measures of BM input-output functions, the test would need to include masker frequencies that were well below the signal frequency, to approximate the off-frequency maskers used in the GOM and TMC paradigms. Thus, to our knowledge, no physiological or behavioral study has yet provided convincing support for the assumption, as used by most psychophysical methods for estimating BM gain and compression in humans, that the recovery from forward masking for a given signal is independent of masker frequency.

The aim of this study was to test the hypothesis that the rate of recovery from forward masking is the same for different masker frequencies, once masker effectiveness has been equated at a given masker-signal delay. A confirmation of the hypothesis would strengthen the theoretical underpinnings of compression measures based on TMCs, which rely on the assumption when deriving BM input-output functions. On the other hand, if the hypothesis is not supported, the accuracy of at least some of the derived behavioral estimates of BM input-output functions may be questioned.

II. FORWARD MASKING FOLLOWING ON- AND OFF-FREQUENCY MASKERS

A. Stimuli and procedure

The signal was always a 10-ms 4-kHz tone, gated on and off with 5 ms raised-cosine ramps (no steady-state portion). Detection thresholds were measured for the signal using a three-alternative forced-choice (3AFC) procedure, coupled with an adaptive tracking technique that estimated the 70.7% correct point on the psychometric function (Levitt, 1971). The threshold level of the signal was measured in quiet, in a simultaneous background noise, and in the presence of sinusoidal forward maskers.

1. Thresholds in quiet

In the quiet condition, the signal appeared randomly in one of the three observations intervals while the other two contained silence. The three intervals were marked by lights on a computer screen. At the beginning of each run, the signal was set to a level at which it was clearly audible. The signal level was decreased by 8 dB after two consecutive correct responses and increased by the same step size after each incorrect response until the second reversal was obtained. The step size was then reduced to 4 dB until the fourth reversal. After that, the step was further reduced to 2 dB for the remaining eight reversals. A run terminated after 12 reversals and a threshold estimate was obtained by averaging the signal levels at the last eight reversals. Three thresholds obtained from single runs were averaged to compute the final threshold estimate for each subject. In this and all subsequently described tasks, visual feedback indicating the correct response was provided after each trial.

2. Thresholds in simultaneous masking by noise

In all the forward-masking experiments, a simultaneous low-level background noise was added to reduce any potential "off-frequency listening" (e.g., O'Loughlin and Moore, 1981). The noise consisted of two bands spectrally placed around the signal frequency. The lower-frequency band extended from 2260 to 3200 Hz (a half-octave band) and the higher-frequency band extended from 5200 to 6200 Hz. As the signal level was varied adaptively, the overall noise level was also varied to remain 20 dB below the level required to mask the signal.

To determine the necessary noise levels during the adaptive procedure, the growth of simultaneous masking for the 10-ms 4-kHz signal was measured as a function of the noise level, prior to the main experiment. The 3AFC procedure and the steps used in adaptive tracking were the same as for measuring the threshold for detecting the signal in quiet. In each trial, the noise was turned on 300 ms before the beginning of the first observation interval. The noise continued throughout the three observation intervals and ended with the offset of the signal in the third interval (or where the signal offset would have been if the signal had been presented in the third interval). A new sample of noise was drawn for every trial. Thresholds were measured for overall noise levels between 10 and 80 dB SPL.

Once thresholds had been measured over this range of masker levels, the masked signal thresholds were plotted as a function of the simultaneous masker level for each subject individually and were fitted by a quadratic equation with the coefficients estimated from a least-squares fit. During the forward-masking experiments, the noise level on each trial was set by computing the level based on the quadratic fit to the current signal level, and subtracting 20 dB from the obtained value.

3. Forward-masking thresholds

The recovery from forward masking was measured for the 4-kHz signal presented after a 150-ms forward masker (total duration), gated on and off with 5-ms raised-cosine ramps. The recovery curves were measured for a 4-kHz (on-frequency) masker, and for three off-frequency maskers, two below the signal, with a frequency of 2.4 kHz, and one above the signal, with a frequency of 4.4 kHz. The levels of the 2.4-kHz maskers (below the signal frequency) were 92 and 83 dB SPL (90 and 83 dB SPL for S1); the level of the 4.4-kHz masker (above the signal frequency) was 92 dB SPL. For each off-frequency masker, the level of the on-frequency masker was found that produced a similar masked threshold at a 0-ms delay between the masker offset and the signal onset. To determine the necessary on-frequency masker level, the following steps were performed separately for each off-frequency masker and each subject:

(1) For the off-frequency masker, the masker-signal delay was set to 0 ms and the signal level was varied adaptively to find the masked threshold. The stepping rule in adapting signal levels was the same as for measuring detection of the signal in quiet. Three threshold estimates were obtained and averaged. When the standard deviation of the mean exceeded 6 dB, three additional estimates were obtained and all the single-run thresholds were averaged to obtain the final estimate. On rare occasions when there was a clear outlier (a threshold that was more than 10 dB higher than any of the other single-run estimates), the discrepant threshold was not included in the mean.
(2) The 4-kHz signal was set to a level corresponding to the masked threshold obtained in step 1. The signal was preceded by a 4-kHz (on-frequency) masker and the delay between the masker offset and the signal onset was 0 ms. In the 3AFC task, the on-frequency masker level was varied adaptively to find the level needed just to mask the signal. The step sizes in adapting the masker level were the same as for measuring detection of the signal in quiet, except that the masker level was increased after two consecutive correct responses and decreased after each incorrect response. The masker level at threshold was computed by averaging three to six single-run estimates. As in step 1, occasional outliers (masker levels lower by more than 10 dB than any other estimate) were excluded from the mean.
(3) The level of the 4-kHz masker was set to the value obtained as the final threshold estimate in step 2. A 4-kHz signal was presented after a 0-ms delay, and its level was varied adaptively as in step 1 to find masked threshold. Three single-run threshold estimates were averaged to compute the final estimate. The final estimate was compared with the mean threshold obtained in step 1. If the estimate fell within ± 2 dB of the value obtained in step 1, the masking produced by the 4-kHz masker was considered equivalent to that of the off-frequency masker. In cases where the masked threshold fell outside that range of signal levels, the measurement was repeated with the level of the 4-kHz masker increased or decreased relative to the level obtained in step 2. The level adjustments were made on a trial-and-error basis, but typically one additional adjustment was sufficient to obtain a signal level at threshold that fell in the ± 2 dB range around the level obtained in step 1. In such cases, the "adjusted" 4-kHz masker level was considered equivalent to that of the off-frequency masker.

Upon completion of steps 1–3, recovery functions were measured with the masker levels fixed and the signal level varied adaptively, for masker-signal offset-onset delays ranging from 0 to 115 ms. For half the listeners, recovery functions were first measured for the 92-dB 2.4-kHz masker and the equivalent (at a 0-ms delay) 4-kHz masker. The recovery functions for the 83-dB 2.4-kHz masker and the equivalent 4-kHz masker were measured subsequently. For the other half of the listeners, testing was carried out in the reverse order. For listener S6, recovery functions were first measured in the condition involving the 4.4-kHz masker. This listener was later tested using the two 2.4-kHz maskers. Listeners S1–S3 were tested with the 4.4-kHz masker after they completed the conditions involving the 2.4-kHz maskers. Thus, the experiment consisted of three masking conditions, each involving the measurement of two recovery functions, one

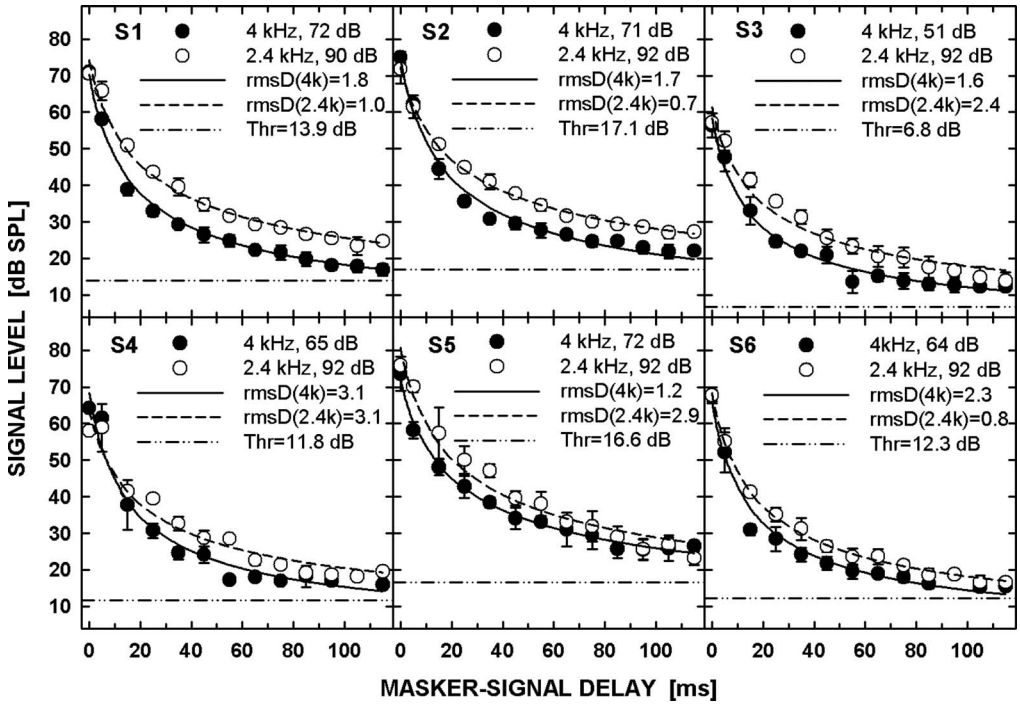


FIG. 1. Recovery from forward masking in the LF_HL condition, i.e., for a 2.4-kHz masker presented at 92 dB SPL (open symbols) and a 4-kHz masker that produced similar masked thresholds at a 0-ms delay (filled symbols). The dashed and solid lines show power-function fits to the data for the 2.4- and 4-kHz maskers, respectively. Value of rms deviations (rmsD) are provided in the legend, for each fitted function. The dashed-dotted line shows threshold for detecting the 4-kHz signal in quiet.

for an off-frequency masker and the other for an on-frequency masker. The condition involving a 92-dB SPL 2.4-kHz masker will be referred to as the low-frequency high-level (LF_HL) condition; the condition involving an 83-dB SPL 2.4-kHz masker will be referred to as the low-frequency low-level (LF_LL) condition, and the condition involving a 92-dB SPL 4.4-kHz masker will be referred to as the high-frequency high-level (HF_HL) condition.

For a given condition, the masker order (on- or off-frequency masker first) was selected randomly for the measurement of each recovery function and for each subject. Once the masker frequency was selected, the recovery function was measured with the masker-signal delay chosen randomly for each repetition. A new random order was selected for each repetition of each condition. Three to six threshold estimates were obtained for each delay. The rule for eliminating outliers was the same as for step 1 listed above. About 5% of runs were discarded as outliers.

Off-frequency listening was controlled by the presence of the background noise, as described above. Precautions were also taken to reduce or eliminate the effect of temporal confusion that has been shown to affect the amount of on-frequency forward masking at short masker-signal delays (Moore and Glasberg, 1982; Neff, 1986). A 70-dB SPL 7-kHz tone was gated on and off with the masker, and was presented to the contralateral ear. The tone served as a cue that marked the temporal beginning and end of the masker. The contralateral tone was used in all conditions, including the measurement of growth of simultaneous masking for the short signal presented in background noise.

All the stimuli were generated digitally on a personal computer with a sampling rate of 48 kHz and played out via a 24-bit LynxStudio Lynx22 sound card. The stimuli were presented via the left earphone of a Sennheiser HD 580 headset, except for the contralateral cuing tone, which was presented via the right earphone. The listeners were tested in a double-walled sound-attenuating booth and responded via a computer keyboard or mouse.

B. Listeners

Seven listeners participated in the study. All had normal hearing as evidenced by their quiet thresholds measured using an ANSI certified audiometer (Madsen Conera). The listeners had thresholds that were below 15 dB HL for audiometric frequencies between 0.25 and 8 kHz. All the listeners were paid for their participation, except S1 who is the first author. Not every listener participated in every condition: Listeners S1–S6 completed conditions LF_HL and LF_LL, and four listeners from that group (S1–S3 and S6) and an additional listener (S7) completed condition HF_HL. The listeners received at least 2 h of practice before the data collection commenced.

C. Results

Functions representing recovery from forward masking in the LF_HL condition for the six listeners are shown in Fig. 1. Each panel shows data from one listener. Masked thresholds for a 92-dB SPL 2.4-kHz masker, plotted as a function of masker-signal delay, are shown by the open circles. Masked thresholds for the 4-kHz masker that produced an equivalent masked threshold at a 0-ms delay are shown by the filled circles. For convenience, such an on-frequency masker will be referred to hereafter as an "equivalent

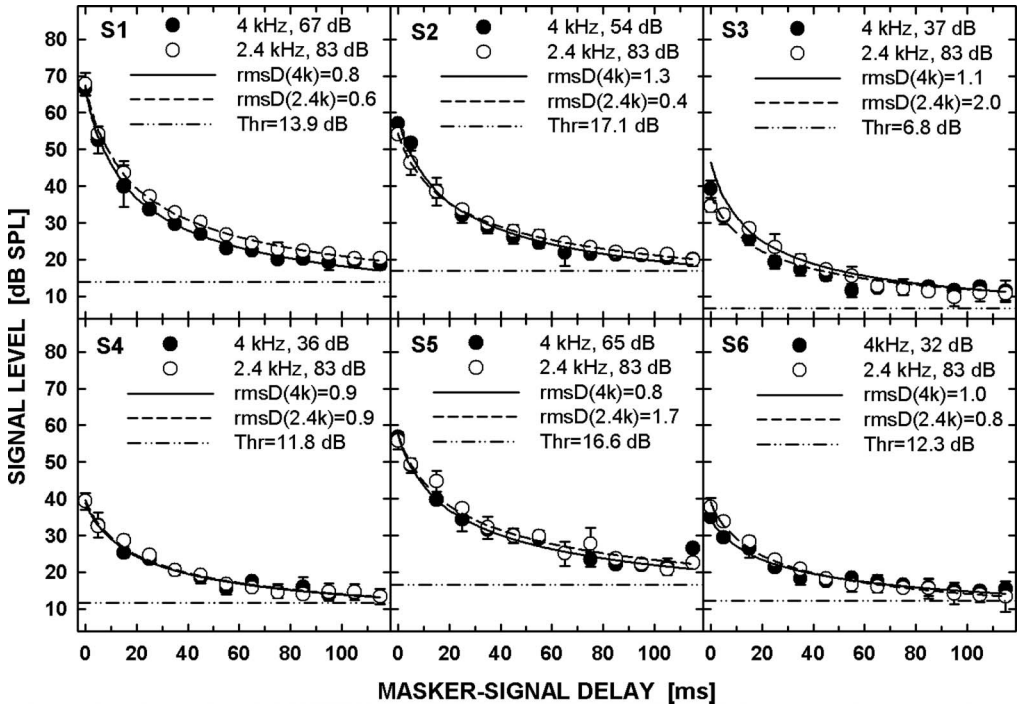


FIG. 2. As Fig. 1 except for the LF_LL condition, in which the 2.4-kHz masker was presented at 83 dB SPL.

masker" even if it did not produce the same masked thresholds at delays other than 0 ms. The levels of the equivalent 4-kHz masker are presented in the legend for each listener. The dashed horizontal line represents the threshold for detecting the 10-ms 4-kHz signal in quiet.

At the 0-ms delay, the two maskers produced very similar amounts of masking, as intended. As the delay increased, the two functions diverged for all listeners, and thresholds for the off-frequency masker fell above those obtained for the on-frequency masker. This indicates that the rate of recovery was slower for the off-frequency masker than for the on-frequency masker. For four listeners (S3–S6), masked thresholds for the two maskers became similar at the longest masker-signal delays. For the on-frequency masker, the masked threshold remained constant for masker-signal delays greater than 55 ms for S3 and S4, and delays greater than 85 ms for S5, despite the fact that recovery was not yet complete. Over the same range of delays, the threshold for the off-frequency masker continued to decrease. Because the on-frequency masking reflected by this residual threshold elevation exhibited extremely slow recovery and was not consistently present in the data of all the listeners, the data should probably not be interpreted as reflecting faster recovery for the off-frequency masker at long delays.[1] The solid lines in Fig. 1 represent power-function fits to the data. The fitted function was defined as

$$L_s = a(d/10)^{-b}, \tag{1}$$

where L_S represents the signal level (dB SPL) at masked threshold predicted by the fitted function, d is the delay (milliseconds) between the masker offset and the signal offset, and a and b are free parameters that were adjusted to produce the best fit using the least-squares method. Parameter b determines the rate of recovery from forward masking. The offset-offset interval (rather than the offset-onset interval) was selected to avoid values of zero, which produce an undefined value of L_s.

Data for the LF_LL condition are presented in Fig. 2. The open symbols show masked thresholds obtained for the 83-dB SPL 2.4-kHz masker and the filled symbols show thresholds for the equivalent 4-kHz masker. In contrast to the LF_HL condition, the two maskers produced similar thresholds at all delays between the masker and the signal, suggesting the same rate of recovery from forward masking for the lower-level (83 dB SPL) off-frequency masker and equivalent on-frequency masker.

Figure 3 shows comparisons between the values of parameter b that determined the recovery rates in Eq. (1) for the on- (filled bars) and off-frequency (unfilled bars) maskers. The left and right panels show the comparisons for the LF_HL and LF_LL conditions, respectively.

For the LF_HL condition, the values of exponent b are smaller for the 2.4-kHz masker, indicating a slower rate of recovery than for the equivalent 4-kHz masker, for all the listeners except S5. A paired t-test comparing b values for the two maskers confirmed that the difference was significant [$t(5)=4.37$, $p=0.007$]. This result indicates that, despite producing the same amount of masking at the selected masker-signal delay of 0 ms, the rate of recovery from forward masking differed between the two maskers. Because the signal frequency was the same in both cases, and because the range of signal levels was very similar, effects of peripheral compression cannot explain the observed differences between the recovery rates.

In the LF_LL condition, three listeners showed a slightly slower recovery for the 2.4-kHz masker (S1, S2, and S3), but the other three showed either no difference between the re-

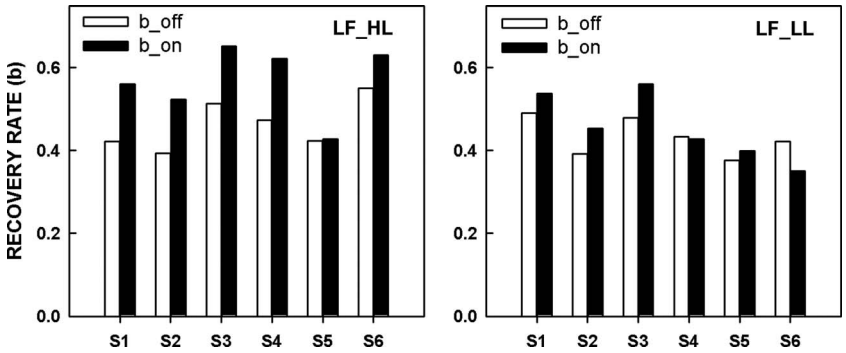


FIG. 3. Values of exponent b in Eq. (1) that characterizes the rate of recovery from forward masking for a 2.4-kHz masker (open bars) and a 4-kHz masker that produced similar masked threshold at a 0-ms delay (filled bars). Values of parameter b are shown for the LF_HL condition (left panel) and the LF_LL condition (right panel).

covery rates for the two maskers or a slight difference in the opposite direction. A paired t-test showed that the difference between the values of exponents b for the two maskers was not significant in this condition [$t(5)=1.02$, $p=0.35$]. Thus, in contrast to the results from the LF_HL condition, data from the lower masker level generally follow a pattern that is consistent with the assumption of equal rates of recovery, independent of masker frequency.

The final condition, HF_HL, allowed us to test whether the different on- and off-frequency decay curves found in the LF_HL condition were due solely to the high sound pressure level of the masker, or whether the results also depended on masker frequency. Figure 4 shows recovery functions for the HF_HL condition obtained for five listeners. Open symbols represent masked thresholds plotted as a function of the delay between the masker offset and the signal onset, for the 92-dB SPL 4.4-kHz masker. Filled symbols show thresholds for the equivalent 4-kHz masker. Equation (1) was fitted to the data and the values of exponent b were estimated. The rates of recovery for the 4- and 4.4-kHz maskers were very similar for all the listeners with the exception of S7. A paired t-test revealed no significant difference between the b-values for the two maskers [$t(4)=0.133$, $p=0.9$]. As with the LF_LL condition, the results from the HF_HL condition are consistent with the hypothesis that the recovery from forward masking is independent of masker frequency. Thus, it appears that a high off-frequency masker level is not sufficient in itself to produce different on- and off-frequency decay curves.

In summary, for two of the three conditions (LF_LL and HF_HL), recovery from forward masking was found to be the same for on- and off-frequency maskers. However, for the LF_HL condition, this was not the case; instead, the 92-dB SPL off-frequency masker produced a slower recovery from forward masking than the on-frequency masker, even though both produced the same amount of masking at a 0-ms

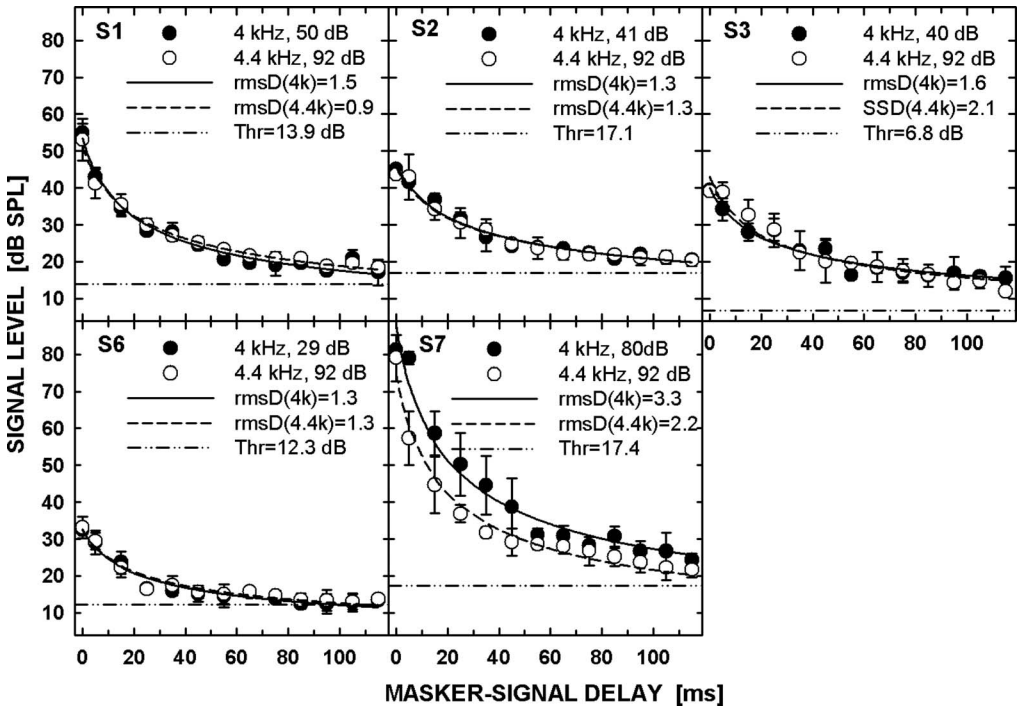


FIG. 4. Recovery from forward masking in the HF_HL condition, i.e., for a 4.4-kHz masker presented at 92 dB SPL (open symbols) and a 4-kHz masker that produced similar masked threshold at a 0-ms delay (filled symbols). Otherwise as Fig. 1.

masker-signal delay. The LF_HL and LF_LL conditions are directly relevant to studies that have estimated BM compression using psychophysical TMC method, as they rely on comparing on-frequency and off-frequency forward-masking thresholds with the off-frequency masker well below the signal frequency and often at high levels (as in our LF_HL condition).

D. Discussion

The listeners in our study exhibited a slower recovery rate for the high-level 2.4-kHz masker than for the 4-kHz masker that produced an equivalent amount of masking at a 0-ms delay. At face value, the results may have important implications for behavioral estimates of BM compression and gain in humans obtained by comparing the slopes of on- and off-frequency TMCs. But first, other potential explanations, not involving different underlying rates of decay of forward masking, should be considered.

One potential explanation is that the contralateral 70-dB SPL 7-kHz tone was not effective in eliminating the effect of temporal confusion on the recovery rates. For the on-frequency masker, perceptual similarity between the masker and the signal could increase the amount of forward masking for the shortest masker-signal delays (Moore and Glasberg, 1982; Neff, 1986). As the delay increased, the effect of similarity would diminish leading to a steeper slope of the recovery function. For the off-frequency masker, the pitch of the signal was distinctly different from that of the masker, and thus the slope of the recovery function would not be affected by temporal confusion at short masker-signal delays. However, it is not clear why temporal confusion could produce a difference between on- and off-frequency maskers in the LF_HL condition, but not in the LF_LL or HF_HL condition. In all conditions, the level of the cuing tone was either similar to or higher than the level of the on-frequency masker, ensuring that the cue tone was clearly audible and salient. Also, comparing across conditions (especially LF_HL and LF_LL), no consistent relationship between the similarity of levels of the masker and the cue and the presence of a difference between recovery rates for the on- and off-frequency maskers is apparent. For example, the on-frequency masker levels for S1, S4, and S6 in Fig. 1, where the difference in recovery was observed, were similar to the on-frequency masker levels for S1, S2, and S5 in Fig. 2, where no difference was observed.

Another potential explanation involves the basalward shift in the peak of the BM traveling wave with increasing stimulus level (e.g., Ruggero *et al.*, 1997). For the high-level 2.4-kHz masker, the peak of the excitation pattern produced by the signal might shift apically with decreasing signal level during the course of recovery toward the peak of the excitation pattern produced by the masker. This could lead to a slower recovery rate. For the on-frequency masker, as the masker-signal delay increased, the peak of the excitation pattern produced by the signal would also shift apically and therefore away from the peak of the excitation pattern produced by the masker, making the signal more detectable at a given masker-signal delay and thus leading to a faster recovery from forward masking. The difference between the rates of recovery for the on- and off-frequency maskers might not be observed in the LF_LL and HF_HL conditions because the maximum signal level at threshold may have been too low for substantial peak migration to occur during the course of recovery. This potential problem would not occur for TMC measurements because the signal is typically kept at a fixed low level, and so reflects the response of a fixed place along the BM, whether or not that place represents the peak of excitation produced by an on-frequency masker.

A number of animal studies have investigated the shift of the peak of BM vibrations with increasing level in the basal end of the cochlea using different paradigms. When isointensity curves are measured, i.e., when a given place on the BM is stimulated with equal-intensity tones of differing frequency, the plot of the BM responses as a function of frequency of the stimulating tone exhibits a gradual shift in the position of the peak response toward lower frequencies with increasing level of the stimulus (Ruggero *et al.*, 1997; Ruggero *et al.*, 2000). These isolevel functions can be considered as equivalent to the responses of the auditory filter for different levels of the stimulus, if they are measured at frequencies for which the contributions of the outer and middle ear transfer functions are negligible. In contrast, when responses to a tone with a fixed frequency are measured at different places on the BM around the CF, for different levels of the test tone, the spatial patterns of BM responses exhibit broadening on the basal side with increasing level but the position of the peak of these patterns shifts toward the base only for levels of 90 dB SPL or higher (Russell and Nilsen, 1997; Ren, 2002). The spatial patterns can be thought of as equivalent to the excitation patterns produced by a tone presented at different levels. Thus, there is consensus regarding peak migration between the different experimental paradigms at very high levels (90 dB SPL or more) but not at lower levels. The lack of consensus over a wide range of levels may simply reflect differences between species since different animals were used with different experimental paradigms. However, the data of Nilsen and Russell (2000) suggest they may reflect differences between animals even within the same species. Nilsen and Russell (2000) measured spatial patterns of BM responses in two animals (guinea pigs) and found that one exhibited a peak shift only at very high levels (90 and 100 dB SPL), consistent with their earlier reports while the other exhibited a peak shift for a stimulus level as low as 60 dB SPL, although the position of the shifted peak did not change for levels between 60 and 90 dB SPL.

Interestingly, psychophysical human data appear to show similar discrepancies depending on the paradigm used. While tuning curves measured at high frequencies typically exhibit a shift of the best frequency, i.e., the frequency corresponding to the position of the tip of the tuning curve, toward lower frequencies (e.g., Lopez-Poveda *et al.*, 2007), other masking paradigms often show spread of excitation toward high frequencies, but no evidence of a change in peak position except at very high levels (see McFadden, 1986 for a review). Moore *et al.* (2002) measured masking patterns in forward masking and showed evidence of a level-dependent

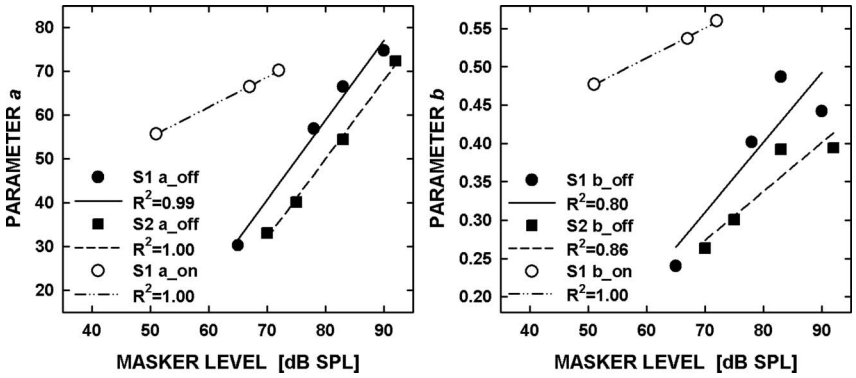


FIG. 5. Values of parameters *a* (left panel) and *b* (right panel) providing the best fit to the forward-masking data by the function described by Eq. (1), plotted as a function of the level of the off-frequency masker (filled symbols) and the on-frequency masker (open symbols). Circles (filled and open) represent the data for S1 and squares for S2. The solid, dashed, and dashed-dotted lines represent linear regression fits to the data.

peak shift for a 6-kHz masker, although the limited number of subjects (two) and the variability in the data preclude strong conclusions. For a 4-kHz masker (the frequency used in our study), their data did *not* exhibit a level-dependent peak shift. Likewise, masking patterns measured by Wojtczak and Viemeister (2006) did not provide evidence for a peak shift, except when the level of a 4-kHz masker was 93 dB SPL.

In all conditions tested in our study, including the LF_HL condition, the signal level at masked threshold did not exceed 80 dB SPL. In view of the previously published data, an explanation in terms of the BM-response peak shift, although appealing, is not supported by data from physiological and psychophysical studies that have explicitly investigated the level dependence of the spectral position of the peak of the excitation pattern.

In summary, two potential factors, "confusion" and BM peak shift, appear unlikely to have contributed to the different masking curves found for the on-frequency and high-level low-frequency maskers. Instead, the results may reflect a true difference in decay of the internal effect of the two maskers. If that is the case, then our findings may have important implications for the BM compression estimates obtained using the TMC method. The following section explores the extent to which the differential forward-masking recovery functions measured here might lead to errors in the estimated BM input-output functions.

III. EVALUATION OF POTENTIAL ERRORS IN COMPRESSION ESTIMATES

When TMCs are measured to derive the BM response and estimate cochlear compression, the signal level is fixed at a low value, typically 10–15 dB SL, and the on- and off-frequency masker levels needed to just mask the signal are measured at different masker-signal delays (Nelson *et al.*, 2001; Lopez-Poveda *et al.*, 2003; Plack *et al.*, 2004). The present experiment showed that the decay of forward masking for a given signal is not always independent of masker frequency. Because this result contradicts an assumption of the procedure that allows BM input-output functions to be derived from TMC data, it is possible that some systematic errors may have resulted. In this section, we attempt to evaluate the extent to which TMC-based estimates of BM compression may be affected by this assumption.

A. Methods

To evaluate the potential error that results from the observed difference in recovery rates, TMCs were simulated using our data. In order to derive TMCs, with a fixed signal level and a variable masker level, from our data, with a fixed masker level and a variable signal level, some interpolation of the data was necessary. In particular, it was assumed that parameters *a* and *b* in Eq. (1) are linear functions of masker level (see below for justification of this). Linear regression fits to the values of *a* and *b*, plotted as a function of masker level, were obtained separately for each masker frequency and each listener. In most cases, the regression lines were fitted to just two data points, i.e., two values of *a* and two values of *b*, obtained separately for the 4-kHz masker and the 2.4-kHz masker in the LF_LL and LF_HL conditions. For listeners S1 and S2, additional data (not shown) were collected. The recovery of forward masking was measured for the 2.4-kHz masker at levels of 78 and 65 dB SPL for listener S1 and at masker levels of 75 and 70 dB SPL for listener S2. For S1, the recovery function was also measured for the 4-kHz masker at a level of 51 dB SPL. These additional data provided further estimates for parameters *a* and *b*. In these selected cases, the regression lines were fitted to three (S1, 4-kHz masker) or four data points (S1 and S2, 2.4-kHz masker). Parameters *a* and *b* obtained from regression lines were used to generate recovery functions described by Eq. (1) for masker levels selected from the range between those used in the LF_LL and LF_HL conditions in 1-dB steps, separately for the 4- and 2.4-kHz masker. For S1 and S2, the range of levels was extended down to the lowest masker levels used to measure recovery functions for these two listeners. The additional data collected for S1 and S2 provided us with some test as to whether our assumption of linear variation of parameters *a* and *b* with masker level was reasonable. Figure 5 shows values of parameters *a* and *b* plotted as a function of masker level along with the linear regression lines, for cases where more than two points per function were available. The regression lines provided excellent fits to parameter *a* plotted as a function of masker level, for both

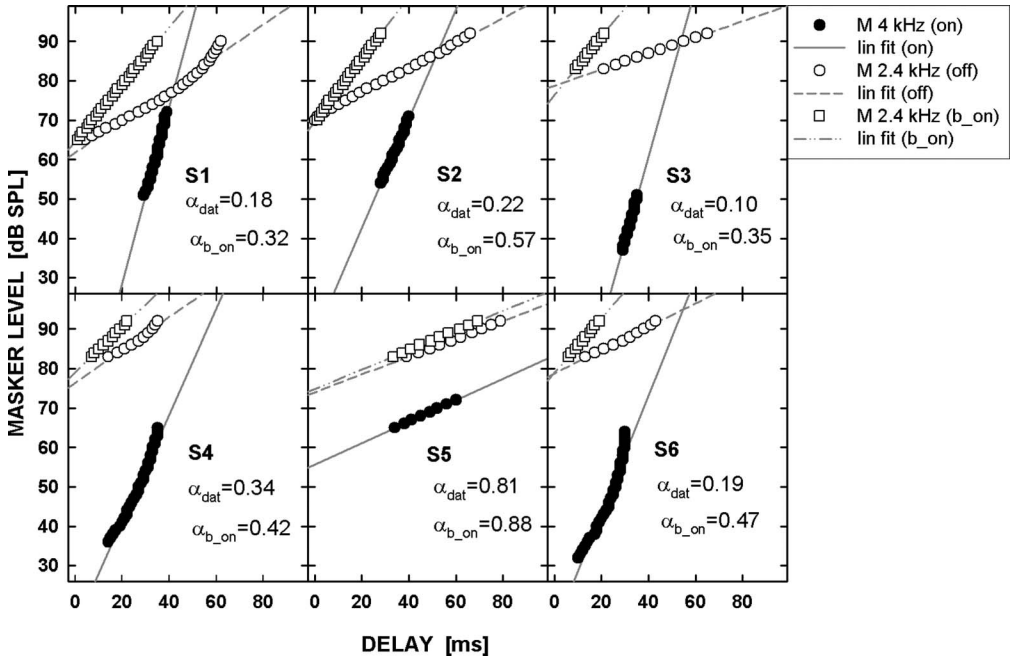


FIG. 6. Temporal masking curves simulated using parameters of best-fitting power functions to the data in Figs. 1 and 2, for the 4-kHz masker (filled circles) and the 2.4-kHz masker (open circles). The temporal masking curve for the 2.4-kHz masker simulated with the value of parameter *b* for the 4-kHz masker is shown by the open squares. The solid, dashed, and dashed-dotted lines show linear regression fits to each curve. The insert shows compression exponent α_{dat}, estimated from the simulated TMCs by computing the ratio of the slopes of the dashed (off-frequency TMC) and solid (on-frequency TMC) lines, and exponent α_{b_on}, estimated from the ratio of the slopes of the dashed-dotted line (off-frequency TMC simulated with parameter *b* for the 4-kHz masker) and the solid line (on-frequency TMC).

the on- and off-frequency maskers. Parameter *b* was well fitted by a regression line only for the on-frequency masker (open circles, dashed-dotted line). For the off-frequency masker, there was a clear departure of *b* from a linear function of level at masker levels between 83 and 92 dB SPL. The results in Fig. 5 demonstrate that, in general, the linear interpolation of the parameters in Eq. (1) was justified. However, by using the values of parameter *b* from the fitted line rather than raw data for the 2.4-kHz masker, we slightly underestimated the potential error in calculating compression for listeners S1 and S2.

B. Results

For each masker level, the delay corresponding to a masked signal threshold of 15 dB SL was found, which effectively simulates the TMC. Figure 6 shows TMCs obtained by plotting the level of the masker against the delay at which that masker produced threshold for detecting the signal at 15 dB SL. The filled circles represent the simulated TMC for a 4-kHz (on-frequency) masker. The open circles represent the simulated TMC for a 2.4-kHz (off-frequency) masker derived directly from the data. This off-frequency TMC would be observed if masker levels at threshold were measured as a function of the masker-signal delay in a typical TMC paradigm. The open squares represent the off-frequency TMC that would be obtained if the recovery rate for the 2.4-kHz masker were the same function of masker level as that for the 4-kHz masker, consistent with the assumption used by the psychophysical methods for deriving compression from TMCs. Thus, the difference between the slopes of the two off-frequency TMCs (open circles versus open squares) directly represents the contribution of the factor that differentially affects psychophysical recovery rates for the on- versus off-frequency masker, but that is *not* related to the difference in peripheral compression.

Straight line fits to the simulated TMCs were used to compute the compression exponents that are provided in the inset in each panel. Exponent α_{dat} represents the compression that would be estimated from the TMC method, provided our simulations accurately represent the results of a task in which the masker level rather than the signal level is measured as a function of the masker-signal delay. For five listeners, the values of α_{dat} were in the range between 0.1 and 0.34, with a mean value of 0.21. This value is in good agreement with compression exponents obtained from TMCs measured for the same signal frequency of 4-kHz, i.e., 0.26 in the study of Lopez-Poveda *et al.* (2003), 0.2 in Plack *et al.* (2004), and 0.23 in Rosengard *et al.* (2005). However, our data indicate that these compression estimates may be affected by a difference between the rates of recovery for the on- and off-frequency maskers that cannot be attributed to differences in compression. Exponent α_{b_on} was obtained from the ratio of the slopes of the straight-line fits to the simulated TMCs shown by the open squares and filled circles. The exponent represents the estimate of compression that would be obtained if the factor that differentially affects the slopes of recovery functions but that is not due to the difference in compression were eliminated in the TMC method. In other words, the compression exponent α_{b_on} would be observed if the assumption of equal recovery rates were valid for the equivalent 2.4- and 4-kHz maskers. The values of α_{b_on} are between 0.32 and 0.57 with the mean value of 0.43, which is

about a factor of 2 larger than the mean value of the compression exponents estimated from a TMC method for the same five listeners. (Data for S5 were excluded from the above comparisons because these data showed very little compression in any condition, which is unusual in a person with normal hearing.) Studies estimating compression based on TMCs have often reported the minimum slope of a third-order polynomial fit to the derived BM input-output functions as a measure of compression. BM input-output functions (not shown here) derived from the simulated TMCs in Fig. 6 were fitted with a third-order polynomial and the minimum slopes were computed for each listener (except S5). The average minimum compression exponent α_{dat} was 0.13, a little lower than the average compression exponents reported in other studies but still well within the range of exponents observed for different individuals. The average "corrected" exponent α_{b_on} was 0.28, which is a little higher but also within the range of individual exponents. The exponents obtained from the third-order polynomial fit to the BM input-output function support the conclusion that compression based on the measured TMCs may be, on average, overestimated by about a factor of 2. It is important to note that this difference does not depend critically on the method used to fit the data. Since the off-frequency TMCs were generally well fitted by a straight line over the range of overlapping delays, the corrected and uncorrected compression estimates decreased by the same factor when the third-order polynomial fit was used.

IV. GENERAL DISCUSSION

A. Comparisons with other studies and implications

Previously reported compression exponents in humans agree with the "uncorrected" estimates, α_{dat}, and also are very similar to the often-reported compression of 0.2 observed in direct measurements of BM responses in guinea pig (Nuttall and Dolan, 1996) and in chinchilla (e.g., Ruggero, *et al.* 1997; Rhode and Recio, 2000). Thus, it may appear that our corrected compression estimates, α_{b_on}, imply responses at the base of the human cochlea that are less compressive than responses in the cochleas of other species. Although differences in the exact amount of compression among different species would not be surprising, given differences in the mechanical characteristics of the BM, such a conclusion is not warranted given the variability of compression estimates across humans and across animals. Compression of 0.2 in animal studies has typically been observed in only one or two cochleas selected from a larger number of sensitive cochleas within the species. In fact, most BM responses measured by Ruggero *et al.* (1997) had slopes between 0.3 and 0.5, and the two cochleas for which the slope was 0.2 at medium input levels exhibited a *decrease* (negative slope) or saturation of the response (no change) with increasing input level [see Fig. 3 of Ruggero *et al.* (1997)].

Our findings are consistent with certain aspects of data from the study by Rosengard *et al.* (2005), which compared slopes of the BM-response functions derived with a 4-kHz signal using two methods, GOM and TMC, in the same listeners. The TMCs measured by Rosengard *et al.* (2005) often contained levels of the off-frequency masker (2.2 kHz) that were in the range between 80 and 100 dB SPL. In contrast, the GOM functions for the off-frequency masker rarely reached levels higher than 85 dB SPL. Interestingly, for four out of five normal-hearing listeners, compression measured at 4 kHz was stronger when estimated from the TMCs than when estimated from GOM. This result is generally consistent with the finding that compression is overestimated when high levels of the off-frequency masker are used.

Psychophysical methods for estimating compression are often used to compare the BM-response growth between listeners with normal and impaired hearing. Because the off-frequency TMC measured at 4 kHz likely represents linear processing (Lopez-Poveda *et al.*, 2003), the slope of the off-frequency TMC should be similar for the two groups of listeners. However, the off-frequency TMCs for hearing-impaired listeners generally exhibit shallower slopes than those measured for listeners with normal hearing (Plack *et al.*, 2004; Rosengard *et al.*, 2005). It is possible that the shallower slopes simply reflect slower temporal processing in hearing-impaired listeners, given that the two groups of listeners were not matched with respect to the listeners' age. The effect of age might have played a role, as a number of studies have shown age-related deficits in temporal auditory processing (e.g., Mendelson and Ricketts, 2001; Fitzgibbons *et al.*, 2007). However, slower temporal processing would likely affect the on- and off-frequency TMCs approximately equally and thus would not affect the estimated compression. In light of our findings, the shallower slopes of the off-frequency TMCs in hearing-impaired listeners may have resulted from higher off-frequency masker levels compared with those used for listeners with normal hearing over the same range of masker-signal delays. In addition, the necessity to use very high off-frequency masker levels may have introduced a difference between the recovery rates for the on- and off-frequency maskers that was not due to the difference in compression. If that were the case, previously reported amounts of compression in the hearing-impaired listeners may also have been overestimated.

Our findings suggest that it would be prudent to avoid using masker levels that are higher than about 83 dB SPL in the measurement of a TMC. However, this may not always be possible, since it would greatly reduce the dynamic range available for deriving the BM response in normal-hearing listeners and would often make it impossible to measure an off-frequency TMC for hearing-impaired listeners.

B. Alternative methods for estimating compression

Because the difference between the recovery rates for the on- and off-frequency maskers is level dependent, compression estimates obtained from GOM may also be affected by the different recovery rates if high off-frequency masker levels are used. GOMs are typically measured using a single short masker-signal delay (Oxenham and Plack, 1997), and thus the potential errors in compression estimates are likely to be smaller than those from the TMCs measured over a wide range of delays. However, the method based on GOM may be inadequate for estimating compression in frequency

regions below about 4 kHz, due to the possibility that the response to the off-frequency masker at the CF_s place is compressive at lower frequencies (e.g., Lopez-Poveda *et al.*, 2003).

Additivity of forward masking may provide a desirable alternative method because it avoids comparisons between the effects of on- and off-frequency maskers (Plack and O'Hanlon, 2003; Plack *et al.*, 2006). The mean compression estimates of 0.17 in the study by Plack and O'Hanlon (2003) and 0.21 in the study by Plack *et al.* (2006) fall in between the average compression exponents α_{dat} and α_{b_on} obtained from the third-order polynomial fits to the BM input-output functions derived from the TMCs shown in Fig. 6. However, it should be noted that the method for estimating compression from additivity of forward masking relies on an assumption of linear summation of the effects of two (or more) forward maskers. At a low signal level (10 dB SL), combining the effects of two equally effective forward maskers leads to a compression estimate that is nearly equal to 1 (Plack and O'Hanlon, 2003). Assuming that the BM-response growth is linear at low levels, this result supports the assumption of linear additivity. It is harder to demonstrate linear additivity of forward masking over a range of medium and high levels, for which the BM response becomes compressive. Some support can be found in the data from listeners with severe cochlear hearing losses (Oxenham and Moore, 1995). Instead of using two forward maskers, a forward and a backward masker were combined in that study and masked thresholds for two equally effective maskers presented together were compared with the masked threshold observed for either masker separately. When equally effective forward and backward maskers were combined, the hearing-impaired listeners exhibited a much smaller shift in threshold relative to the threshold for a single masker than the normal-hearing listeners, consistent with more linear processing. However, in many cases the shift was less than 3 dB at lower SLs and in most of the hearing-impaired listeners, the threshold shift exceeded 3 dB at higher levels. The latter might indicate some residual compression of the BM response even in the presence of a cochlear hearing loss (Plack *et al.*, 2004; Rosengrad *et al.*, 2005). The interpretation of the data for the hearing-impaired subjects in the study of Oxenham and Moore (1995) may also be complicated by the fact that backward masking is poorly understood and tends to diminish after training. Thus, an equally effective backward masker may have become less effective during the course of the experiment. Perhaps the strongest support for linear additivity of the effects of two forward maskers is from a recent study of Plack *et al.* (2007). That study demonstrated that a combined effect of masker M1 presented at a level necessary to mask a signal with a level of L_s+x dB and masker M2 presented at a level needed to mask the signal with a level of L_s dB was the same as the combined effect of M1 presented at a level necessary to mask an L_s dB signal and M2 presented at a level needed to mask an L_s+x dB signal. This commutative behavior is consistent with linear additivity of the effects of two forward maskers. Plack *et al.* (2007) showed that commutation worked for levels up to 70 dB SPL. Thus, at present there is no clear reason to suspect that additivity of masking experiments yield inaccurate estimates of compression. Comparison of compression estimates from additivity of forward masking and the "revised" TMC estimates should be made using the same listeners to determine if the two measures do indeed produce consistent results.

C. Potential mechanisms

The difference in recovery was observed in the LF_HL condition but not in the LF_LL and HF_HL conditions, suggesting that factors such as off-frequency listening and temporal confusion probably cannot account for this result.

It is possible that the difference between recovery rates in the LF_HL condition is mediated by mechanisms at higher levels of the auditory pathways. However, without knowing the site of origin and the mechanisms that contribute to forward masking it is hard to pinpoint the cause of the observed effect. Further research is needed to assess a possible involvement of efferent activity (Shore, 1998; Guinan, 2006) or activation of the middle ear acoustic reflex (Møller, 2000). Higher-level interactions between excitatory and inhibitory inputs characterized by different delays (e.g., Oertel, 1983) could also lead to frequency dependence of the rate of recovery from forward masking.

V. CONCLUSIONS

The purpose of this study was to test the hypothesis that the rate of recovery from forward masking for a given signal is the same for different masker frequencies, provided that the maskers are equally effective at a given masker-signal delay. This hypothesis is at the core of some psychoacoustic methods for estimating BM compression in humans, in particular, the TMC method. The following conclusions can be drawn from the data.

(1) When a high-level (90–92 dB SPL) off-frequency masker was nearly an octave below the signal frequency (LF_HL condition), different recovery rates were observed for the equivalent on- and off-frequency maskers. In contrast, no difference between the recovery rates was found for lower levels of the low-frequency masker (83 dB SPL) or for high-frequency maskers (4.4 kHz) at the same high level (92 dB SPL).
(2) Simulated TMCs, obtained using the parameters of recovery functions fitted to masked thresholds, provide compression estimates which suggest that compression from TMCs may be overestimated by as much as a factor of 2 when high-level off-frequency maskers are used.
(3) Since the difference between the rates of recovery from forward masking depended on the level of a 2.4-kHz masker, the TMC and GOM methods may overestimate compression when high levels of the off-frequency masker are used, although the potential error is likely to be smaller for the GOM-based estimates.
(4) The results suggest that caution should be exercised in interpreting TMC data in terms of BM compression, particularly when the off-frequency masker levels exceed about 85 dB SPL.

ACKNOWLEDGMENTS

The study was supported by NIH Grant No. R01DC03909. The authors thank the Associate Editor Brian C. J. Moore and two anonymous reviewers for their thoughtful comments on earlier versions of the manuscript.

[1]A function defined by a sum of two exponentials was initially fitted to each listener's data. The shorter time constants were always greater for the off-frequency masker than for the on-frequency masker. The longer time constants did not consistently follow the same trend across listeners. For listeners S3, S4, and S5 the longer time constant was greater for the on-frequency masker, but it was extremely long ($>10^{10}$ ms). This suggests that a mechanism other than that underlying forward masking at shorter delays played a role in threshold elevation for these three listeners.

Fitzgibbons, P. J., Gordon-Salant, S., and Barrett, J. (**2007**). "Age-related differences in discrimination of an interval separating onsets of successive tone bursts as a function of interval duration," J. Acoust. Soc. Am. **122**, 458–466.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Guinan, J. J., Jr. (**2006**). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," Ear Hear. **27**, 589–607.

Jesteadt, W., Bacon, S., and Lehman, J. R. (**1982**). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950–962.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am. **113**, 951–960.

Lopez-Poveda, E. A., Barrios, L. F., and Alves-Pinto, A. (**2007**). "Psychophysical estimates of level-dependent best-frequency shifts in the apical region of the human basilar membrane," J. Acoust. Soc. Am. **121**, 3646–3654.

McFadden, D. (**1986**). "The curious half octave shift: Evidence for a basalward migration of the traveling-wave envelope with increasing intensity," in *Basic and Applied Aspects of Noise-Induced Hearing Loss*, edited by R. J. Salvi, D. Henderson, R. P. Hamernik, and V. Colletti (Plenum, New York), pp. 295–312.

Mendelson, J. R., and Ricketts, C. (**2001**). "Age-related temporal processing speed deterioration in auditory cortex," Hear. Res. **158**, 84–94.

Møller, A. R. (**2000**). *Hearing: Its Physiology and Pathology* (Academic, San Diego).

Moore, B. C. J., and Glasberg, B. R. (**1982**). "Contralateral and ipsilateral cuing in forward masking," J. Acoust. Soc. Am. **71**, 942–945.

Moore, B. C. J., Alcántara, J. I., and Glasberg, B. R. (**2002**). "Behavioral measurement of level-dependent shifts in the vibration pattern on the basilar membrane," Hear. Res. **163**, 101–110.

Murugasu, E., and Russell, I. J. (**1995**). "Salicylate ototoxicity: The effects on basilar membrane displacement, cochlear microphonics, and neural responses in the basal turn of the guinea pig cochlea," Aud. Neurosci. **1**, 139–150.

Neff, D. L. (**1986**). "Confusion effects with sinusoidal and narrow-band noise forward maskers," J. Acoust. Soc. Am. **79**, 1519–1529.

Nelson, D. A., and Pavlov, R. (**1989**). "Auditory time constants for off-frequency forward masking in normal-hearing and hearing-impaired listeners," J. Speech Hear. Res. **32**, 298–306.

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). "A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 2045–2064.

Nilsen, K. E., and Russell, I. J. (**2000**). "The spatial and temporal representation of a tone on the guinea pig basilar membrane," Proc. Natl. Acad. Sci. U.S.A. **97**, 11751–11758.

Nuttall, A. L., and Dolan, D. F. (**1996**). "Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig," J. Acoust. Soc. Am. **99**, 1556–1565.

Oertel, D. (**1983**). "Synaptic responses and electrical properties of cells in brain slices of the mouse anteroventral cochlear nucleus," J. Neurosci. **3**, 2043–2053.

O'Loughlin, B. J., and Moore, B. C. J. (**1981**). "Off-frequency listening: Effects on psychoacoustical tuning curves obtained in simultaneous and forward masking," J. Acoust. Soc. Am. **69**, 1119–1125.

Oxenham, A. J., and Moore, B. C. J. (**1995**). "Additivity of masking in normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **98**, 1921–1934.

Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.

Plack, C. J., and O'Hanlon, C. G. (**2003**). "Forward masking additivity and auditory compression at low and high frequencies," J. Assoc. Res. Otolaryngol. **4**, 405–415.

Plack, C. J., Carcagno, S., and Oxenham, A. J. (**2007**). "A further test of the linearity of temporal summation in forward masking (L)," J. Acoust. Soc. Am. **122**, 1880–1883.

Plack, C. J., Drga, V., and Lopez-Poveda, E. A. (**2004**). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am. **115**, 1684–1695.

Plack, C. J., Oxenham, A. J., and Drga, V. (**2006**). "Masking by inaudible sounds and the linearity of temporal summation," J. Neurosci. **26**, 8767–8773.

Ren, T. (**2002**). "Longitudinal pattern of basilar membrane vibration in the sensitive cochlea," Proc. Natl. Acad. Sci. U.S.A. **99**, 17101–17106.

Rhode, W. S., and Robles, L. (**1974**). "Evidence from Mösbauer experiments for nonlinear vibration in the cochlea," J. Acoust. Soc. Am. **55**, 588–596.

Rhode, W. S., and Recio, A. (**2000**). "Study of mechanical motions in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 3317–3332.

Rosengard, P. S., Oxenham, A. J., and Braida, L. D. (**2005**). "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **117**, 3028–3041.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.

Ruggero, M. A., Narayan, S. S., Temchin, A. N., and Recio, A. (**2000**). "Mechanical bases of frequency tuning and neural excitation at the base of the cochlea: Comparison of basilar-membrane vibrations and auditory-nerve-fiber responses in chinchilla," Proc. Natl. Acad. Sci. U.S.A. **97**, 11744–11750.

Russell, I. J., and Nilsen, K. E. (**1997**). "The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane," Proc. Natl. Acad. Sci. U.S.A. **94**, 2660–2664.

Ryugo, D. K., and Parks, T. N. (**2003**). "Primary innervation of the avian and mammalian cochlear nucleus," Brain Res. Bull. **60**, 435–456.

Shore, S. E. (**1998**). "Influence of centrifugal pathways on forward masking of ventral cochlear nucleus neurons," J. Acoust. Soc. Am. **104**, 378–389.

Stainsby, T. H., and Moore, B. C. J. (**2006**). "Temporal masking curves for hearing-impaired listeners," Hear. Res. **218**, 98–111.

Wickesberg, R. E., and Oertel, D. (**1988**). "Tonotopic projection from the dorsal to the anteroventral cochlear nucleus of mice," J. Comp. Neurol. **268**, 389–399.

Wojtczak, M., and Viemeister, N. F. (**2006**). "Estimating the rate of response growth using above characteristic-frequency masker (A)," J. Acoust. Soc. Am. **119**, 3331.

Within- and across-channel factors in the multiband comodulation masking release paradigm

John H. Grose,[a] Emily Buss, and Joseph W. Hall III
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7070

(Received 12 November 2007; revised 20 October 2008; accepted 21 October 2008)

Maskers made up of comodulated narrow bands of noise can result in a signal detection advantage due to both within- and across-channel processes. The purpose of this study was to determine whether contributions from these processes could be differentiated on the basis of two stimulus manipulations: (1) onset/offset asynchrony across bands and (2) introduction of a random temporal fringe surrounding the comodulated bands. The hypothesis was that only masking release due to across-channel processing would be disrupted by these manipulations. Five-band comodulated maskers were constructed, and the availability of within- and across-channel cues was varied by adjusting the frequency spacing of the bands; both logarithmic and linear spacings were tested. The signal was a 1 kHz pure tone. Onset/offset asynchrony had different effects depending on the characteristics of the asynchrony. The results were consistent with an interpretation that across-channel, but not within-channel, masking release was disrupted when the flanking bands were presented continuously and the on-signal band was gated. However, the results suggested that both the across-channel and the within-channel masking release were disrupted in conditions where the on-signal band was continuous and the flanking bands were gated on, as well as in conditions where a random temporal fringe was present.
 [DOI: 10.1121/1.3023067]

PACS number(s): 43.66.Dc, 43.66.Mk [BCM] Pages: 282–293

I. INTRODUCTION

The detection of a tonal signal masked by a narrow band of noise centered at the signal frequency can be facilitated by the presence of additional narrow bands of noise that share the same fluctuation pattern as the on-signal masking band. This detection advantage is termed comodulation masking release (CMR), and the measurement technique of employing a complex of comodulated narrow bands of noise will be referred to here as the multiband CMR paradigm. The phenomenon of CMR has been used as an exemplar of across-channel processing in the auditory system, where such processing refers to the synthesis of information transduced in independent (peripheral) frequency channels. In the case of multiband CMR, across-channel processing refers to the use of signal detection cues that derive from the comparison of concurrent stimulus patterns in spectrally discrete regions (for review, see Verhey *et al.*, 2003; Grose *et al.*, 2005a). However, as will be expanded on below, it is evident that within-channel cues can provide some, if not most, of the detection information under some conditions. Here, the detection cues are derived from information contained within the output of a single auditory filter. It is therefore important for the understanding of CMR to be able to distinguish the contributions of within- and across-channel cues to signal detection. The purpose of this study is to assess two manipulations that have been purported to assist in this differentiation using the multiband CMR paradigm.

The initial demonstration of CMR contributed to a growing interest in across-channel mechanisms in signal detection (Hall *et al.*, 1984), but it was noted early on that within-channel cues could contribute strongly to the detection advantage in the multiband CMR paradigm under some conditions (Schooneveldt and Moore, 1987). In their study, Schooneveldt and Moore (1987) demonstrated that the magnitude of masking release was sensitive to the frequency proximity of the (single) comodulated flanking band (CFB), with the largest masking releases occurring for close spacing of the two bands. It was argued that these local threshold minima reflected sensitivity to signal-induced disruptions in temporal beating patterns between the two comodulated bands; i.e., a within-channel cue. One proposed manifestation of this cue, among several considered by Schooneveldt and Moore (1987), was a change in the pattern of phase locking brought about by the signal dominating the neural synchrony during the minima of the beating masker. Single-channel cues for CMR based on temporal envelope interactions have also been proposed by Berg (1996).[1] The extent to which masking release measured with the multiband CMR paradigm represents an across-channel process, therefore, depends on the spectral relations (and levels) of the comodulated noise bands.

In many studies, the assumption of across-channel processing in CMR is implicit in the selection of the frequency spacings of the multiple narrow bands of noise making up the comodulated masker: placement of the individual bands in putatively independent frequency channels is usually assumed to facilitate primarily across-channel processing. However, the assumption of channel independence is often

[a]Electronic mail: jhg@med.unc.edu

not systematically tested. One approach to differentiating within- and across-channel factors is based on the supposition that facets of auditory grouping and across-channel processing are intrinsically linked, whereas this is not the case for within-channel processing. That is, the perceptual organization of a complex sound involves in part the extraction of common features from across the spectrum (e.g., common amplitude modulation), and this extraction constitutes a form of across-channel processing. The notion of an interdependency between across-channel processing and auditory grouping implies that factors resulting in a disruption of auditory grouping have their effect by disrupting across-channel processing but should have no effect on within-channel processing. This rationale has been applied in several contexts where the role of across-channel processing has been explored by manipulating strength of auditory grouping through the use of temporal asynchrony. For example, the question of whether modulation detection interference (MDI)[2] results from a perceptual fusion of the modulating probe and the distal interferer has been tested by introducing gating asynchronies between probe and interferer to weaken their perceptual fusion (Hall and Grose, 1991; Moore and Shailer, 1992; Oxenham and Dau, 2001). Similarly, gating asynchronies have been used to assess the contribution of within- and across-channel processing to comodulation difference detection (CDD)[3] (Moore and Borrill, 2002; Hall *et al.*, 2006).

For multiband CMR, Grose and Hall (1993) showed that the magnitude of masking release was dependent on the gating synchrony between the on-signal masking band and the accompanying CFBs: onset asynchrony across bands elevated threshold for detecting the signal despite the fact that the masker configuration during the listening interval was unaffected by this manipulation. They interpreted this effect as being due to a disruption of the perceptual fusion between the on-signal band (OSB) and the CFBs, and concluded that auditory grouping processes are integral to the across-channel mechanisms underlying CMR. Dau *et al.* (2005) extended this result by applying the gating asynchrony manipulation to complexes of comodulated narrow bands of noise that were specifically selected to be either proximal in frequency (increasing availability of within-channel cues) or well separated in frequency (decreasing availability of within-channel cues). In some conditions the on-signal and flanking masker bands were synchronously gated, and in other conditions the flanking bands were gated on 100 ms prior to OSB onset. Results demonstrated that this gating asynchrony reduced masking release for the widely spaced complexes but not for the narrowly spaced complexes. This finding was also interpreted in the context of auditory grouping, viz., across-channel mechanisms that contribute to CMR are integral to auditory grouping processes in perceptual organization.

Another approach that has been used to shed light on the nature of multiband CMR involves manipulation of masker modulation coherence before and after the comodulated listening interval (Mendoza *et al.*, 1998; Grose *et al.*, 2005b). In this approach, signal detection is measured in a complex of synchronously gated comodulated bands of noise as a function of the characteristics of the noise presented during the interstimulus and intertrial intervals. When this temporal fringe consists of noise bands that are in all respects the same as the comodulated bands, with the exception that their envelopes are independent of each other (random), masking release declines relative to that measured in the absence of the fringe (or relative to that measured with a comodulated fringe). Note that this effect relies on a seamless perceptual continuity between the fringe and the comodulated interval; i.e., the transition from fringe to comodulated interval is perceptually transparent (Mendoza *et al.*, 1998). One way to view the effect of the random temporal surround is that, during the periods outside of the observation intervals, the uncorrelated envelope fluctuations across the random masker might closely resemble across-channel difference cues used to detect the presence of a signal. Because these spurious signal-like cues occur so frequently in the fringe condition, the perceptual weight given to such cues could be reduced, an underweighting that persists into the comodulated observation intervals. Grose *et al.* (2005b) proposed that this sensitivity to the random temporal surround was a trait of across-channel CMR and could, in consequence, be used as a criterion test. In that study, the bands of noise were well separated in frequency, and so it was implicitly assumed that the masking release was due to across-channel processing. However, before it can be concluded that the random temporal fringe manipulation constitutes a criterion test for across-channel CMR, it should be demonstrated that this manipulation does not affect within-channel processing. That is, if the temporal fringe manipulation is a test for across-channel CMR, then it should have no effect on the masking release measured using closely spaced noise bands, which presumably results mainly from within-channel processing. This reasoning is analogous to the logic used by Dau *et al.* (2005) concerning the role of gating asynchronies and auditory grouping in CMR.

The purpose of this study was to further test manipulations that differentially affect within- versus across-channel masking release. Two specific issues were addressed. First, the effect of a random temporal fringe on within-channel processing was assessed to determine the validity of the proposition that this manipulation constitutes a test for across-channel CMR. The second specific issue is the effect of gating asynchronies on within-channel processing. As noted above, Dau *et al.* (2005) provided an important extension to the Grose and Hall (1993) study by testing stimulus configurations designed to accentuate within-channel interactions. However, they tested only conditions where the CFBs were gated on before, and gated off after, the OSB. Whereas they found a reduction—but not an elimination—of the masking release in the across-channel conditions, as did Grose and Hall (1993), they did not test the inverse conditions where the OSB was gated on before, and gated off after, the flanking bands.[4] This configuration was particularly informative in the Grose and Hall (1993) study since it showed a complete elimination of the masking release for asynchronies greater than about 50 ms. A goal of this experiment, therefore, was to include this configuration in a test of

gating asynchrony effects across a range of frequency spacings that would encompass both within- and across-channel processes.

In summary, this study tests the hypothesis that multiband CMR sometimes occurs as a result of across-channel processes, and that these processes are intrinsically linked with auditory grouping. The hypothesis was tested by assessing the effects of two specific stimulus manipulations on masking release: (1) onset/offset asynchronies across the bands making up the comodulated complex and (2) embedding the gated comodulated complex into a random temporal fringe. Both manipulations were tested using comodulated complexes where the noise bands were variably spaced in frequency to preferentially provide either within- or across-channel cues. Two experiments were undertaken. The first adopted and extended the approach of Dau *et al.* (2005) which employed predominantly logarithmic spacing of masker bands. The second experiment used linear spacing of masking bands. The investigation of linear spacing was of interest because within-channel cues in such configurations are likely to be more salient than in logarithmic spacing, an issue developed further in the preamble to experiment 2, below.

II. EXPERIMENT 1. MULTIBAND CMR FOR LOGARITHMIC SPACING

A. Method

1. Observers

Observers were five highly practiced normal-hearing adults (two females). All had pure tone thresholds ⩽20 dB HL at octave frequencies from 250–8000 Hz, and none reported a history of significant ear disease. Ages ranged from 24 to 54 years (mean of 38 years).

2. Stimuli

The signal in all conditions was a 1000 Hz pure tone, 280 ms in duration including 40 ms raised-cosine onset/offset ramps. In all conditions, the signal was masked by a 20 Hz wide band of Gaussian noise centered at 1000 Hz. This OSB was either presented alone or was accompanied by four additional 20 Hz wide bands of noise, two centered above and two centered below the OSB frequency. The bands were logarithmically spaced. The center frequencies were 944, 972, 1000, 1029, and 1059 Hz (1/24th-octave); 794, 891, 1000, 1122, and 1260 Hz (1/6th-octave); 630, 794, 1000, 1260, and 1587 Hz (1/3rd-octave); and 250, 500, 1000, 2000, and 4000 Hz (1-octave). These frequency spacings were selected to span the range from wholly subcritical (falling within the bandwidth of a single auditory filter centered at 1000 Hz) to maximal frequency independence (each band falling within the passband of a separate auditory filter). The 1/6th-octave spacing and the 1-octave spacing are the same as those used by Dau *et al.* (2005) to exemplify within- and across-channel configurations, respectively. Each masker band was presented at 60 dB sound pressure level (SPL).

A total of 30 masking conditions were constructed: 6 reference conditions and 24 conditions employing comodulated five-band maskers during the observation intervals. These conditions are summarized in Table I and shown schematically in Fig. 1. The nomenclature for all conditions used in this study pairs the characteristics of the masker during the observation intervals (the Core) with the characteristics of the masker during the periods outside of the observation intervals (the Surround). Thus, each condition is designated by a [Core/Surround] label. Two of the reference conditions consisted of the OSB presented alone. In one (Cond 1), it was gated on in the listening intervals for 280 ms, including 40 ms onset/offset raised-cosine ramps (Core=OSB; Surround=absent masker, or ∅; i.e., [OSB/∅]). The signal, when it occurred, was therefore gated concurrently with the masker. The second reference condition (Cond 2) consisted of the OSB presented continuously [OSB/OSB]. The remaining four reference conditions (Conds 3–6) consisted of five-band maskers presented continuously, where each band was random with respect to the others [RAN/RAN]. There was one [RAN/RAN] condition for each of the four frequency spacings (1/24th-, 1/6th-, 1/3rd-, and 1-octave).

TABLE I. Summary of masker conditions. Core/Surround denotes the characteristics of the masker during the observation interval (Core) and during the period outside of the observation intervals (Surround). These characteristics were either absent masker (∅), on-signal band alone (OSB), comodulated flanking bands alone (CFB), five-band comodulated complex (COM), five-band codeviant complex (COD), or five-band random complex (RAN). Where four groups of conditions are indicated in a row, this indicates that the same Core/Surround was applied for each of the four frequency spacings (1/24th-, 1/6th- 1/3rd-, and 1-octave). The table also indicates the condition classification.

Cond	Core/Surround	Classification
1	OSB/∅	Reference (gated OSB)
2	OSB/OSB	Reference (continuous OSB)
3-6 (Exp 1)	RAN/RAN	Reference (continuous random bands)
3–6 (Exp 2)	RAN/∅	Reference (gated random bands)
7–10	COM/∅	Baseline (gated comodulated bands)
11–14	COM/OSB	Asynchrony (continuous OSB)
15–18	COM/CFB	Asynchrony (continuous CFBs)
19–22	COM/RAN	Fringe (random)
23–26	COM/COD	Fringe (codeviant)
27–30	COM/COM	Fringe (comodulated)

The remaining 24 conditions fell into three sets. The first set consisted of the baseline comodulated conditions (Conds 7–10). Here, the five-band comodulated masker complexes were gated on for 280 ms, including 40 ms onset/offset raised-cosine ramps, during the listening intervals [COM/∅]. There were four conditions in this set, one for each of the four frequency spacings. The second set of eight conditions addressed the factor of onset/offset synchrony across bands. In four of the conditions (Conds 11–14), one for each of the frequency spacings, the OSB was presented continuously and the CFBs were gated on during the 280 ms listening intervals [COM/OSB]. In the remaining four of the eight conditions (Conds 15–18), the reverse occurred: for each of the four frequency spacings, the CFBs were presented continuously and the OSB was gated on during the 280 ms listening intervals [COM/CFB]. The third set of conditions addressed the factor of a temporal fringe and contained 12

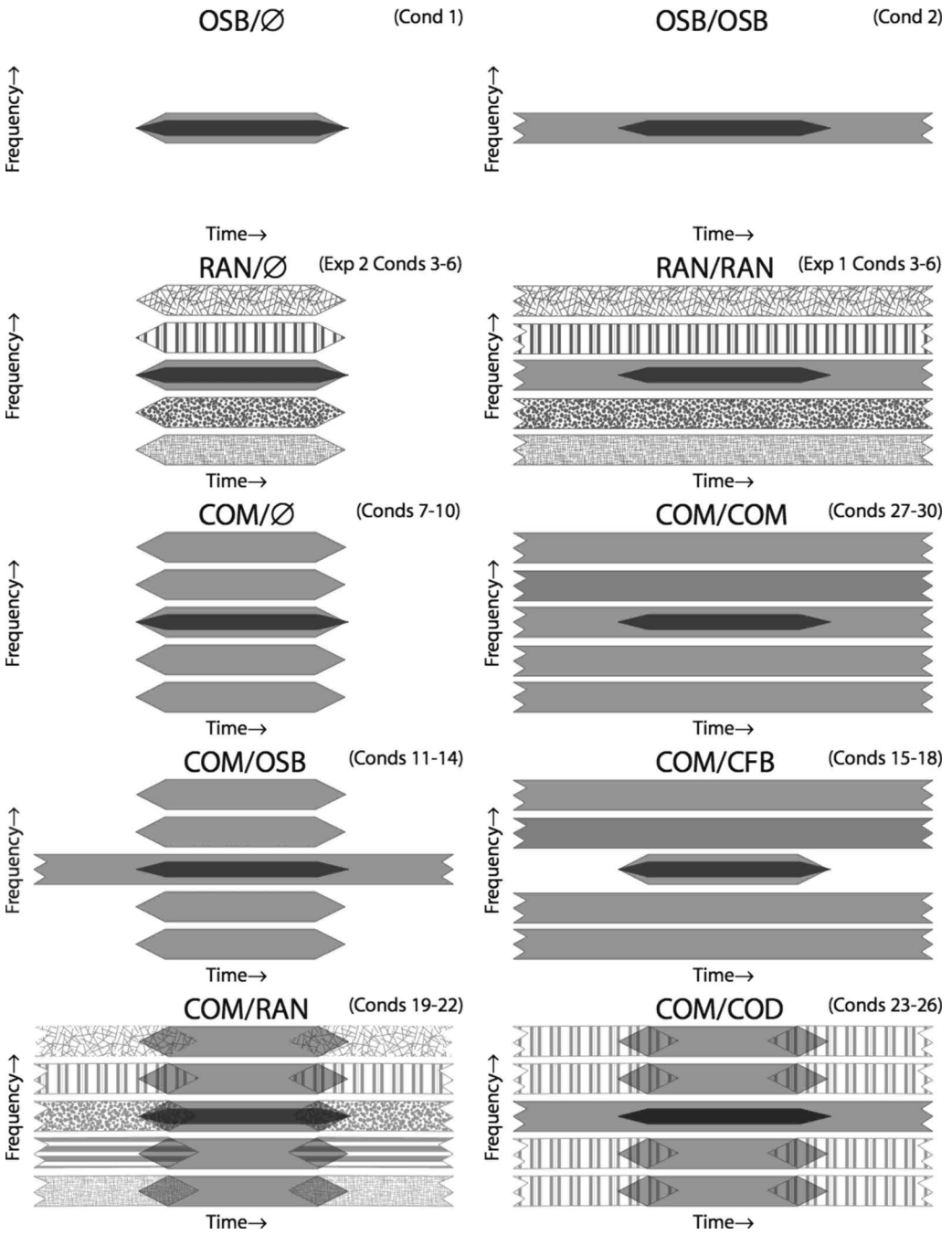


FIG. 1. Schematic representations of the stimulus conditions. Each panel shows a Core/Surround configuration for the masker (see also Table I). The signal is denoted by a black trapezoid; identical fill patterns across masker bands indicate comodulation.

conditions. In 4 of the 12 conditions (Conds 19–22), one for each of the frequency spacings, the five-band comodulated masker was gated on during the listening intervals, but these core comodulated intervals were temporally surrounded by a continuous presentation of five random noise bands that were in all respects the same as the comodulated bands except that each band had an independent fluctuation pattern from all others [COM/RAN]. The transition from random fringe to comodulated listening interval was perceptually seamless and was implemented by overlapping 40 ms gates; i.e., the fringe was gated off as the comodulated core was gated on to begin the listening interval, and the random fringe was gated on as the comodulated core was gated off at the end of the interval. In the next four conditions (Conds 23–26), the temporal fringe consisted of codeviant bands [COM/COD]. Here, the four flanking bands were comodulated with respect to each other, but independently of the OSB. The rationale for this configuration was that, under conditions where across-frequency processing facilitated auditory grouping behavior, the flanking bands might form a separate auditory stream from the OSB, thus perceptually isolating the OSB. Thus, performance in the [COM/COD] configuration might show a dependency on frequency spacing that reflected within- versus across-frequency processing. In the final four conditions (Conds 27–30), the temporal fringe consisted of five comodulated bands [COM/COM]. These conditions are functionally equivalent to the presentation of continuous comodulated noise bands because there was no perceptual dis-

tinction between the comodulated temporal fringe and the comodulated core. This configuration allowed for a comparison between gated and continuous comodulated bands as a function of frequency spacing. A gated/continuous difference is characteristic of CMR (Hatch *et al.*, 1995).

Maskers were generated in the frequency domain based on 2^{17} points. When played out at a sampling rate of 12 207 Hz, this resulted in a waveform segment 10.7 s in duration that repeated seamlessly when played continuously. The coherence or independence of bands across frequency was controlled via the assignment of values to the real and imaginary components in the spectral domain. To generate a single noise band, those points in the complex spectrum corresponding to components within the desired passband were assigned values drawn from a Gaussian distribution; all other points were set to zero. To generate comodulated bands, the same set of random draws was used to define the real and imaginary components of each of the five bands. To generate bands with independent envelope fluctuations across bands, each of the five bands received random values for the real and imaginary components (i.e., independent draws from a Gaussian distribution for each band). To generate bands for the codeviant fringe conditions, one set of random draws was used to define the OSB and a second set of independent draws was used to define the remaining four flanking maskers. Two sets of maskers were generated prior to each threshold estimation track, one set for presentation during the listening intervals and one set for presentation during the interstimulus and intertrial intervals. In conditions where no fringe stimulus was required, the array defining the fringe masker was filled with zeros. Stimulus output and gating were controlled by a digital signal processing platform [Tucker-Davis Technologies (TDT) RP2]. The stimulus was routed through a headphone buffer (TDT HB7) and presented to the left phone of a Sennheiser 265 Linear headset.

3. Procedure

Stimuli were presented in a three-alternative forced-choice paradigm, with a 280 ms observation interval and a 500 ms interstimulus interval. Each observation interval was visually marked by a light on a response box. Observers entered their response after each trial by means of a button on the response box and correct interval feedback was provided by means of lights. Signal thresholds were estimated using a three-down one-up track that converged on the 79% correct point on the psychometric function. Signal level was adjusted in steps of 4 dB, reduced to 2 dB after the second track reversal, and a track continued until eight reversals were obtained. The threshold estimate for a track was the mean signal level at the last six reversals. Three threshold estimates were collected for each condition. If the range of these estimates exceeded 3 dB, a fourth estimate was collected. The final threshold value for a condition was taken as the arithmetic mean of the three (or four) estimates.

B. Results and discussion

The results of the reference and baseline conditions are shown in Fig. 2. All five observers exhibited similar patterns

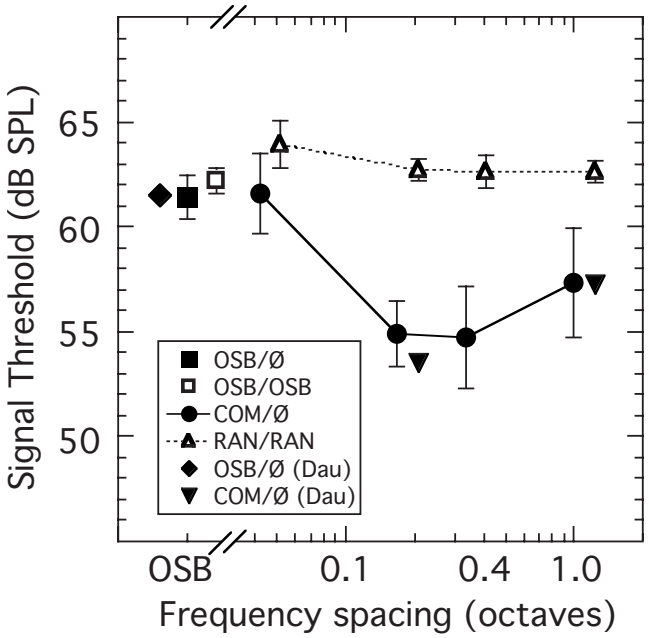


FIG. 2. Mean results for reference and baseline conditions from experiment 1. OSB alone thresholds are shown as filled square for gated presentation (i.e., [OSB/⊘]) and unfilled square for continuous presentation (i.e., [OSB/OSB]); for comparison, filled diamond is [OSB/⊘] from Dau *et al.* (2005). Thresholds in continuous random maskers [RAN/RAN] as a function of band spacing are shown by open triangles. Thresholds in gated comodulated maskers [COM/⊘] are shown as circles; for comparison, filled inverted triangles are [COM/⊘] from Dau *et al.* (2005). Error bars are ±1 standard deviation.

of performance, and so only group means and standard deviations are shown. The filled and unfilled squares indicate, respectively, thresholds in the OSB masker presented alone in gated [OSB/⊘] and continuous [OSB/OSB] modes. The unfilled triangles indicate thresholds in the continuous random bands [RAN/RAN]. Filled circles are thresholds for the gated comodulated bands [COM/⊘]. Also shown are corresponding data from Dau *et al.* (2005) for the gated OSB alone (diamond) and gated comodulated bands (inverted triangles). A preliminary consideration concerns the choice of reference condition for assessing masking release associated with the comodulated masker. For the OSB alone, thresholds were equivalent for the gated and continuous presentation modes ($t_4=1.589$; $p=0.19$). Masking release referenced to the gated OSB was assessed by comparing thresholds for the [OSB/⊘] condition and the four [COM/⊘] conditions using a repeated measures analysis of variance (ANOVA). The analysis showed a significant overall effect of condition ($F_{4,16}=32.95$; $p<0.001$). Simple contrasts indicated that the masking release was significant for the three wider flanking band spacings ($F_{1,4}$ ranging from 19.32 to 109.59; $p<0.012$); no masking release occurred for the narrowest spacing ($F_{1,4}=0.08$; $p=0.80$). Masking release referenced to the continuous random bands was assessed by comparing thresholds for the [RAN/RAN] and [COM/⊘] conditions. The repeated measures ANOVA indicated a significant effect of masker type ($F_{1,4}=63.825$; $p<0.001$), a significant effect of frequency spacing ($F_{1,4}=26.523$; $p<0.001$), and a significant interaction between these two factors ($F_{3,12}=21.911$; $p<0.001$). Analysis of simple effects showed that for each

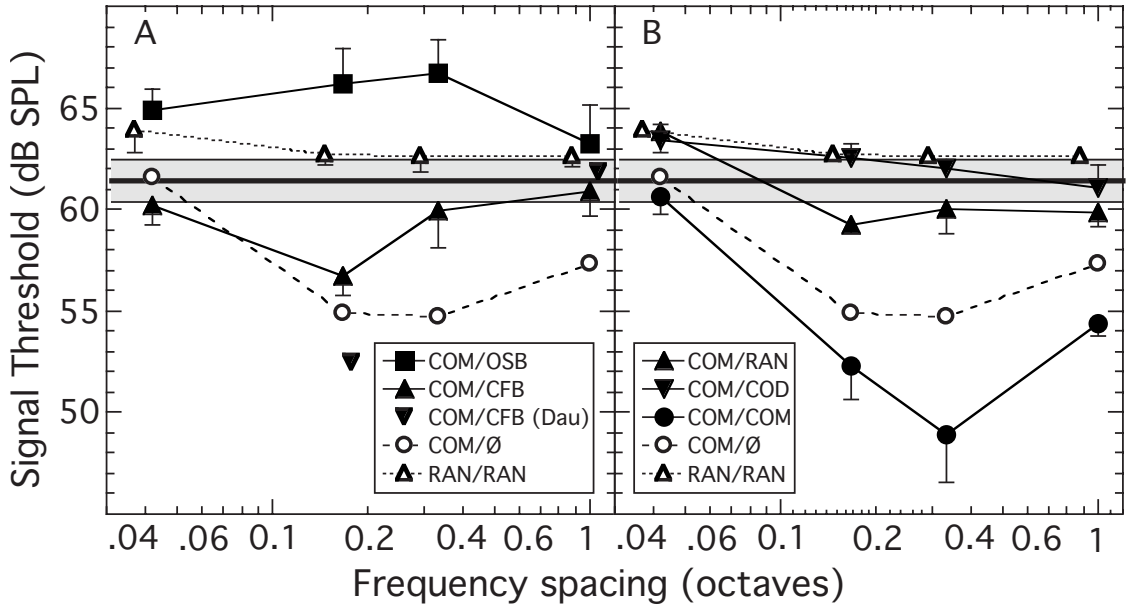


FIG. 3. Mean results for asynchrony conditions (Panel A) and temporal fringe conditions (Panel B) from experiment 1. Panel A: filled squares =[COM/OSB]; filled triangles=[COM/CFB]; filled inverted triangles=[COM/CFB] from Dau *et al.* (2005). Panel B: filled triangles=[COM/RAN]; filled inverted triangles=[COM/COD]; filled circles=[COM/COM]. In each panel, replotted from Fig. 2, are the [OSB/⊘] reference (the thick horizontal line bounded by a ±1 standard deviation shaded region), the [RAN/RAN] reference (open triangles), and the [COM/OSB] baseline (open circles). Error bars are 1 standard deviation.

frequency spacing thresholds in the comodulated masker were significantly lower than in the random masker ($p < 0.02$). For the three wider frequency spacings (1/6th-, 1/3rd-, and 1-octave), therefore, a significant masking release was observed irrespective of whether the reference was the OSB alone or the five-band random masker. However, for the narrowest frequency spacing (1/24th-octave), a significant masking release was observed only when the five-band random masker was used as the reference. (The higher signal threshold in the five-band random masker spaced at 1/24th-octave relative to the single-band OSB masker presumably reflects increased energetic masking.) Because of the precedence of using the OSB as the reference for masking release in this paradigm (Grose and Hall, 1993; Dau *et al.*, 2005), this reference will also serve as the primary anchor in this study. However, where the pattern of data differs depending on which reference is used, consideration of the five-band random masker will also be included.

One stimulus manipulation designed to reduce the strength of auditory grouping among the comodulated noise bands was that of introducing an onset/offset asynchrony across the bands. The asynchrony was introduced in two ways: (1) the OSB was presented continuously and the CFBs were gated on [COM/OSB]; and (2) the CFBs were presented continuously and the OSB was gated on [COM/CFB]. The results for this pair of conditions are shown in Fig. 3, Panel A (filled symbols). For comparison, reference thresholds and baseline thresholds for synchronously gated comodulated bands are replotted from Fig. 2: the [OSB/⊘] reference threshold is shown as a thick horizontal line with the shaded region indicating ±1 standard deviation; the [RAN/RAN] reference thresholds are shown as open triangles; the [COM/⊘] baseline thresholds are shown as open circles. Relative to the baseline gated masking release conditions, signal thresholds appear generally elevated when the OSB is presented continuously and the flanking bands are gated ([COM/OSB], filled squares). This increased masking appears not only to negate any masking release but to increase thresholds above reference thresholds for most spacings. When the flanking bands were presented continuously and the OSB was gated ([COM/CFB], filled triangles), thresholds appear higher than for the baseline masking release conditions in some cases and approach the reference thresholds for the two widest masker spacings. For comparison, data from Dau *et al.* (2005) for [COM/CFB] are also shown (inverted triangles).

In order to test the significance of these data patterns, separate repeated measures ANOVAs were undertaken for each mode of onset asynchrony. For the [OSB/⊘] reference and the asynchrony with the OSB presented continuously [COM/OSB], there was a significant overall effect of condition ($F_{4,16}=13.94$; $p<0.001$). *Post hoc* simple contrasts indicated that thresholds in the asynchronous conditions were elevated relative to that for the OSB alone for all frequency spacings except the widest 1-octave spacing ($F_{1,4}=20.27–36.08$; $p<0.01$). A similar pattern was found using the [RAN/RAN] reference except that a threshold elevation was absent also for the narrowest 1/24th-octave spacing. The [COM/OSB] thresholds were significantly higher than the baseline [COM/⊘] thresholds for all spacings ($F_{1,4}=11.6–133.4$; $p<0.03$).

For the asynchrony conditions with the CFBs presented continuously [COM/CFB], there was a significant overall effect of condition ($F_{4,16}=11.47$; $p<0.001$) using the [OSB/⊘] reference. *Posthoc* simple contrasts indicated that thresholds for the narrowest and widest frequency spacings (1/24th- and 1-octave) did not differ from that for the OSB alone, but thresholds for the 1/6th- and 1/3rd-octave spacings did

($F_{1,4}$=30.16–9.03; $p<0.05$). Thresholds for the [COM/CFB] condition were significantly lower than the [RAN/RAN] reference for all frequency spacings. A repeated measures ANOVA comparing thresholds in the baseline [COM⊘] condition to the asynchronous [COM/CFB] condition indicated a significant effect of masker synchrony ($F_{1,4}$=8.27; $p<0.05$), a significant effect of frequency spacing ($F_{1,4}$=29.26; $p<0.001$), and a significant interaction between these two factors ($F_{3,12}$=11.98; $p=0.001$). Analysis of simple effects showed that significant threshold differences existed only for the two widest spacings ($F_{1,4}$=7.4–36.1; $p\leq0.05$). That is, thresholds in the [COM/CFB] condition were elevated with respect to the baseline [COM/⊘] condition only for the 1/3rd- and 1-octave spacings.

In summary, the pattern of results from the asynchrony manipulation indicates that the effect of onset/offset asynchrony between OSB and CFBs depends in part on the mode of the asynchrony. When the OSB is presented continuously, any signal detection advantage associated with the presence of CFBs during the observation intervals is eliminated, and in most cases thresholds are elevated relative to the reference [OSB/⊘] condition. When the CFBs are presented continuously, the signal detection advantage is reduced only for the wider spacings.

The other stimulus manipulation designed to disrupt use of across-channel cues was that of embedding the gated comodulated complex into a random temporal fringe [COM/RAN]. In this condition, each of the five bands making up the temporal fringe had an envelope that was independent of that of the others; the results of this condition are shown in Fig. 3, panel B (filled upward triangles). As a comparison to the random temporal fringe, two other temporal fringe conditions were also examined. In one, the outer four bands (flanking bands) of the fringe were comodulated with respect to each other, but independently of the envelope pattern of the band centered at the signal frequency ([COM/COD], inverted triangles). The purpose of this condition was to create a temporal fringe where, in the context of auditory grouping mechanisms, the OSB might be perceptually segregated from the accompanying noise bands. In the other fringe condition, all five bands remained comodulated, and this condition amounted to a continuous presentation of comodulated noise bands ([COM/COM], filled circles). The purpose of this condition was to determine the extent to which gated/continuous differences existed for comodulated complexes over the range of frequency spacings tested. It is apparent that signal thresholds in both the random and codeviant fringe conditions were similar to those for the reference [OSB/⊘] condition, indicating little or no masking release. For the continuous comodulated conditions, thresholds were further reduced for the three wider spacings, such that the masking release for continuous comodulated noise exceeded that for baseline gated comodulated noise.

To test the significance of these patterns, separate repeated measures ANOVAs were undertaken for each fringe condition. Using the [OSB/⊘] reference, results for the random temporal fringe [COM/RAN] indicated an overall effect of condition ($F_{4,16}$=21.53; $p<0.001$), and *post hoc* simple contrasts indicated that signal threshold in the random fringe was different from that for the OSB alone for all frequency spacings except the 1/3rd-octave spacing ($F_{1,4}$=11.33–43.67, $p<0.05$). Note that signal threshold was elevated relative to the OSB reference for the 1/24th-octave spacing and lower than this reference for the 1/6th- and 1-octave spacings. Relative to the [RAN/RAN] reference, the [COM/RAN] thresholds were significantly lower for all but the narrowest frequency spacing. Results for the codeviant temporal fringe [COM/COD] relative to the [OSB/⊘] reference indicated a significant overall effect of condition ($F_{4,16}$=6.03; $p<0.005$), and *post hoc* simple contrasts indicated that only the signal threshold for the narrowest spacing (1/24th-octave) was different from reference ($F_{1,4}$=9.61; $p<0.05$). Again, this difference reflected an elevation in signal threshold above reference. The [COM/COD] thresholds did not differ significantly from the [RAN/RAN] reference thresholds for any frequency spacing. Results for the comodulated temporal fringe [COM/COM] relative to the [OSB/⊘] reference indicated a significant overall effect of condition ($F_{4,16}$=91.09; $p<0.001$), and *post hoc* simple contrasts indicated that signal threshold was lower than that for the OSB reference for all frequency spacings ($F_{1,4}$=34.72–199.93, $p<0.005$). The [COM/COM] thresholds were also significantly lower than the [RAN/RAN] reference thresholds for all frequency spacings. An ANOVA comparing thresholds in the gated comodulated maskers [COM/⊘] with those in the continuous gated maskers [COM/COM] indicated that thresholds were lower in the continuous mode of presentation for the 1/6th- and 1/3rd-octave spacing ($F_{1,4}$=12.37 and 32.81, respectively; $p<0.03$); the additional masking release failed to reach significance for the 1/24th-octave spacing ($p=0.28$) and the 1-octave spacing ($p=0.07$).

In summary, this pattern of results indicates that a temporal fringe that does not consist of the same comodulated noise bands as the gated intervals generally eliminates any masking release associated with the gated comodulated bands. When the temporal fringe consists of the same comodulated bands as the gated intervals (i.e., continuous comodulation), the masking release tends to increase for all but the narrowest spectral spacing.

The purpose of this experiment was to determine whether the factors of temporal asynchrony and the presence of a noncomodulated temporal fringe differentially affect multiband masking release depending on whether the frequency spacing of the multiple bands provides predominantly within-channel or across-channel cues. Relative to the [OSB/⊘] reference, significant masking release was found with synchronously gated comodulated maskers for the 1/6th-, 1/3rd-, and 1-octave frequency spacings. A temporal asynchrony wherein the CFBs were presented continuously and the OSB was gated on largely eliminated the masking release for the wider spacings but had minimal effect for the 1/6th-octave spacing. This result is consistent with Dau *et al.* (2005). For the other mode of temporal asynchrony, wherein the OSB was presented continuously and the flanking bands were gated on, an elevation in threshold above that for the OSB reference occurred for all frequency spacings—even the 1/24th-octave spacing where no baseline masking

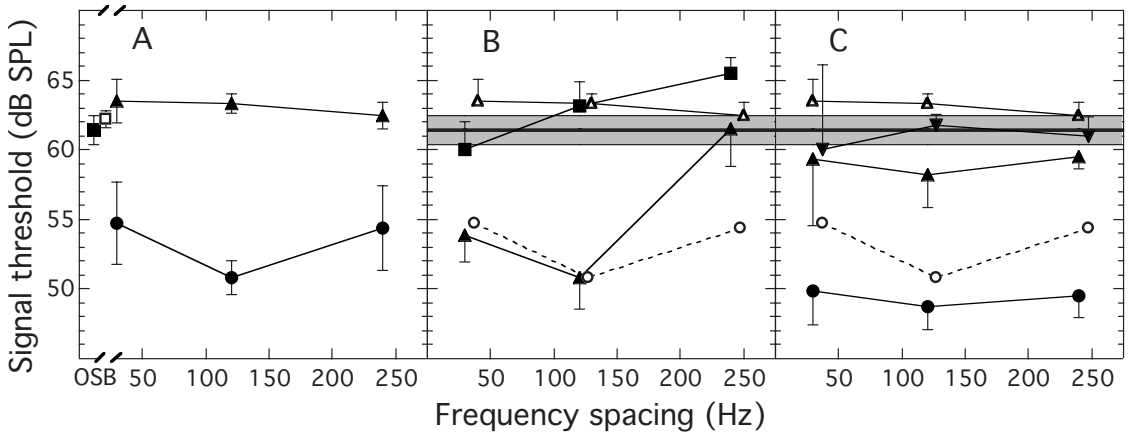


FIG. 4. Mean results from experiment 2. Panel A shows results from the baseline gated comodulated conditions. OSB alone thresholds are shown as filled squares for gated presentation (i.e., [OSB/⊘]) and unfilled squares for continuous presentation (i.e., [OSB/OSB]); thresholds in continuous random maskers [RAN/RAN] as a function of band spacing are shown by triangles. Thresholds in gated comodulated maskers [COM/⊘] are shown as circles. Panel B shows results from the asynchrony conditions, and panel C shows results from the temporal fringe conditions. Panel B: filled squares=[COM/OSB]; filled triangles=[COM/CFB]. Panel C: filled triangles=[COM/RAN]; filled inverted triangles=[COM/COD]; filled circles=[COM/COM]. In each panel, replotted from panel A, are the [OSB/⊘] reference (the thick horizontal line bounded by a ±1 standard deviation shaded region), the [RAN/RAN] reference (open triangles), and the [COM/OSB] baseline (open circles). Error bars are 1 standard deviation.

release was observed. The baseline masking release—including that for the 1/6th-octave spacing—was also largely, or completely, eliminated by the presence of a random or codeviant temporal fringe. A comodulated fringe generally served to increase the magnitude of masking release when it was present in the baseline conditions. For the [RAN/RAN] reference, where a masking release is now observed for the narrowest 1/24th-octave spacing, a broadly similar picture emerges of the effects of temporal asynchrony/fringe as a function of frequency spacing.

Before considering the implications of these findings for the hypothesis being tested in detail, it is prudent to note a limitation in the stimulus configurations used in experiment 1. As in Dau *et al.* (2005), the frequency spacings were defined on a log scale. Whereas this allowed an adjustment in frequency proximity that was approximately uniform in cochlear space, it did not result in configurations that accentuated the cues thought to be important for within-channel processing. As noted in the Introduction, a primary cue for the presence of a signal in a complex of closely spaced comodulated bands is thought to be its disruption of the inherent beating pattern between bands (Schooneveldt and Moore, 1987). If so, then the use of logarithmic spacing is not favorable to this cue since the beat frequencies are not uniform across bands. Instead, linear spacings would be more conducive to the generation of a signal-induced, beat-disruption cue since the pattern of component interactions between neighboring bands are uniform across bands. A companion experiment was therefore undertaken that employed linear spacings of comodulated bands. Its purpose was to improve the cues thought to result in masking release for proximal frequency spacings in order to further clarify the effects of temporal asynchrony and random temporal fringe on within-channel processing of comodulated bands. It was anticipated that a comparison of the results of the two experiments would give a more complete picture of within- and across-channel processing in CMR, and address more fully the hypothesis that the factors of temporal asynchrony and random temporal fringe undermine masking release associated with across-channel processes but not within-channel processes.

III. EXPERIMENT 2. MULTIBAND CMR FOR LINEAR SPACING

A. Method

The stimulus parameters and procedure were the same as experiment 1 except that the frequency spacings of the five-band maskers were linear rather than logarithmic. In addition, the random reference condition was gated [RAN/⊘] instead of continuous [RAN/RAN].[5] The frequency spacings were 30, 120, and 240 Hz, resulting in center frequencies, respectively, of 940, 970, 1000, 1030, and 1060 Hz; 760, 880, 1000, 1120, and 1240 Hz; and 520, 760, 1000, 1240, and 1480 Hz. These separations were selected to be globally similar to the 1/24th-, 1/6th- and 1/3rd-octave spacings of experiment 1. The same observers participated as in experiment 1.

B. Results and discussion

The results of experiment 2 are shown in Fig. 4, following the same labeling conventions as Figs. 2 and 3. In panel A, it can be seen that signal thresholds did not differ between the gated [OSB/⊘] and continuous [OSB/OSB] references ($t_4=1.49$; $p=0.19$). It can also be seen that thresholds in the baseline gated comodulated conditions [COM/⊘] were reduced relative to either the OSB alone reference or the random band reference [RAN/⊘]; i.e., a masking release occurred for all frequency spacings. This was confirmed with repeated measures ANOVAs which indicated that thresholds in the comodulated maskers were each lower than either the [OSB/⊘] reference ($F_{1,4}=28.63-147.87$; $p<0.01$) or the [RAN/⊘] reference ($F_{1,4}=101.1$; $p<0.001$).

The effect of onset asynchrony is shown in panel B of Fig. 4. When the OSB was presented continuously and the

flanking bands were gated on during the listening intervals ([COM/OSB], filled squares), thresholds were comparable to, or higher than, the [OSB/⊘] reference threshold. When the flanking bands were presented continuously and the OSB was gated ([COM/CFB], filled triangles), thresholds were not changed for the two narrowest spacings (30 and 120 Hz), but rose to the reference levels for the widest spacing (240 Hz). This pattern was confirmed with repeated measures ANOVAs. For the [COM/OSB] asynchrony relative to the [OSB/⊘] reference, there was a significant effect of condition ($F_{3,12}=13.21$; $p<0.001$). *Post hoc* contrasts indicated that thresholds did not differ from the reference threshold for the 30 and 120 Hz spacings ($F_{1,4}=2.22$ and 2.41; $p\geqslant 0.20$) but did differ significantly for the 240 Hz spacing ($F_{1,4}=200.05$; $p<0.001$). Relative to the [RAN/⊘] reference, there was a significant interaction between the [COM/OSB] and [RAN/⊘] conditions as a function of frequency spacing ($F_{2,8}=11.27$; $p<0.005$); analysis of simple effects indicated that thresholds for the asynchronous masker were lower for the 30 Hz spacing and higher for the 240 Hz spacing ($F_{1,4}=14.6$ and 23.6, respectively; $p<0.02$), but did not differ for the 120 Hz spacing. For the [COM/CFB] asynchrony referenced to [OSB/⊘] thresholds, there was a significant effect of condition ($F_{3,12}=30.53$; $p<0.001$). *Post hoc* contrasts showed that thresholds were significantly lower than the reference threshold for the two narrowest spacings ($F_{1,4}=66.41$ and 58.28; $p<0.01$) but not for the 240 Hz spacing ($F_{1,4}=0.01$; $p=0.94$). Relative to the [RAN/⊘] reference, there was a significant interaction between the [COM/CFB] and [RAN/⊘] conditions as a function of frequency spacing ($F_{2,8}=21.71$; $p<0.001$); analysis of simple effects indicated that thresholds were lower for the 30 Hz spacing and the 120 Hz spacing ($F_{1,4}=75.81$ and 101.77, respectively; $p<0.001$), but not for the 240 Hz spacing.

The effect of temporal fringe is shown in panel C of Fig. 4. For the random temporal fringe ([COM/RAN], upward triangles), thresholds approach the reference thresholds, although some masking release appears to remain. For the codeviant fringe ([COM/COD], inverted triangles), thresholds largely coincide with the reference thresholds. For the comodulated fringe ([COM/COM], filled circles), thresholds are further reduced relative to baseline. The relatively large standard deviations associated with the data for 30 Hz spacing of the random and codeviant fringe conditions should be kept in mind when considering the data analysis. Individual differences were striking for these conditions and, to highlight this, the individual results are shown in Fig. 5. It can be seen that, for three of the five observers, the random and deviant temporal fringe manipulations resulted in elevated thresholds relative to the baseline gated comodulated condition. However, for the remaining two listeners, the fringe manipulations had little effect. This degree of individual variability in the effect of the temporal fringe for the narrowest 30 Hz spacing suggests that the threshold elevation—when it is observed—is not due to factors associated with energetic masking.

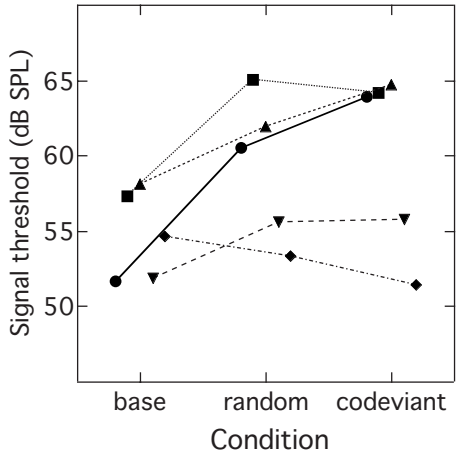


FIG. 5. Individual data for the 30 Hz spacing of experiment 2. Thresholds are shown for the baseline masking release condition, the random temporal fringe, and the codeviant temporal fringe.

IV. GENERAL DISCUSSION

The stimulus conditions of experiments 1 and 2 were designed to provide thresholds under comparable conditions for logarithmic and linear masker spacing. The motivation for experiment 2 was to augment the within-channel beating cue by using linear, rather than logarithmic, spacing. The 30 Hz linear spacing resulted in center frequencies that were similar to those for the 1/24th-octave spacing of experiment 1; whereas no masking release was observed for the log spacing, a 6.7 dB masking release was observed for the linear spacing. This suggests that linear spacing provides a stronger within-channel beating cue than logarithmic spacing does. This conclusion is supported by thresholds obtained at the next wider spacing in both experiments, which were also thought to provide robust within-channel cues. Whereas the 1/6th-octave spacing of experiment 1 led to about a 6.5 dB masking release, the 120 Hz linear spacing of experiment 2 resulted in a masking release of about 10.5 dB. In contrast, the 1/3rd-octave spacing of experiment 1 and the 240 Hz spacing of experiment 2 both gave similar magnitudes of masking release (6.7 and 7.0 dB, respectively). These two spacings had similar spectral distributions, and both were expected to provide primarily across-channel cues. The similar masking releases for these linear and logarithmic spacings suggest that the detection cue here was not associated with a disruption of beat patterns due to the addition of a signal tone.

A final piece of evidence which supports the use of a beating cue in configurations where the frequency spacing is both narrow and linear comes from a supplementary condition that used a modified OSB. Recall that the CFBs were constructed by assigning to their real and imaginary components the same values as assigned to the corresponding components in the OSB. As a result of this, the beating patterns between the components of any neighboring linearly spaced bands were identical. In the supplemental condition, the OSB was modified such that its spectral profile was changed yet its envelope remained comodulated with that of the flanking bands. To achieve this, the frequency spectrum of the OSB was modified by multiplying the phase values by −1 and

reversing the assignment of both magnitude and phase values as a function of component number (see Richards, 1988). Despite the maintenance of comodulation, the beating patterns between the OSB components and those of neighboring comodulated bands were now different from those between the components of the flanking bands themselves. The aim of this manipulation was to undermine the strength of the cue wherein the signal disrupts the uniform beating pattern of the comodulated masker complex. The modified masker was presented in continuous mode, and therefore the appropriate comparison is to the original comodulated complex also presented continuously (i.e., [COM/COM]). The average signal threshold masked by the original comodulated complex was 49.9 dB SPL, whereas the average threshold for the comodulated complex with the "mirror image" OSB was 58.4 dB SPL. This 8.5 dB threshold elevation when the beating cue was diminished provides further support for the contention that the linear spacings of experiment 2 facilitated the use of a beating disruption cue for signal detection under conditions of proximal spacing. In summary, the aim of accentuating the availability of temporal beating cues by using linear spacing in experiment 2 appears to have been successful.

The greater salience of within-channel processing cues in experiment 2 contributes to the interpretation of the comparative data patterns across the two experiments. For the four spectral spacings of experiments 1 and 2 where predominantly within-channel processing was expected (1/24th-octave, 1/6th-octave, 30 Hz, 120 Hz), a temporal asynchrony wherein the flanking bands were presented continuously and the OSB was gated had minimal effect on signal threshold. In contrast, for the spacings conducive to across-channel processing in the two experiments (1/3rd-octave, 1-octave, 240 Hz), this type of temporal asynchrony largely eliminated any masking release. Taken together, this pattern of results is consistent with the contention that within-channel processing is not sensitive to asynchrony among comodulated noise bands, whereas across-channel processing is. This, in turn, supports the hypothesis that across-channel contributions to auditory grouping can be disrupted by onset asynchrony.

However, the association between the onset asynchrony effect and a reliance on across-channel cues is not supported by results obtained with the alternative type of onset asynchrony, wherein the OSB was presented continuously and the flanking bands were gated. Here, irrespective of frequency spacing, thresholds were elevated to levels as high as, or higher than, the reference threshold. The elevation of threshold above the OSB reference threshold has been observed in previous work testing both adults (Grose and Hall, 1993) and children (Hall *et al.*, 1997) but is not uniformly found (Grose and Hall, 1996). It should be noted that in all these previous studies the signal onset was delayed relative to the masker onset by 100–200 ms; here, the signal onset was synchronous with the gated masker onset. As such, the increased masking bears some resemblance to the "transient masking" associated with the onset of a remote masker as measured by Bacon and Moore (1987), although that study pertained mainly to brief signals/maskers. One ramification of this resemblance to transient masking is that it suggests that the threshold elevation associated with the gated flanking bands might not be restricted to CFBs; that is, similar threshold elevations might be observed with gated random flanking bands. However, this condition [RAN/OSB] was not tested in this study. The observation that the elevation of threshold above the reference threshold tended to be largest for the 1/3rd-octave (logarithmic) and 240 Hz (linear) spacings might suggest that some form of across-channel masking contributes to the effect.

The temporal fringe manipulation had a generally uniform effect across the two experiments. When the fringe consisted of a set of random noise bands, or a set of CFBs plus an OSB with an independent fluctuation pattern (codeviant), masking release was largely eliminated irrespective of the frequency spacing of the noise bands. This general result runs counter to the suggestion that the disruptive effects of a noncomodulated temporal fringe are specific to across-channel cues and should therefore not affect signal threshold in cases where performance is based on within-channel cues. Note again that for the linear spacing of bands there were large individual differences at the narrowest spacing of the noncomodulated fringe, with two of the five observers being minimally affected by the fringe.

In order to gain insight into the overall pattern of results, it is useful to focus separately on possible within-channel and across-channel contributions to that pattern. Beginning first with within-channel contributions, recall that one hypothesis was that masking release related to within-channel cues would not be reduced by either the random temporal fringe or by onset asynchrony. The results of the narrow frequency spacing conditions were not entirely consistent with this hypothesis. Whereas the hypothesis was supported for the asynchrony condition where the flanking bands were continuous, it was not supported for the asynchrony condition where the OSB was continuous, nor for the condition where there was a random temporal fringe. We will briefly consider a possible account for the pattern of results obtained for the narrow spacing conditions, where performance is probably related to the processing of a within-channel cue. In this account, it is assumed that the listener monitors the regular beating pattern arising from the interaction among comodulated bands and bases detection on the change in beating pattern that occurs when the signal is added. In the asynchrony case where the CFBs are present continuously, the listener can monitor the pattern of the interactions among the CFBs, a pattern that is regular during both the no-signal intervals and the interstimulus intervals. This is consistent with the finding that performance was relatively good in the condition where the flanking bands were presented continuously and the OSB was gated on. Use of a beating cue is also consistent with the relatively *poor* performance in the asynchrony condition where the OSB was presented continuously, in that the stimulus surrounding the listening intervals contained no regular beating pattern due to the fact that a single noise band was present. The different effects of onset asynchrony as a function of masker spacing observed here could depend on the number of flanking bands. For example, if there were only two noise bands, one centered on the signal and one flanking band, then the different effects of gating

mode would not be expected to occur because beating between comodulated bands would not occur either for a continuous flanking band or a continuous OSB.

Use of a cue based on envelope beats is also consistent with the finding that performance with narrow frequency spacing was poor for the case where the fringe consisted of either random bands or CFBs plus an OSB with an independent fluctuation pattern (codeviant). In the case of the random fringe, the stimulus surrounding the listening intervals contained no consistent beating pattern due to the fact that the flanking bands had random envelopes. In the codeviant fringe condition, the OSB would diminish the regularity of the beating cue even in the absence of a pure tone signal. In both cases, stimulus features outside the listening interval fail to form a stream of consistent envelope beats.

Note that in the above accounts, it is assumed that the listener cannot simply base detection on the information within the observation intervals (where all bands are comodulated), due to some sort of "carry over" effect wherein the pattern of energy surrounding the listening intervals prevents the listener from optimally processing the beating pattern within the listening interval. Moreover, a listener's susceptibility to this carry over varies markedly as demonstrated by the extensive individual variability seen for the narrow 30 Hz linear spacing (see Fig. 5). One of the observers who was unaffected by the random and codeviant fringes in this configuration reported that the detection cue for the signal was a disruption in the regularity of the beat percept during the comodulated core interval, and that this cue remained salient even in the presence of a noncomodulated temporal fringe.

Turning now to across-channel contributions to the overall pattern of results, recall that part of the motivation for this study was based on the idea that auditory grouping is a precondition to across-channel CMR, and that the effects of onset asynchrony would be relatively great for masking release based on across-channel analysis. In contrast, the presence of random bands surrounding the observation intervals was thought to disrupt the listener's ability to make use of coherent modulation during the observation intervals independent of grouping. For the wider masker spacings, all manipulations tested elevated thresholds (except for the continuous comodulated condition). This effect could be due to reduced grouping strength in the case of onset asynchrony and some other factors in the temporal fringe conditions. However, these effects could also be attributed to some other common factor. One possibility is the regularity of spectrotemporal coherence. Here, the cue in across-channel CMR is characterized as a discontinuity in the dynamic spectrum of the stimulus. Addition of a pure tone signal can result in a spectral peak at the signal frequency when energy at other frequencies is at a modulation minimum, a cue that could be the basis of a detection process sometimes referred to as *listening in the dips*. It is possible that such an across-channel cue can be disrupted by spectrotemporal irregularities that occur outside the observation intervals. For example, random bands surrounding the observation intervals expose the listener to a variable spectrum, where peaks in the spectrum are associated with different frequency regions as a function of time. These peaks could impede the formation, or stability, of a no-signal template (in this case a flat spectrum that coherently fluctuates in level as a function of time). Similarly, asynchronous gating introduces radical changes in the spectrum of the masker that could interfere with cues based on more subtle spectral cues associated with addition of a pure tone signal.

The importance of spectrotemporal coherence for the use of an across-channel cue and the importance of envelope beat regularity as a function of time for the use of within-channel cues could be described in similar terms. Whereas spectrotemporal coherence may facilitate detection of a change in the spectrum, consistency of envelope statistics in a single auditory channel output could facilitate detection of a change in that temporal pattern. Features of a random stimulus surrounding the observation intervals, such as reduced modulation depth (within-channel) or spectral variability (across-channel), could resemble the stimulus features associated with addition of a pure tone signal. Asynchronous gating across bands could likewise interfere with formation of a no-signal baseline against which to compare the observation interval. It is therefore possible that performance based on either within- or across-channel cues is disrupted by manipulation of the spectrotemporal coherence of the stimulus surrounding the observation intervals for rather similar reasons.

The implication of these results for previous studies deserves comment. Dau *et al.* (2005) manipulated temporal asynchrony to differentiate within- and across-channel contributions to CMR. They hypothesized that the use of across-channel cues contributing to masking release for comodulated noise would be affected by across-frequency grouping mechanisms, but that within-channel mechanisms would not. In their experiment, they used onset asynchrony wherein the flanking bands were gated on before, and gated off after, the OSBs. The CMR declined under conditions of temporal asynchrony only for stimulus complexes made up of widely spaced noise bands, a result that supported the hypothesis. The results of the present investigation replicate this result, but indicate that support for the hypothesis should be tempered in light of the results obtained with the other mode of onset asynchrony. Using similar reasoning as Dau *et al.* (2005), Grose *et al.* (2005b) incorporated the manipulation of a random temporal fringe as a test for across-channel processing in CMR. They proposed that the presence of a random temporal fringe should reduce the likelihood of using available across-channel cues and should therefore diminish across-channel CMR. The results of the present investigation caution that sensitivity of signal threshold to this manipulation is not a specific indicator of across-channel processing, since a disruption to masking release due to within-channel processing also results from the presence of a noncomodulated temporal fringe.

V. SUMMARY AND CONCLUSIONS

The purpose of this investigation was to determine whether within- and across-channel contributions to masking release in the multiband CMR paradigm could be differentiated via manipulations of the stimulus characteristics in the

interstimulus interval. Two stimulus manipulations were examined: (1) onset asynchrony between the OSB and the CFBs, and (2) the presence of a random temporal fringe surrounding the core comodulated bands. The availability of within- and across-channel cues was manipulated by varying the frequency spacing of the narrow bands of noise making up the comodulated complex. Both logarithmic and linear spacings were employed. It was hypothesized that the two stimulus manipulations would disrupt masking release due to across-channel cues but not masking release due to within-channel cues.

The overall pattern of results did not wholly support the hypothesis. When the OSB was gated on during the observation intervals and the CFBs were presented continuously the results were compatible with the hypothesis: masking release was not affected for stimuli providing robust within-channel cues, but it was disrupted for stimuli providing robust across-channel cues. However, presenting the OSB continuously and gating the flanking bands elevated thresholds, irrespective of frequency spacing. The second stimulus manipulation (random temporal fringe) also generated results that did not support the hypothesis: masking release was largely eliminated by the presence of a random temporal fringe irrespective of the frequency spacing of the noise bands making up the masking complex.

In conclusion, this investigation has demonstrated that a detection advantage for a tone masked by a complex of comodulated narrow bands of noise can be observed under conditions of predominantly within- or cross-channel processing. Identifying the mechanisms underlying the masking release remains a challenge. Manipulation of stimulus features in the interstimulus interval that were previously thought to disrupt only across-channel processes has been shown to also disrupt within-channel processes under some conditions.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD R01-DC01507 and R01-DC007391. Helpful comments on previous drafts were provided by Brian Moore and two anonymous reviewers.

[1]For the band-widening method of measuring CMR, Verhey *et al.* (1999) showed that within-channel cues can provide the predominant cues for signal detection.

[2]MDI refers to the phenomenon wherein detection of modulation carried by one tone is disrupted by the presence of a spectrally remote tone that is modulated at a suprathreshold level (Yost and Sheft, 1989).

[3]CDD refers to the phenomenon wherein detection threshold for a narrow band of noise is typically higher in the presence of a complex of comodulated narrow bands of noise than a complex of noise bands that do not share the same fluctuation pattern as the target band (Cohen and Schubert, 1987; McFadden, 1987).

[4]Although Dau *et al.* (2005) did not test conditions where the OSB was gated on before, and gated off after, the CFBs, they extracted data for this condition from the Grose and Hall (1993) study. Thus, their identification of a discrepancy between their findings and those of Grose and Hall (1993) is due to a comparison of dissimilar conditions. Like Dau *et al.*, (2005) Grose and Hall (1993) did not observe a total elimination of masking release when flanker onset preceded on-signal masker onset.

[5]Pilot data indicated that there was no difference in masked threshold between gated and continuous random masking bands.

Bacon, S. P., and Moore, B. C. (**1987**). "Transient masking and the temporal course of simultaneous tone-on-tone masking," J. Acoust. Soc. Am. **81**, 1073–1077.

Berg, B. G. (**1996**). "On the relation between comodulation masking release and temporal modulation transfer functions," J. Acoust. Soc. Am. **100**, 1013–1023.

Cohen, M. F., and Schubert, E. D. (**1987**). "The effect of cross-spectrum correlation on the detectability of a noise band," J. Acoust. Soc. Am. **81**, 721–723.

Dau, T., Ewert, S. D., and Oxenham, A. J. (**2005**). "Effects of concurrent and sequential streaming in comodulation masking release," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer, New York).

Grose, J. H., and Hall, J. W., III. (**1996**). "Cochlear hearing loss and the processing of modulation: Effects of temporal asynchrony," J. Acoust. Soc. Am. **100**, 519–527.

Grose, J. H., and Hall, J. W., III. (**1993**). "Comodulation masking release: Is comodulation sufficient?," J. Acoust. Soc. Am. **93**, 2896–2902.

Grose, J. H., Hall, J. W., III, and Buss, E. (**2005a**). "Across-channel spectral processing," Int. Rev. Neurobiol. **70**, 87–119.

Grose, J. H., Hall, J. W., III, Buss, E., and Hatch, D. R. (**2005b**). "Detection of spectrally complex signals in comodulated maskers: Effect of temporal fringe," J. Acoust. Soc. Am. **118**, 3774–3782.

Hall, J. W., III, Buss, E., and Grose, J. H. (**2006**). "Comodulation detection differences for fixed-frequency and roved-frequency maskers," J. Acoust. Soc. Am. **119**, 1021–1028.

Hall, J. W., and Grose, J. H. (**1991**). "Some effects of auditory grouping factors on modulation detection interference (MDI)," J. Acoust. Soc. Am. **90**, 3028–3036.

Hall, J. W., Grose, J. H., and Dev, M. B. (**1997**). "Auditory development in complex tasks of comodulation masking release," J. Speech Lang. Hear. Res. **40**, 946–954.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectrotemporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Hatch, D. R., Arne, B. C., and Hall, J. W. (**1995**). "Comodulation masking release (CMR): Effects of gating as a function of number of flanking bands and masker bandwidth," J. Acoust. Soc. Am. **97**, 3768–3774.

McFadden, D. (**1987**). "Comodulation detection differences using noise-band signals," J. Acoust. Soc. Am. **81**, 1519–1527.

Mendoza, L., Hall, J. W., III, and Grose, J. H. (**1998**). "Comodulation masking release: The effect of the characteristics of noisebands presented before and after a signal," J. Acoust. Soc. Am. **103**, 2843.

Moore, B. C., and Borrill, S. J. (**2002**). "Tests of a within-channel account of comodulation detection differences," J. Acoust. Soc. Am. **112**, 2099–2109.

Moore, B. C., and Shailer, M. J. (**1992**). "Modulation discrimination interference and auditory grouping," Philos. Trans. R. Soc. London, Ser. B **336**, 339–346.

Oxenham, A. J., and Dau, T. (**2001**). "Modulation detection interference: Effects of concurrent and sequential streaming," J. Acoust. Soc. Am. **110**, 402–408.

Richards, V. M. (**1988**). "Components of monaural envelope correlation perception," Hear. Res. **35**, 47–58.

Schooneveldt, G. P., and Moore, B. C. J. (**1987**). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation flanking band," J. Acoust. Soc. Am. **82**, 1944–1956.

Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," J. Acoust. Soc. Am. **106**, 2733–2745.

Verhey, J. L., Pressnitzer, D., and Winter, I. M. (**2003**). "The psychophysics and physiology of comodulation masking release," Exp. Brain Res. **153**, 405–417.

Yost, W. A., and Sheft, S. (**1989**). "Across-critical-band processing of amplitude-modulated tones," J. Acoust. Soc. Am. **85**, 848–857.

The decision process in forward-masked intensity discrimination: Evidence from molecular analyses[a)]

Daniel Oberfeld[b)]
Department of Psychology, Johannes Gutenberg–Universität Mainz, 55099 Mainz, Germany

(Received 15 May 2007; revised 8 October 2008; accepted 9 October 2008)

In a two-interval forced-choice intensity discrimination task presenting a fixed increment, the level of the forward masker in interval 1 and interval 2 was sampled independently from the same normal distribution on each trial. Mean and standard deviation of the distribution were varied. Correlational analyses of the trial-by-trial data revealed different decision strategies depending on the relation between mean masker level and standard level. If the two levels were identical, listeners tended to select the interval containing the higher-level masker, behaving like an energy detector at the output of a temporal window of integration. For mean masker level higher than the standard level, most listeners showed a negative correlation between the masker level in a given interval and the probability of selecting this interval, indicating a strategy of comparing the masker loudness and the target loudness in each of the two observation intervals, and voting for the interval where the loudness difference was smaller. Implications for models of forward-masked intensity discrimination and differences from decision strategies reported for forward-masked detection tasks [Jesteadt *et al.*, (2005). "Effect of variability in level on forward masking and on increment detection," J. Acoust. Soc. Am. **118**, 325–337] are discussed.
 [DOI: 10.1121/1.3021296]

PACS number(s): 43.66.Fe, 43.66.Dc, 43.66.Ba, 43.66.Mk [RAL] Pages: 294–303

I. INTRODUCTION

Nonsimultaneous masking produces a rather complex pattern of effects on intensity resolution (e.g., Carlyon and Beveridge, 1993; Plack *et al.*, 1995; Zeng, 1998; Oberfeld, 2008b). The present study for the first time not only examined the effects of a forward masker on performance levels or intensity-difference limens (DLs) ("molar psychophysics;" see Green, 1964) but also assessed the decision process in a forward-masked intensity discrimination task by introducing within-trial variability in masker level and analyzing the trial-by-trial data ("molecular psychophysics;" see Green, 1964; Gilkey and Robinson, 1986; Berg, 2004). The results demonstrate different decision strategies for different combinations of masker level and standard level.

To summarize previous findings, an important result is that with an intense forward masker [e.g., 90 dB SPL (sound pressure level)], intensity DLs are strongly elevated for a midlevel standard, relative to the DL in quiet. On the other hand, there is only a small effect of the masker on the DLs for standards presented at low and high levels, resulting in the so-called *midlevel hump in intensity discrimination* (Zeng *et al.*, 1991).

Three explanations have been proposed for these effects (for an in-depth discussion see Oberfeld, 2008b). Zeng *et al.* (1991) suggested that the effect is due to adaptation of the (small) population of low spontaneous-rate (SR) auditory-nerve neurons showing slower recovery from prior stimulation than high-SR neurons (Relkin and Doucet, 1991). However, in subsequent experiments a midlevel hump was also found for backward maskers and contralaterally presented maskers (e.g., Plack and Viemeister, 1992; Plack *et al.*, 1995; Schlauch *et al.*, 1999), which precludes mechanisms in the auditory periphery as the origin of the effect.

The *referential encoding hypothesis* by Plack and Viemeister (1992) and Carlyon and Beveridge (1993) can account for the effects of backward maskers and contralaterally presented maskers. It assumes that the masker presented between the targets in a two-interval (2I) task degrades the memory trace of the target presented in the first observation interval, so that the listener is forced to switch to the "context-coding mode" (Durlach and Braida, 1969; Braida and Durlach, 1988), in which he or she remembers a categorical/verbal representation of the sensation, based on a comparison with internal or external references.[1] Referential encoding is assumed to work efficiently at low and high standard levels, where the detection threshold, the discomfort threshold, or the level of the intense forward masker can be used as a coding reference (Braida *et al.*, 1984; Carlyon and Beveridge, 1993). At intermediate standard levels, however, the perceptual distance to these references is large, and discrimination performance is thus predicted to be poor (Braida *et al.*, 1984). Consequently, for a midlevel standard, the model predicts a strong effect of an intense masker. For low-level and high-level standards, on the other hand, it predicts essentially no effect of the masker due to the assumed efficiency of referential encoding.

[a)] A portion of this work was presented at the 22nd Annual Meeting of the International Society for Psychophysics (Fechner Day), June 2006.
[b)] Electronic mail: oberfeld@uni-mainz.de

Thus, the referential encoding hypothesis can account for the midlevel hump. Yet, it cannot easily explain the reduced midlevel humps observed if a masker differing from the standard in duration or spectral content is presented (Schlauch *et al.*, 1997, 1999). Even more important, the referential encoding hypothesis predicts essentially no effect of the masker at low and high standard levels. However, Oberfeld (2008b) found significant masker-induced DL elevations at standard levels of 25 and 85 dB SPL. At the low standard level, a *mid-difference* hump was observed: the DL elevation was larger at intermediate than at large masker-standard level differences. The latter result and the effects of masker duration and spectrum indicate that the *perceptual similarity* between the masker and the standard modulates the effect of the masker. Such similarity effects can be integrated into a third explanation proposed for the midlevel hump, which is based on the finding that a forward masker has an effect not only on intensity resolution but also on loudness. Maskers higher in level than a temporally proximal target result in an increase in target loudness (loudness enhancement; see Zwislocki and Sokolich, 1974 and Oberfeld, 2007). Now, Carlyon and Beveridge (1993) suggested that the masker-induced change in target loudness introduces loudness variability, which in turn results in impaired performance in an intensity discrimination task (*loudness enhancement hypothesis*). In fact, both a midlevel hump (Zeng, 1994; Plack, 1996) and a mid-difference hump (Oberfeld, 2008b) have been reported for loudness enhancement. One possible explanation for loudness enhancement is that listeners cannot access the "isolated" loudness of the target but will instead use a weighted average of masker loudness and target loudness when making their decision ("mergence;" see Elmasian *et al.*, 1980 and Oberfeld, 2007). Oberfeld (2007, 2008b) proposed that the observed similarity effects can be incorporated into such a model by assuming that maskers strongly differing from the target in at least one dimension (e.g., level, duration, or spectrum) will receive only a relatively small weight. Thus, the third model considered in this paper is the combination of the loudness enhancement hypothesis and the *similarity hypothesis* proposed by Oberfeld (2008b).

The "molar" data (i.e., DLs) collected in previous studies are compatible with the latter model, but as Oberfeld (2008b) noted, the referential encoding hypothesis could also be extended by the assumption that a perceptual difference between masker and standard results in less memory trace interference, accounting for the observed similarity effects. The current study provides a "molecular" analysis of the data from an intensity discrimination task. The levels of the masker in interval 1 and the masker in interval 2 were independently and randomly perturbed on each trial. The influence of the masker intensity information on the decision of the listener was measured in terms of the correlation between the within-trial difference in masker level and the response *(increment in interval 1 or increment in interval 2)*.

Now the three different explanations for the effect of a forward masker on intensity resolution predict rather distinct patterns of correlations. For instance, a decision strategy compatible with the loudness enhancement hypothesis combined with the similarity hypothesis (Oberfeld, 2008b) would be that the listeners behave as an energy detector, comparing the output level of a "temporal window" for the first interval (representing a weighted sum of masker level and target level in that interval) with the output level of a temporal window for the second interval and responding that the increment was presented in the interval where the output level was higher (Plack and Oxenham, 1998).[2] These assumptions result in the hypothesis that the response be positively correlated with the within-trial difference in masker level. For example, listeners should tend to respond "*Increment in interval 2*" if the masker presented in interval 2 is higher in level than the masker in interval 1. The recovery-rate model and the referential encoding hypothesis predict different patterns of correlations, which will be detailed in Sec. IV.

The assumptions about the decision process in forward-masked intensity discrimination have not yet been tested using a molecular approach. Concerning intensity discrimination in quiet, Jesteadt *et al.* (2005) reanalyzed data from an experiment by Jesteadt *et al.* (2003) in which external variability was added by randomly varying the pedestal level in each of the two observation intervals. The relation between the level of the tone in the interval containing the standard only and in the interval containing the standard-plus-increment and performance was compatible with the pattern an energy detector would produce (see Green and Swets, 1966). Listeners voted for the interval containing the tone higher in level. For a forward-masked detection task with randomly perturbed masker levels, on the other hand, analysis of the trial-by-trial data demonstrated a decision strategy incompatible with energy detection at the output of a temporal window of integration (Plack and Oxenham, 1998; Nizami, 2003) because the subjects did not vote for the interval containing the higher-level masker.

It was hoped that an insight into the decision process could also provide an explanation for an additional aspect of the previous data currently not well understood, namely, the considerable intersubject variability, which is most notable for the case of an intense masker combined with a low-level standard. As pointed out above, most listeners show no or only a small DL elevation in this condition. In several studies, however, a large increase in the DL was observed for some listeners (Zeng *et al.*, 1991, listener RB; Zeng and Turner, 1992, listener RB; Carlyon and Beveridge, 1993, listener LW; Schlauch *et al.*, 1997, listener S4; Schlauch *et al.*, 1999, listener 2; Oberfeld, 2008b, listener BS). Could these differences be due to the use of different decision strategies?

II. METHOD

A 2I, two-alternative forced-choice (2AFC) intensity discrimination procedure was used. The levels of the standards remained constant at 25 dB SPL throughout the experiment. Listeners were tested in quiet and with forward maskers presented at mean levels of 25, 55, and 85 dB SPL. Three different values of the masker level variance were presented. Either the masker level was identical in the two observation intervals (i.e., the masker level was constant) or a small or large within-trial masker level variability was intro-

duced. The performance of an energy detector operating at the output of a temporal window of integration would decrease with the variability in masker level because the latter variability would contribute to the variance of the decision variable, which is assumed to be the difference in output level of the temporal window for the two observation intervals.

In order to make the task more similar to the "classical" 2I intensity discrimination task used in previous experiments, the level of the target tones (standard and standard-plus-increment) was not perturbed. In an adaptive procedure, for example, there are only two alternative target levels (standard or standard-plus-increment), standard level is constant, and increment level remains relatively stable toward the end of a run. To approximate these conditions, a fixed level increment was presented in each block. The task obviously still differed from a classical intensity discrimination task due to the within-trial variation in masker level. This level variation of the to-be-ignored masker seemed less likely to introduce a decision strategy other than in the classical task than if the target levels had been perturbed. Second, the presentation of only two fixed target levels (standard and standard-plus-increment) per condition made it possible to use a signal detection approach for calculating sensitivity.

A. Listeners

Four students at the Johannes Gutenberg–Universität Mainz participated in the experiment voluntarily (three females, one male, age 19–24 years). They either received partial course credit or were paid for their participation. All listeners reported normal hearing. For the right ear (the ear tested), detection thresholds measured by a 2I, 2AFC adaptive procedure with a 3-down, 1-up rule (Levitt, 1971) were better than 10 dB HL at all octave frequencies between 0.5 and 8 kHz. Listeners were naïve with respect to the hypotheses under test. Only listener KD had previous experience in an intensity discrimination task.

B. Stimuli and apparatus

The standard and the masker were 1-kHz pure tones with a steady-state duration of 20 ms, gated on and off with 5-ms cosine-squared ramps. Each sinusoid started at zero phase. On each trial, there were two observation intervals. Except in no-increment trials (see below), an increment—that is, a pure tone of the same frequency, duration, and temporal envelope—was added in-phase to the standard in one of the observation intervals (selected with an equal *a priori* probability). In the forward masking conditions, a masker was presented in both intervals. On each trial, the sound pressure level of the masker presented in interval 1 and of the masker presented in interval 2 was sampled independently from the same normal distribution. Mean masker level μ_M was 25, 55, or 85 dB SPL. The standard deviation (SD) was 0 (fixed masker level), 2, or 6 dB. Masker level was limited to a range of $\mu_M \pm 2.5$ SD.

The silent interval between masker offset and standard onset was 100 ms. The interval between the offset of the first target and the onset of the second target was 650 ms. A similar stimulus configuration has been used in previous experiments (e.g., see Plack *et al.*, 1995; Zeng, 1998; Oberfeld, 2008b).

A trial started with a visual attention signal. The targets (standard and standard-plus-increment) were also marked by visual signals. The intertrial interval was 2000 ms, with the restriction that the next trial never started before the response and the feedback to the preceding trial had been given.

The stimuli were generated digitally, played back via one channel of an RME ADI/S digital/analog converter (f_s = 44.1 kHz, 24-bit resolution), attenuated by a TDT PA5 programmable attenuator, buffered by a TDT HB7 headphone buffer, and presented to the right ear via Sennheiser HDA 200 circumaural headphones calibrated according to IEC 318 (1970). The attenuator setting remained constant within a trial. The experiment was conducted in a single-walled IAC sound-insulated chamber. Listeners were tested individually.

C. Procedure

The listeners participated in two training sessions, followed by three sessions in which intensity DLs were measured using an adaptive procedure. On the basis of these DLs, individual increments corresponding to 70%–85% correct were selected for each combination of mean masker level and masker level SD. The so-selected increments were used in the main experiment in which the increment was constant within each block.

1. Adaptive measurement of intensity-difference limens

Prior to the main experiment, intensity DLs were measured using a 2I, 2AFC adaptive procedure with a 3-down, 1-up tracking rule (Levitt, 1971). A level increment was added to the standard in one of the two observation intervals (selected randomly). Listeners were instructed to ignore the maskers. Visual trial-by-trial feedback was provided. The initial level of the in-phase intensity increment, $10 \log_{10}(\Delta I/I)$, was 8 dB. The step size was 5 dB until the fourth reversal and 2 dB for the remaining eight reversals. For each track, the arithmetic mean of $10 \log_{10}(\Delta I/I)$ at the eight final reversals was converted to $\Delta L_{\mathrm{DL}} = 10 \log_{10}[1 + \Delta I/I]$. A track was discarded if the SD of $10 \log_{10}(\Delta I/I)$ at the eight final reversals was greater than 5 dB. At least three tracks were obtained for each Mean Masker Level × SD combination. Time permitting, additional tracks were run if the SD of the DLs estimated in the first three tracks exceeded 5 dB.

2. Intensity discrimination task with a fixed level increment

In a 2I, 2AFC procedure, a level increment was added to the standard in one of the two observation intervals (selected randomly). The increment was fixed within each block. Listeners indicated the interval containing the louder target. They were instructed to ignore the maskers.

Based on the DLs obtained in the adaptive procedure, a level increment ΔL was selected individually for each Mean Masker Level × SD combination that would correspond to percent correct in the range from 70% to 85%. Across listen-

ers and conditions, ΔL ranged from 1.3 to 7.2 dB. Due to sizeable individual DL elevations caused by the 55-dB SPL and the 85-dB SPL maskers, it was not always possible to test a single intensity level increment resulting in the targeted performance level for all conditions. Therefore, for most listeners, the conditions differed not only in mean masker level and masker level variability, but also in increment level. This variation in the increment level presents a potential problem for the analyses of the trial-by-trial data because different correlation coefficients observed at, e.g., two different mean masker levels could be either due to the difference in masker level or to the difference in increment level. For example, Dai *et al.* (1996) demonstrated that the increment level has a pronounced effect on the expected and observed correlations in a spectral-shape discrimination task. To solve this problem, additional trials presenting the standard in both observation intervals (i.e., trials without an increment) were included in each block (Green, 1964; Dai *et al.*, 1996), except for the in-quiet condition. If the two targets are identical, any correlation between, e.g., the response and the difference between masker level in interval 2 and masker level in interval 1, can be attributed to the variation in masker level.

Only one Mean Masker Level × SD combination was presented in each block. Each block comprised 35 trials with the level increment presented in the first interval, 35 trials with the increment presented in the second interval, and 35 trials without an increment. Visual trial-by-trial feedback was provided, except following a no-increment trial. Additionally, the percentage of correct responses was displayed at the end of each block. Listeners were informed that they would receive both trials with feedback and trials without feedback, but they were not told that there would be no-increment trials without any difference in target level. Within each block, the three types of trials were presented in random order. For each combination of mean masker level and masker level SD, six blocks of 105 trials each were run in separate sessions, resulting in a total of 210 trials per condition (Mean Masker Level × SD × Increment Position). The main experiment comprised nine sessions. A testing session lasted approximately 1 h. Listeners took one short break in each session.

III. RESULTS AND DISCUSSION

A. Effect of mean masker level and masker variability on intensity resolution

The data from the intensity discrimination task using a fixed level increment were analyzed in terms of a signal detection theory model assuming equal-variance Gaussian distributions (Green and Swets, 1966). The no-increment trials were excluded from the analysis. For each observer, the sensitivity index applicable to a 2I forced-choice (2IFC) task, $d'_{2IFC}=(1/\sqrt{2})(z_{hit}-z_{FA})$,[3] was computed for each block.[4]

In 7 of the total of 240 blocks obtained across listeners and conditions, the proportion of hits was 1.0. As d' is not defined in this case, the "log-linear correction" for extreme proportions was used (see Goodman, 1970; Hautus, 1995; Jesteadt, 2005).

As the level increment was not constant across listeners and conditions, it was not possible to analyze the sensitivity in terms of d' directly. Instead, the level increment resulting in $d'=1.16$ was estimated for each block. This value of d' corresponds to the performance level targeted by a 3-down, 1-up adaptive procedure (79.4% correct; see Macmillan and Creelman, 2005), under the assumption of unbiased responding. In the first step, resolution-per-dB was computed as $\delta'=d'/\Delta L$ (Durlach and Braida, 1969), where ΔL is the size of the level increment presented in a given block. The estimate of the level increment corresponding to $d'=1.16$ was then computed as $\Delta L_{DL}=1.16/\delta'$, with the subscript DL denoting the correspondence to the DL in a 3-down, 1-up adaptive intensity discrimination task. Note that this analysis rests on the assumption that d' is proportional to the difference in level (ΔL) between standard-plus-increment and standard (Rabinowitz *et al.*, 1976; Buus and Florentine, 1991; Jesteadt *et al.*, 2003; for a discussion see Green, 1993).

In one block, d' was slightly negative so that no meaningful value of δ' could be computed. This block was excluded from the analysis. If for a given block the distance between the estimated ΔL_{DL} and the closer of the two quartiles of the six individual estimates obtained in the respective condition was larger than three interquartile ranges, this data point would be as an outlier (Lovie, 1986), resulting in the exclusion of four data points. Across the remaining 235 blocks, the percentage of correct responses ranged between 0.56 and 0.96 ($M=0.73$, SD=0.084). Mean d' was 0.95 (SD=0.40).

Figure 1 displays the individual estimates of the level increment ΔL_{DL} corresponding to $d'=1.16$, as a function of mean masker level and masker level variance. In the condition with fixed masker level (SD=0; open symbols in Fig. 1), the effect of the masker was small (maximum DL elevation of 2.6 dB), except for listener LE who produced a DL elevation of more than 7 dB at the two higher masker levels. For comparison, Oberfeld (2008b) using an adaptive procedure observed a mean DL elevation of 2.7 dB for a 55-dB SPL masker combined with a 25-dB SPL standard. For listeners KD and LE, the effect of the midlevel masker was slightly larger than the effect of the intense masker, representing weak evidence for a *mid-difference hump* pattern (Oberfeld, 2008b), which was also present in the mean data shown in Fig. 2. Listeners KD and LE showed a mid-difference hump in the presence of variability in masker level, while in these conditions ΔL_{DL} tended to increase monotonically with the masker level for listeners AZ and NH. The data were analyzed via repeated-measures analyses of variance (ANOVAs). The Huynh–Feldt correction for the degrees of freedom (Huynh and Feldt, 1976) was used where applicable and the value of $\tilde{\varepsilon}$ is reported. For the conditions with fixed masker level (SD=0), a one-factorial repeated-measures ANOVA indicated no significant effect of masker level [$F(3,9)=2.24$]. There was a significant quadratic trend, however, [$F(1,3)=10.26$, $p=0.049$], compatible with the observation of a mid-difference hump. For the data obtained under forward masking, the DLs tended to be larger if the masker levels were varied rather than fixed (Fig. 2). Such a pattern is compatible with energy detection at the output of a temporal window of integration. However, an ANOVA with the within-subject factors mean masker level (25, 55, and

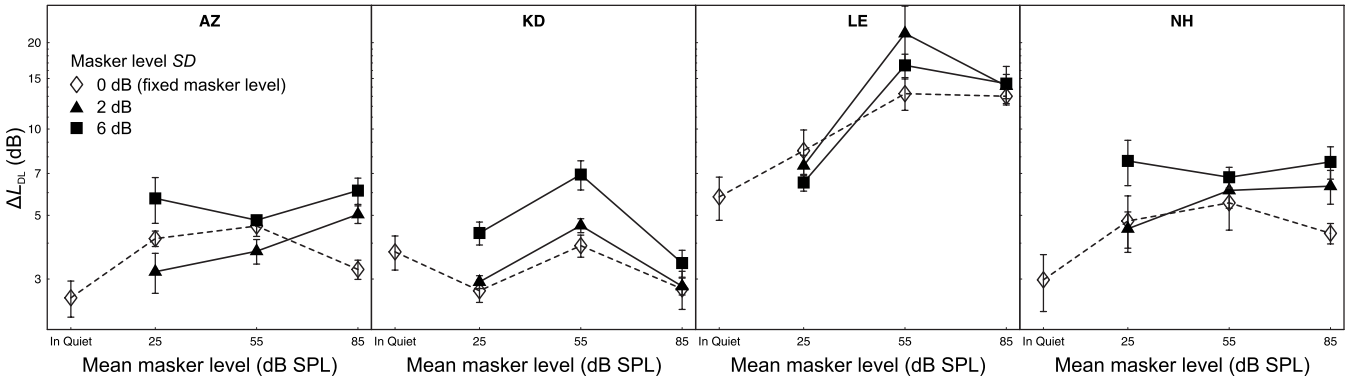


FIG. 1. Individual level increments (ΔL_{DL}, plotted on a log scale) corresponding to $d'=1.16$ as a function of mean masker level, estimated from resolution-per-dB in a 2IFC intensity discrimination task presenting a fixed level increment. The different symbols indicate different masker level SDs. Open diamonds: SD=0 dB (fixed masker level). Triangles: SD=2 dB. Boxes: SD=6 dB. Panels represent listeners. Error bars show ±1 standard error of the mean (SEM) of the six estimates per data point.

85 dB SPL) and masker level SD (0, 2, and 6 dB) showed neither a significant effect of mean masker level [$F(2,6)=1.64$] nor an effect of masker level SD [$F(2,6)=3.50$] nor a Mean Masker Level × SD interaction [$F(4,12)=0.94$].

B. Correlational analyses of the trial-by-trial data

A correlational approach was used for molecular analyses of the trial-by-trial data (see Richards and Zhu, 1994 and Lutfi, 1995; see also Berg, 1989).[5] In the first analysis, point-biserial correlations were computed between the within-trial difference between the masker levels in interval 2 and interval 1 ($L_{M2}-L_{M1}$) and the binary response of the listener (1 or 2, indicating that he or she detected the increment in interval 1 or interval 2, respectively). This made it possible to test the hypothesis that the listeners use energy detection at the output of a temporal window of integration, that is, tend to respond that the increment was presented in the interval in which the masker was higher in level.

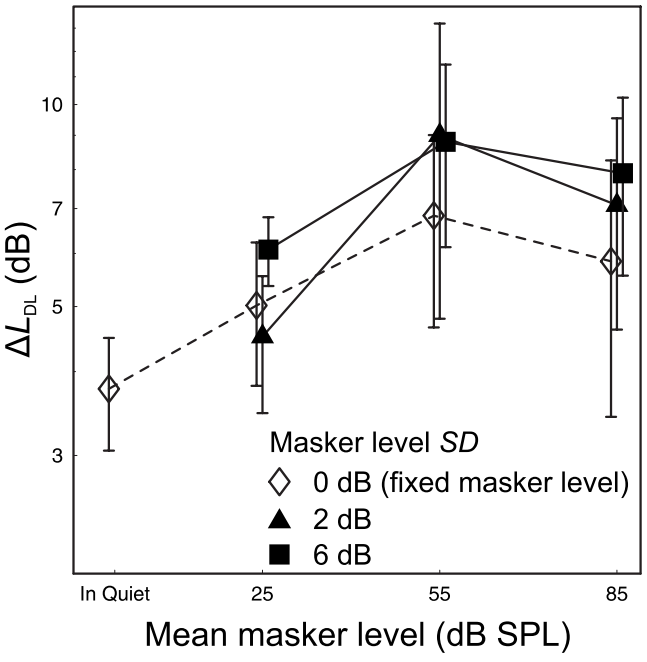


FIG. 2. Mean estimated level increments (ΔL_{DL}) corresponding to $d'=1.16$ as a function of mean masker level. Same symbols as in Fig. 1. Error bars show ±1 SEM of the four individual estimates.

In the second analysis, correlations of the correctness of the response (correct or incorrect) on individual trials with the masker level in the interval containing the standard and the level of the masker in the interval containing the standard-plus-increment were computed separately in order to answer the question whether the pattern Jesteadt *et al.* (2005) found for a forward-masked detection task also applies to a forward-masked intensity discrimination task, or if detection and discrimination show different characteristics in this regard.

1. Correlations between the within-trial difference in masker level and the response

Separate point-biserial correlations between the within-trial difference in masker level and the binary response were computed for the increment presented in interval 1, for the increment presented in interval 2, and for the no-increment trials.

An energy detector showing temporal integration will vote for the interval containing the higher overall level. As can be seen in Fig. 3, all listeners showed positive correlations if the mean masker level equaled the standard level (25 dB SPL), compatible with the behavior of an energy detector. All correlation coefficients were significantly different from 0 at the masker level SD condition of 6 dB (each coefficient based on 210 trials; *t*-test for correlation, $p<0.05$, two tailed; see Hays, 1988, p. 589). With a masker level SD of 2 dB, 8 of the 12 coefficients differed significantly from 0.

The data obtained in the conditions where the mean masker level was higher than the standard level indicated a decision strategy not compatible with energy detection, except for listener LE, who showed positive correlations between $L_{M2}-L_{M1}$ and the response at all mean masker levels. For the remaining listeners, 28 of the 36 correlation coefficients were significantly *smaller* than 0 at the 55-dB SPL and the 85-dB SPL mean masker levels (Fig. 3). A negative correlation between the difference in masker level and the response means that the listeners tended to vote for the interval

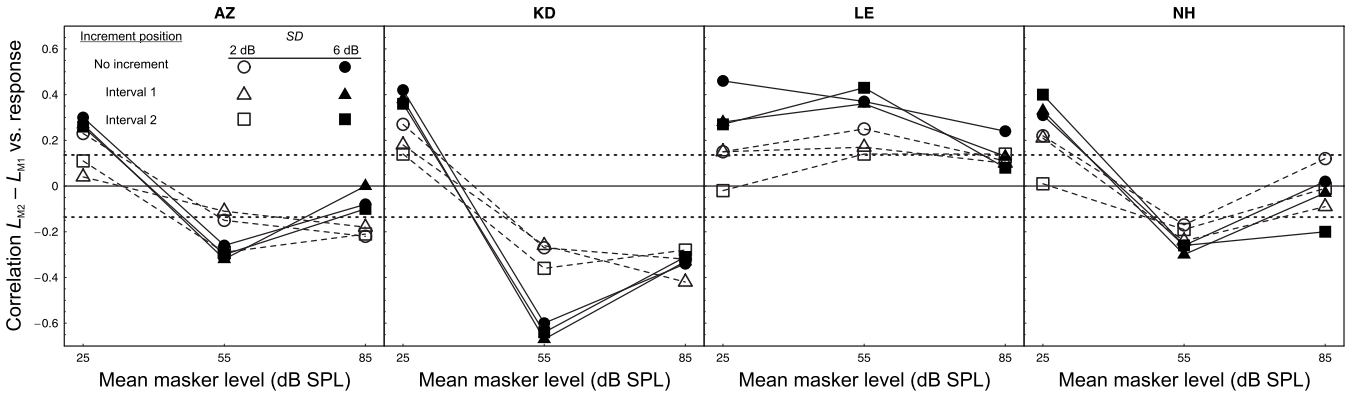


FIG. 3. Individual point-biserial correlations between the within-trial difference in masker level ($L_{M2}-L_{M1}$) and the response of the listener (1 or 2, indicating that he or she detected the increment in interval 1 or interval 2, respectively). Open symbols: SD=2 dB. Filled symbols: SD=6 dB. Circles: no increment. Triangles: increment in interval 1. Boxes: increment in interval 2. Panels represent listeners. Each data point is based on 210 trials. Correlation coefficients above or below the dotted horizontal lines are significantly different from zero (t-test for correlation, $p<0.05$, two tailed).

containing the *softer* masker. Such a behavior is compatible with a decision strategy of comparing the masker loudness and the target loudness in each of the two observation intervals and voting for the interval in which this difference in loudness was smaller.

A simple explanation for the zero correlations observed in several cases would be that the listener was effectively ignoring the maskers.

The correlations were stronger at a mean masker level of 55 dB SPL, especially so for the larger masker level SD. At the intermediate mean masker level, 23 of the 24 coefficients were significantly different from 0. At the largest mean masker level, only 12 of the 24 correlation coefficients differed significantly from 0. In principle, this pattern is consistent with the idea that a large loudness difference between masker and target should reduce the influence of the masker level on the decision (Oberfeld, 2008b), although the loudness enhancement hypothesis combined with the similarity hypothesis, of course, predicts positive rather than negative correlations.

At this point, a general cautionary note concerning the interpretation of the correlations as decision weights is in order. As Richards and Zhu (1994) showed, the correlation coefficients are identical in sign to the decision weight assigned to the parameter of interest (in the current case, the within-trial difference in masker level). Their magnitude, on the other hand, depends not only on the decision weight but also on parameters such as response bias, the (external) variability in masker level, and additional sources of variability in the decision variable (internal noise). Therefore, for a meaningful comparison of the magnitudes of the weight assigned to masker levels between, e.g., different mean masker levels, it would be necessary to analyze estimates of the decision weight rather than correlation coefficients. Unfortunately, the weight estimation would require rather strong assumptions about, for example, sources of internal noise and the weight assigned to target level. In part, these assumptions could be relaxed if the target levels were also randomly perturbed. In the current experiment, this was not done in order to make the task more similar to the classical intensity discrimination task. As a consequence, only the qualitative patterns of the correlations, most importantly the signs of the correlation coefficients, provide unequivocal information concerning differences in decision strategies. The *magnitude* of the correlations does not necessarily reflect the magnitude of the decision weights assigned to masker level information, however.

Mean data are displayed in Fig. 4. As discussed, the use of different level increments in the different Mean Masker Level × SD combinations poses a problem for the comparison of the correlations (see Dai *et al.*, 1996). Therefore, only the data from the common no-increment condition (circles in 4) were analyzed using a repeated-measures ANOVA with the factors mean masker level (25, 55, and 85 dB SPL) and masker level SD (2 and 6 dB). There was a significant effect of mean masker level [$F(2,6)=6.48$, $p=0.032$, $\tilde{\varepsilon}=1.0$], indicating different decision strategies for mean masker level

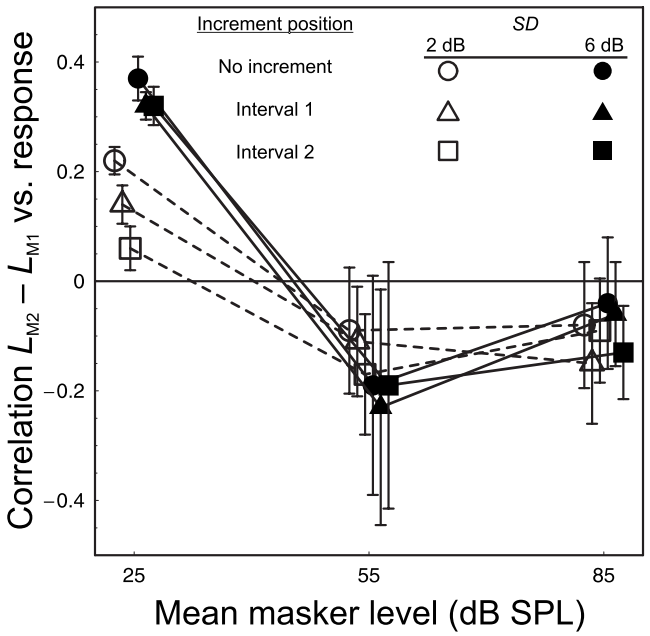


FIG. 4. Mean point-biserial correlations of the difference between masker level in interval 2 and masker level in interval 1 ($L_{M2}-L_{M1}$) with the response of the listener. Same format as Fig. 3. Error bars show ±1 SEM of the four individual correlation coefficients.

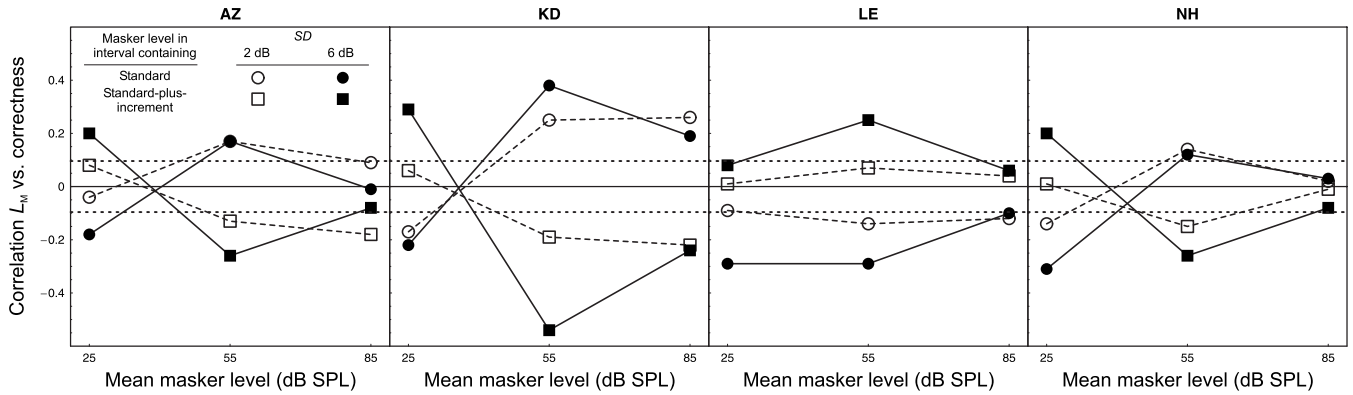


FIG. 5. Individual point-biserial correlations between the masker level and the correctness of the response as a function of interval and mean masker level. Circles: masker level in the interval containing the standard (M_S). Boxes: masker level in the interval containing the standard-plus-increment (M_{S+I}). Open symbols: masker level SD=2 dB. Filled symbols: SD=6 dB. Panels represent listeners. Each data point is based on 420 trials. Correlation coefficients above or below the dotted horizontal lines are significantly different from zero ($p<0.05$, two tailed).

equal to or greater than standard level, respectively. The observation of a stronger correlation at the larger masker level SD was confirmed by a significant Mean Masker Level $\times$SD interaction [$F(2,6)=6.53$, $p=0.031$, $\tilde{\varepsilon}=1.0$]. The main effect of masker level SD was not significant [$F(1,3)=0.275$], reflecting the fact that the correlation coefficients obtained at the two higher masker levels were more negative in the 6-dB SD condition for listeners AZ and KD, but more positive for listener LE.

2. Correlations between the masker level in the interval containing the standard or the standard-plus-increment and the correctness of the response

Correlations of the correctness of the response (correct or incorrect) on individual trials with the masker level in the interval containing the standard (M_S) and the masker level in the interval containing the standard-plus-increment (M_{S+I}) were computed separately. Jesteadt *et al.* (2005) reported that the level of the masker in the interval containing the signal was negatively related to performance, while the level of the masker in the nonsignal interval had only a very small effect on the correctness of the response. The performance of an energy detector at the output of a temporal window of integration would be equally affected by the masker levels in both intervals, albeit in opposite directions. A higher masker level in the interval containing the increment would increase the probability of voting for this interval, thereby increasing the probability of a correct response. A higher masker level in the interval containing the standard would also increase the probability of voting for this interval, but as a result *reduce* the probability of a correct response. For the current experiment, this raises the question whether the corresponding correlations differed as a function of mean masker level. As shown above, for mean masker level equal to standard level, the correlations between the within-trial difference in masker level and the response (increment in interval 1 or 2) were compatible with energy detection. Thus, it was expected that the masker level in both intervals be correlated with the correctness of the response in this condition. In contrast, for the higher masker levels, where a different decision strategy was observed, it seemed conceivable that the correctness was correlated mainly with the masker level in the interval containing the increment, which would parallel the results by Jesteadt *et al.* (2005).

An incorrect answer was coded as 0, a correct answer as 1. The no-increment trials were excluded from the analysis. The individual data are shown in Fig. 5. At a mean masker level of 25 dB SPL, all correlation coefficients for the relation between M_{S+I} and the correctness of the response were positive, although only three of the eight coefficients differed significantly from 0 (each coefficient was computed on the basis of 420 trials). In contrast, all correlation coefficients for the relation between M_S and the correctness of the response were negative (six of the total of eight were significantly different from zero). The positive correlations for the interval containing the increment and the negative correlations for the intervals containing the standard are compatible with energy detection.

For the two higher mean masker levels, the opposite relations were found for all listeners except LE (e.g., a *positive* correlation between M_S and the correctness of the response). For the interval containing the standard (S) and the interval containing the standard-plus-increment ($S+I$), 13 and 10, respectively, of the 16 correlation coefficients were significantly different from 0.

To test whether the strength of association between masker level and correctness differed between the standard interval and the standard-plus-increment interval, the *absolute values* of the correlation coefficients were analyzed in a repeated-measures ANOVA with the factors mean masker level (25, 55, and 85 dB SPL), SD (2 and 6 dB) and interval (interval containing the standard and interval containing the standard-plus-increment). Mean data are displayed in Fig. 6. There was no significant effect of interval [$F(1,3)=0.11$, $p=0.76$], showing that the correctness of the response was equally affected by the masker level in the interval containing the standard (S) as well as by the masker level in the interval containing the increment ($S+I$). The association between masker level and correctness was stronger at the larger masker level SD, confirmed by a significant effect of SD [$F(1,3)=22.63$, $p=0.018$]. There was also a significant Mean Masker Level$\times$SD interaction [$F(2,8)=8.51$, $p=0.026$, $\tilde{\varepsilon}=0.85$] because the difference between the two SDs was not present at the highest mean masker level. Mean

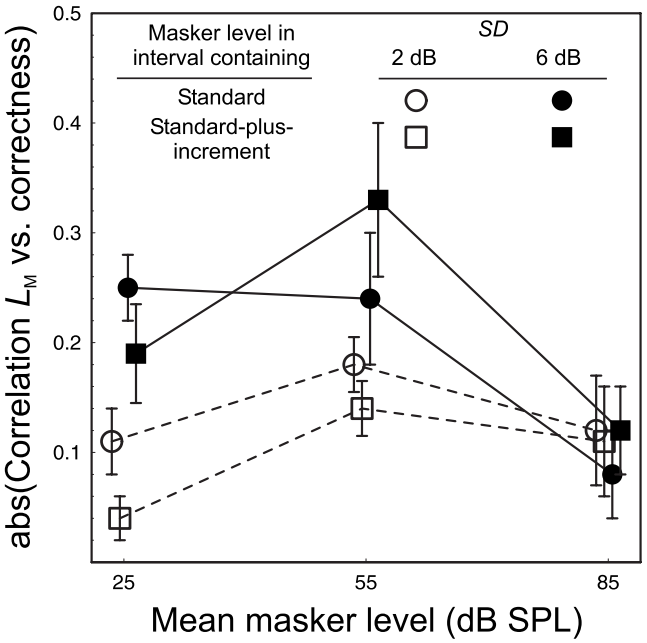


FIG. 6. Mean absolute values of the point-biserial correlation coefficients for the relation between masker level and the correctness of the response, as a function of interval and mean masker level. Same symbols as in Fig. 5. Error bars show ±1 SEM of the four individual values.

masker level had a significant effect, with the correlation being largest at the intermediate masker level [$F(2,6)=8.77$, $p=0.046$, $\tilde{\varepsilon}=0.59$]. The remaining effects were not significant at an alpha level of 0.05.

Taken together, the correlations between masker level in the two types of interval and the correctness of the response were incompatible with the behavior of an energy detector for conditions where mean masker level was higher than the standard level. Different from what Jesteadt *et al.* (2005) observed for a forward-masked detection task, the influence of masker level in the nonsignal interval (i.e., the interval containing the standard *S*) on the correctness of the response was not significantly smaller than the influence of masker level in the interval containing the increment. It can thus be concluded that different decision strategies are used in forward-masked detection and forward-masked intensity discrimination tasks.

IV. GENERAL DISCUSSION

The decision process in a 2I, 2AFC forward-masked intensity discrimination task was studied by introducing within-trial variability in masker level and analyzing the trial-by-trial data. Correlations between the within-trial difference in masker level and the response of the listener indicated the use of different decision strategies for different masker-standard level combinations. If the mean masker level equaled the standard level, the correlation was positive, compatible with a strategy of integrating the loudness of the masker and the target in each observation interval and voting for the interval in which the overall loudness was larger (i.e., energy detection at the output of a temporal window of integration). On the other hand, for three of the four listeners, the correlations were *negative* if the mean masker level was higher than the standard level. This result is compatible with a decision strategy of comparing the masker loudness and the target loudness in each of the two observation intervals and voting for the interval in which this difference in loudness was smaller. One of the listeners used the same strategy ("energy detection") at all mean masker levels. For this listener, the forward masker also had the strongest effect on intensity resolution. It remains to be shown whether and why using the "energy detection" rather than the "differencing" strategy could be related to this interindividual difference in sensitivity.

How do the results relate to the three explanations for the effects of a forward masker on intensity resolution discussed in the Introduction? If the effect of the masker was due to peripheral adaptation, as Zeng *et al.* (1991) assumed, then in principle negative correlations between the masker level presented in a given interval and the probability of selecting this interval should result. The reduction in the neural response (spike count) of auditory-nerve fibers to a test tone is proportional to the response to the masker (Smith, 1977; Abbas and Gorga, 1981). Therefore, the masker-induced reduction in the neural response to the target should be weaker in the interval containing the masker lower in level, introducing a tendency to vote for the interval containing the softer masker. On the other hand, as the output from high-SR fibers saturates at moderate sound pressure levels, a variation of ±5 dB in the level of a forward masker with a mean level of 85 dB SPL (as in the SD=2 dB condition of the current experiment) should have virtually no effect on the neural response to the test tone. However, based on a model of auditory-nerve responses that explicitly takes into account the differences between high-SR and low-SR neurons (Sumner *et al.*, 2002), Meddis and O'Mard (2005) concluded that for low-SR neurons, the amount of depression in the response to the test tone does not saturate even at masker levels of 90 dB SPL. In any case, the *positive* correlations between the within-trial level difference and the response observed for mean masker level equal to standard level are at odds with the physiological characteristics of the auditory nerve, where maskers identical in level to the test tone produce adaptation rather than enhancement (Smith, 1977).

Negative correlations are in principle also compatible with the referential encoding hypothesis, but only if it is assumed that the listeners use the intense masker as the coding reference. If now the intensity of the targets is coded by "measuring" the distance to the reference intensity (Braida *et al.*, 1984), then the distance will be larger in the interval containing the masker higher in level, so that the target presented in this interval will be coded as being softer. It appears very unlikely, though, that the listeners used the masker as a coding reference. As the level of the standard was fixed to 25 dB SPL throughout the experiment, this was the lowest level in all blocks presenting the 55 and the 85-dB SPL mean masker level, thus representing a sharp gradient in the stimulus set. This should have made the standard level useful as a coding reference according to the model by Braida *et al.* (1984), specifically so as the perceptual distance between the targets and the standard was much smaller than the distance between the targets and the intense masker. Second, to explain the different correlations observed for mean

masker level equal to standard level, it could be assumed that the masker causes no trace degradation. On the other hand, only zero correlations would have been directly compatible with the referential encoding hypothesis in this situation, while the observed positive correlations cannot be explained in terms of the model.

The loudness enhancement hypothesis (Carlyon and Beveridge, 1993) combined with the idea that loudness enhancement occurs because the percept of target loudness is the weighted average of the separate sensation masker and target would produce if presented in isolation (mergence; see Elmasian *et al.*, 1980; Oberfeld, 2008b) predicts positive correlations between the within-trial level difference and the response. Compatible results were found only for mean masker level equal to standard level, however, while the negative correlations at higher masker levels are evidence against this hypothesis. The data show that at least in the latter conditions the listeners do not behave as an energy detector at the output of a temporal window of integration.

Taken together, it has to be concluded that none of the three explanations for the effects of a nonsimultaneous masker on intensity resolution is strictly compatible with the results of the present experiment. Instead, the decision strategies identified via molecular analyses should motivate the formulation of alternative models. To this end, it would, of course, be desirable to apply the molecular approach to a wider variety of conditions than in the current experiment. Generally, the findings indicate that it is not appropriate to assume that one and the same decision strategy is used in all conditions and by all listeners.

In order to gain meaningful information about not only the direction of the influence of masker level on the decision but also about the strength of the association between the within-trial difference in masker level and the response, it would be necessary to estimate decision weights rather than to analyze the correlations (see Richards and Zhu, 1994). In this context, a sensible step for future experiments would be to also perturb the target levels, which could simplify the derivation of the weights. The magnitudes of the decision weights in different conditions (e.g., mean masker levels) could provide further information about, for example, the relation to inter-and intraindividual differences in sensitivity.

Correlations of the correctness of the response on individual trials with either the masker level in the interval containing the standard or the masker level in the interval containing the standard-plus-increment showed that the masker level in both types of interval had an effect on performance, contrary to what Jesteadt *et al.* (2005) observed for a forward-masked detection task. The data are thus further evidence for the characteristics of discrimination and detection differing from one another (see Laming, 1986; Zeng *et al.*, 1991).

ACKNOWLEDGMENTS

I express my appreciation to Walt Jesteadt for thoughtful comments on an earlier draft of this manuscript. Two anonymous reviewers provided helpful suggestions concerning previous versions.

[1]If trace degradation prevents across-interval comparisons, then the 2I task is effectively transformed into a one-interval (1I) task. Sensitivity (in terms of d') in a 1I task is approximately two times lower than in a 2I task [e.g., see Viemeister (1970), Jesteadt and Bilger (1974), and Braida *et al.* (1984)]. As there are two observations per trial (one per interval), however, sensitivity should decrease only by a factor of $\sqrt{2}$ if across-interval comparisons are prevented [see Green and Swets (1966), p. 238ff]. If the masker-induced decline in intensity resolution was only due to the latter decrease in sensitivity predicted by signal detection theory, a forward masker should have no effect in a 1I task, however, where the same decision mode would be assumed in quiet and in forward masking. Yet, Oberfeld (2006) found that the effect of a forward masker on intensity resolution in a 1I task was very similar to the effect in a 2I task. Thus, it is necessary to consider additional factors such as an increased variance in the coding process introduced by the change in decision mode [see Braida *et al.* (1984) and Plack and Viemeister (1992)].

[2]The temporal window effective here must be considerably longer than the window with a duration of approximately 40 ms assumed for a detection task (Plack and Oxenham, 1998) because an effect of a forward masker on intensity resolution is observed at masker-target ISIs up to 400 ms (Zeng and Turner, 1992). Therefore, it is not likely that the temporal integration operates at early processing stages. Instead, it is conceivable that the integration of masker and target level information represents properties of the decision process operating on memory representations of the targets.

[3]For d' derived from a one-point receiver operating characteristic (ROC) curve to be a valid measure of sensitivity, it must be assumed that the ROC curve is symmetric (i.e., linear and of unity slope in z-coordinates; see Macmillan and Creelman, 2005). As Green and Swets (1966) (p. 68) have shown on theoretical grounds, the ROC curve for a 2IFC experiment can be expected to be symmetric about the negative diagonal, even if the yes-no ROC curve for the same condition is asymmetric. This prediction was supported empirically (see Norman,1964; Atkinson and Kinchla, 1965; Markowitz and Swets, 1967; see also Luce and Green, 1974).

[4]Alternatively, the frequencies of hits and false alarms could have been pooled across all blocks presenting the same condition. Hautus (1997) showed that computing d' from pooled data can be more efficient than computing average d' across blocks, albeit not in all situations (for a discussion see Macmillan and Creelman, 2005). It is likely that for a given condition the sensitivity of a listener varies from session to session. In fact, Jesteadt (2005) showed that the variability of repeated measurements of d' is larger than the variance associated with the binomial distributions of the proportions underlying the computation of d', indicating additional sources of variance. Computing d' per block and presenting the average sensitivity and the associated SD provide information about the block-to-block variability that would be concealed when computing d' from pooled data.

[5]The techniques used to estimate perceptual weights range from maximum-likelihood estimation of the weights based on the slopes of psychometric functions (Berg, 1989) to point-biserial or Pearson correlations (Ahumada and Lovell, 1971; Lutfi, 1995) and multiple logistic regression (Alexander and Lutfi, 2004; Oberfeld, 2008a). All techniques are based on a similar decision model (for a recent discussion see Lutfi and Jesteadt, 2006). Multiple logistic regression has the advantage of controlling for (spurious) correlations between multiple predictors. In the present experiment, however, only one predictor (the within-trial difference in masker level) was involved. Therefore, the correlational approach was selected because it is computationally simple and its properties have been thoroughly analyzed by Richards and Zhu (1994). Plank (2005) and Tang *et al.* (2005) demonstrated that the weight estimates produced by the different techniques are very similar.

Abbas, P. J., and Gorga, M. P. (**1981**). "AP responses in forward-masking paradigms and their relationship to responses of auditory-nerve fibers," J. Acoust. Soc. Am. **69**, 492–499.

Ahumada, A., and Lovell, J. (**1971**). "Stimulus features in signal detection," J. Acoust. Soc. Am. **49**, 1751–1756.

Alexander, J. M., and Lutfi, R. A. (**2004**). "Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights," J. Acoust. Soc. Am. **116**, 2234–2247.

Atkinson, R. C., and Kinchla, R. A. (**1965**). "A learning-model for forced-choice detection experiments," Br. J. Math. Stat. Psychol. **18**, 183–206.

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

Berg, B. G. (**2004**). "A molecular description of profile analysis: Decision

weights and internal noise," J. Acoust. Soc. Am. **115**, 822–829.
Braida, L. D., and Durlach, N. I. (**1988**). "Peripheral and central factors in intensity perception," in *Auditory Function*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 559–583.
Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context coding," J. Acoust. Soc. Am. **76**, 722–731.
Buus, S., and Florentine, M. (**1991**). "Psychometric functions for level discrimination," J. Acoust. Soc. Am. **90**, 1371–1380.
Carlyon, R. P., and Beveridge, H. A. (**1993**). "Effects of forward masking on intensity discrimination, frequency discrimination, and the detection of tones in noise," J. Acoust. Soc. Am. **93**, 2886–2895.
Dai, H., Nguyen, Q., and Green, D. M. (**1996**). "Decision rules of listeners in spectral-shape discrimination with or without signal-frequency uncertainty," J. Acoust. Soc. Am. **99**, 2298–2306.
Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.
Elmasian, R., Galambos, R., and Bernheim, A. (**1980**). "Loudness enhancement and decrement in four paradigms," J. Acoust. Soc. Am. **67**, 601–607.
Gilkey, R. H., and Robinson, D. E. (**1986**). "Models of auditory masking: A molecular psychophysical approach," J. Acoust. Soc. Am. **79**, 1499–1510.
Goodman, L. A. (**1970**). "The multivariate analysis of qualitative data: Interactions among multiple classifications," J. Am. Stat. Assoc. **65**, 226–256.
Green, D. M. (**1964**). "Consistency of auditory detection judgments," Psychol. Rev. **71**, 392–407.
Green, D. M. (**1993**). "Auditory intensity discrimination," in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 13–55.
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Hautus, M. J. (**1995**). "Corrections for extreme proportions and their biasing effects on estimated values of d'," Behav. Res. Methods Instrum. Comput. **27**, 46–51.
Hautus, M. J. (**1997**). "Calculating estimates of sensitivity from group data: Pooled versus averaged estimators," Behav. Res. Methods Instrum. Comput. **29**, 556–562.
Hays, W. L. (**1988**). *Statistics*, 4th ed. (Holt, Rinehart and Winston, Fort Worth, TX).
Huynh, H., and Feldt, L. S. (**1976**). "Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs," J. Educ. Stat. **1**, 69–82.
IEC 318 (**1970**). *An IEC Artificial Ear, of the Wide Band Type, for the Calibration of Earphones Used in Audiometry* (International Electrotechnical Commission, Geneva).
Jesteadt, W. (**2005**). "The variance of d' estimates obtained in yes-no and two-interval forced choice procedures," Percept. Psychophys. **67**, 72–80.
Jesteadt, W., and Bilger, R. C. (**1974**). "Intensity and frequency discrimination in one- and two-interval paradigms," J. Acoust. Soc. Am. **55**, 1266–1276.
Jesteadt, W., Nizami, L., and Schairer, K. S. (**2003**). "A measure of internal noise based on sample discrimination," J. Acoust. Soc. Am. **114**, 2127–2157.
Jesteadt, W., Schairer, K. S., and Neff, D. L. (**2005**). "Effect of variability in level on forward masking and on increment detection," J. Acoust. Soc. Am. **118**, 325–337.
Laming, D. (**1986**). *Sensory Analysis* (Academic, London).
Levitt, H. (**1971**). "Transformed up-down procedures in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Lovie, P. (**1986**). "Identifying outliers," in *New Developments in Statistics for Psychology and the Social Sciences*, edited by A. D. Lovie(The British Psychological Society, Leicester, UK), pp. 45–69.
Luce, R. D., and Green, D. M. (**1974**). "Detection, discrimination, and recognition," in *Handbook of Perception, Volume II. Psychophysical Judgment and Measurement*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), pp. 299–342.
Lutfi, R. A. (**1995**). "Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks," J. Acoust. Soc. Am. **97**, 1333–1334.
Lutfi, R. A., and Jesteadt, W. (**2006**). "Molecular analysis of the effect of relative tone level on multitone pattern discrimination," J. Acoust. Soc. Am. **120**, 3853–3860.
Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide*, 2nd ed. (Lawrence Erlbaum, Mahwah).
Markowitz, J., and Swets, J. A. (**1967**). "Factors affecting the slope of empirical ROC curves: Comparison of binary and rating responses," Percept. Psychophys. **2**, 91–100.
Meddis, R., and O'Mard, L. P. (**2005**). "A computer model of the auditory-nerve response to forward-masking stimuli," J. Acoust. Soc. Am. **117**, 3787–3798.
Nizami, L. (**2003**). "Afferent response parameters derived from postmasker probe-detection thresholds: 'The decay of sensation' revisited," Hear. Res. **175**, 14–35.
Norman, D. A. (**1964**). "Sensory thresholds, response biases, and the neural quantum-theory," J. Math. Psychol. **1**, 88–120.
Oberfeld, D. (**2006**). "Forward-masked intensity discrimination: Evidence from one-interval and two-interval tasks," in *Fortschritte der Akustik—DAGA '06*, edited by Deutsche Gesellschaft für Akustik, DEGA, Berlin, pp. 309–310.
Oberfeld, D. (**2007**). "Loudness changes induced by a proximal sound: Loudness enhancement, loudness recalibration, or both?," J. Acoust. Soc. Am. **121**, 2137–2148.
Oberfeld, D. (**2008a**). "Does a rhythmic context have an effect on perceptual weights in auditory intensity processing?," Can. J. Exp. Psychol. **62**, 24–32.
Oberfeld, D. (**2008b**). "The mid-difference hump in forward-masked intensity discrimination," J. Acoust. Soc. Am. **123**, 1571–1581.
Plack, C. J. (**1996**). "Loudness enhancement and intensity discrimination under forward and backward masking," J. Acoust. Soc. Am. **100**, 1024–1030.
Plack, C. J., Carlyon, R. P., and Viemeister, N. F. (**1995**). "Intensity discrimination under forward and backward masking: Role of referential encoding," J. Acoust. Soc. Am. **97**, 1141–1149.
Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**, 1598–1608.
Plack, C. J., and Viemeister, N. F. (**1992**). "Intensity discrimination under backward masking," J. Acoust. Soc. Am. **92**, 3097–3101.
Plank, T. (**2005**). "Auditive Unterscheidung von zeitlichen Lautheitsprofilen (Auditory discrimination of temporal loudness profiles)," Ph.D. dissertation, Universität Regensburg, Germany.
Rabinowitz, W. M., Lim, J. S., Braida, L. D., and Durlach, N. I. (**1976**). "Intensity preception [sic]. VI. Summary of recent data on deviations from Weber's law for 1000-Hz tone pulses," J. Acoust. Soc. Am. **59**, 1506–1509.
Relkin, E. M., and Doucet, J. R. (**1991**). "Recovery from prior stimulation. I. Relationship to spontaneous firing rates of primary auditory neurons," Hear. Res. **55**, 215–222.
Richards, V. M., and Zhu, S. (**1994**). "Relative estimation of combination weights, decision criteria, and internal noise based on correlation coefficients," J. Acoust. Soc. Am. **95**, 423–434.
Schlauch, R. S., Clement, B. R., Ries, D. T., and DiGiovanni, J. J. (**1999**). "Masker laterality and cueing in forward-masked intensity discrimination," J. Acoust. Soc. Am. **105**, 822–828.
Schlauch, R. S., Lanthier, N., and Neve, J. (**1997**). "Forward-masked intensity discrimination: Duration effects and spectral effects," J. Acoust. Soc. Am. **102**, 461–467.
Smith, R. L. (**1977**). "Short-term adaptation in single auditory nerve fibers: Some poststimulatory effects," J. Neurophysiol. **40**, 1098–1112.
Sumner, C., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2002**). "A revised model of the inner-hair cell and auditory-nerve complex," J. Acoust. Soc. Am. **111**, 2178–2188.
Tang, Z., Richards, V. M., and Shih, A. (**2005**). "Comparing linear regression models applied to psychophysical data (A)," J. Acoust. Soc. Am. **117**, 2597.
Viemeister, N. F. (**1970**). "Intensity discrimination: Performance in three paradigms," Percept. Psychophys. **8**, 417–419.
Zeng, F.-G. (**1994**). "Loudness growth in forward masking: Relation to intensity discrimination," J. Acoust. Soc. Am. **96**, 2127–2132.
Zeng, F.-G. (**1998**). "Interactions of forward and simultaneous masking in intensity discrimination," J. Acoust. Soc. Am. **103**, 2021–2030.
Zeng, F.-G., and Turner, C. W. (**1992**). "Intensity discrimination in forward masking," J. Acoust. Soc. Am. **92**, 782–787.
Zeng, F.-G., Turner, C. W., and Relkin, E. M. (**1991**). "Recovery from prior stimulation II: Effects upon intensity discrimination," Hear. Res. **55**, 223–230.
Zwislocki, J. J., and Sokolich, W. G. (**1974**). "On loudness enhancement of a tone burst by a preceding tone burst," Percept. Psychophys. **16**, 87–90.

Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans

Astrid Klinge[a] and Georg M. Klump
Animal Physiology and Behavior Group, Department of Biology and Environmental Sciences, Carl-von-Ossietzky University Oldenburg, Postfach 2503, 26111 Oldenburg, Germany

(Received 7 March 2008; revised 10 October 2008; accepted 16 October 2008)

Frequency difference limens (FDLs) for pure tones between 200 and 6400 Hz and for the first, the second, or the eighth harmonic of an 800 Hz complex in four Mongolian gerbils (*Meriones unguiculatus*) were determined using a Go/NoGo-procedure. The 12 harmonics of the complex started either in sine phase or at a random phase. Gerbils showed very high pure tone FDLs ranging from 17.1% Weber fraction (200 Hz) to 6.7% (6400 Hz). They performed much better in detecting mistuning of a harmonic in the complex in the sine phase condition with FDLs decreasing from 0.07% for the first harmonic to 0.02% for the eighth harmonic. FDLs were about one order of magnitude higher when temporal cues were degraded by randomizing the starting phase of every component in the harmonic complex for every stimulus. These results are strikingly different from those obtained in four human subjects who needed about four times higher frequency shifts than gerbils for detecting a mistuned component in a sine phase complex and showed similar detection of mistuning in the random phase condition. The results are discussed in relation to possible processing mechanisms for pure tone frequency discrimination and for detecting mistuning in harmonic complex stimuli. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3021315]

PACS number(s): 43.66.Fe, 43.66.Gf, 43.66.Nm, 43.66.Hg [WPS] Pages: 304–314

I. INTRODUCTION

The natural environment is acoustically complex. The auditory system receives acoustic information from many different sound sources that often overlap in time and in frequency. The auditory system has the capability to segregate these sound sources and to group together sounds that belong to one source. Many of the natural sounds are not simple pure tones but are harmonic complexes composed of tonal components with a frequency that is an integer multiple of a fundamental frequency f_0. Such harmonic complexes are produced by the human voice, musical instruments, and animal calls, but can even be generated incidentally by mechanical action of water or wind. The harmonic complex is generally perceived as one sound with a pitch corresponding to the fundamental frequency. Thus, harmonicity is a strong cue to group together frequencies aiding the segregation of sounds from different sources and the grouping of sounds from a single source to a perceived auditory object.

If a component of a harmonic complex is shifted in frequency, the percept of the harmonic complex changes for small frequency shifts until the mistuned component falls out of the harmonic series producing a separate auditory object at a large frequency shift. The "mistuned" harmonic paradigm is one way of studying the influence of harmonicity on the segregation of concurrent auditory objects (e.g., Moore *et al.* 1984, 1985, 1986; Lin and Hartmann, 1998). In such an experiment, a harmonic complex serves as a reference stimulus and in the inharmonic test stimulus one of the components of the harmonic complex is shifted in frequency. Moore *et al.* (1985) reported that the cues for detecting the inharmonicity differed for the lower and higher harmonics. Mistuned harmonics of low frequencies that are resolved by the auditory system "stand out" as a separate pure tone whereas harmonics with high unresolved frequencies are heard as a "beating" of the stimulus. These results indicate that in a harmonic complex with a mistuned component not only changes in the spectral domain but also changes in the temporal structure can be used to detect inharmonicity.

In this paper, we evaluate the perception of mistuned components in an otherwise harmonic complex in the Mongolian gerbil and compare these results to human data. Mongolian gerbils are a widely used animal model for studying auditory processing mechanisms in mammals since their hearing sensitivity in the low-frequency region is similar to that of humans. We measured the gerbil's ability to detect a frequency shift in three behavioral experiments. In the first experiment, we obtained frequency difference limens (FDLs) for pure tones at different frequencies that were used as a reference for the two subsequent experiments. In the second and third experiment, FDLs for mistuned components of a harmonic complex stimulus were measured. The harmonic complex in the second experiment consisted of components that all started in sine phase. This resulted in a periodic temporal pattern that may provide a reference for detecting a change caused by the mistuning of a harmonic. If gerbils are especially sensitive to such temporal cues, we expect them to perform much better in detecting a frequency shift leading to the inharmonicity in a complex stimulus than in frequency discrimination of pure tones as has been found in birds (e.g., Lohr and Dooling, 1998). By randomizing the phase of every component in the harmonic complex in every presented stimulus during the third experiment, we provided a refer-

[a]Author to whom correspondence should be addressed. Electronic mail: astrid.klinge@uni-oldenburg.de

ence with a constant spectral composition but a varying temporal structure. We thereby made a change in the temporal pattern less useful as a cue. The comparison of the thresholds for detecting a frequency shift in the three conditions will provide insight into the mechanisms of processing of harmonic complexes that affects the formation of auditory objects.

II. MATERIALS AND METHODS

A. Animal subjects

Four adult Mongolian gerbils (*Meriones unguiculatus*)—three females, one male—were trained to serve as subjects in the experiments. For one of the females, we only obtained hearing thresholds and FDLs of mistuned harmonics in a complex with all harmonics starting in sine phase. During the time of testing, the animals were between 15 and 27 months old. All four gerbils had normal hearing within the frequency range of 100 Hz–8 kHz (see below for further explanation). Animals were housed in individual cages (Type III, $42\times26\times15$ cm^3) and had unlimited access to water. The daily amount of food was restricted to maintain a weight of approximately 80% of their *ad libitum* weight.

The care and treatment of the gerbils were in accordance with the procedures of animal experimentation approved by the Government of Lower Saxony, Germany. All procedures were performed in compliance with the NIH Guide on Methods and Welfare Considerations in Behavioural Research with Animals (NIH Publication No. 02-5083).

B. Apparatus and stimulus generation

Experiments were conducted in a single-walled sound attenuating booth (Industrial Acoustics, Type IAC 401-A). The interior of the chamber was lined with a 15 cm thick layer of sound absorbing foam with a pyramidal surface (Illbruck Illtec Pyramide 100/50, Illbruck Illtec PLANO Type 50/0). The reverberation time T_{60} of a broadband white noise measured through a microphone placed at the location where the head of the gerbil would normally be during the experiment was 12 ms and thus can still be referred to as anechoic.

The experimental cage was mounted on a rack in the middle of the chamber. The rack was built out of 5 mm thin metal bars that produced no substantial reflections. The cage was doughnut shaped and consisted of wire mesh (mesh size of 1 cm) on all sides. The floor of the cage was covered with fine wire mesh (mesh size of 1 mm) to provide for an even surface. Inside the cage, there was an elevated platform where the gerbils had to sit and wait for the test stimulus being played. Two light barriers were installed, one monitoring the pedestal and one the access to it. By forcing the gerbil to interrupt the light barriers in a specific sequence, we ensured that test stimuli were only presented if the gerbil was facing the loudspeaker. The loudspeaker (Canton Plus XS, frequency range: 150 Hz–21 kHz) was mounted about 30 cm in front of the elevated platform at 0° azimuth and 0° elevation in relation to the head of the gerbil. For correct responses, food rewards (20 mg round pellets, Bioserve Dustless Precision Pellets Formula F0163-J50) were dispensed from a custom-built feeder. It was connected via a flexible tube to a food tray mounted close to the platform. A red feeder light emitting diode (LED) served as an additional reinforcer. Custom software controlled feeder and feeder light, registered switches of the light barriers, and generated and delivered the stimuli. Experiments took place with no visible light in the chamber. The sessions could be monitored by a charge coupled device camera and infrared LEDs.

All stimuli were generated using a Linux workstation with an AMD processor and an RME sound card (Hammerfall DSP Multiface II). The analog signal from the sound card (sample frequency of 44.1 kHz) was delivered to a manual attenuator (Kenwood type RA-920A, presetting of overall signal level) and from there to a programmable attenuator (PA5 from Tucker Davis Technologies System 3, for computer controlled variations in signal level). The signal was then passed through an acoustic isolation transformer (1:1) to an amplifier (Rotel High Current 8 Channel Power Amplifier RMB-1048) and from there to the free-field loudspeaker. Stimuli were monitored on an oscilloscope (Conrad Voltcraft Type 630-2).

The setup was calibrated before every testing day using a spectrum analyzer (Stanford Research SR780) in conjunction with a G.R.A.S. Type 21SP Acoustic Transducer (microphone Type 40AF plus preamplifier) which was permanently mounted next to the pedestal.

We conducted three experiments with three different stimuli. In the first experiment, we obtained FDLs for pure tones. Test frequencies were 200, 400, 800, 1200, 1600, 3200, 6000, and 6400 Hz. Tone duration was 400 ms including a 25 ms Hanning ramp at stimulus onset and offset. The overall sound-pressure level (SPL) of 60 dB was roved by ±3 dB to eliminate any possible level cue for the gerbils.

In the second and third experiment, we obtained FDLs for mistuned components in a harmonic complex. The harmonic complex consisted of the first 12 harmonics with a fundamental frequency f_0 of 800 Hz and a SPL of 60 dB per component (with levels adjusted to the frequency response of the loudspeaker). The SPL of the overall complex was varied randomly by ±3 dB. As in the first experiment, the duration of the stimulus was 400 ms including a 25 ms Hanning ramp at onset and offset. In the reference stimuli, all components were harmonically related whereas in the test stimuli one of the harmonics was mistuned. For the second experiment, all 12 components of the complex started in sine phase [0°, Fig. 1(a)]. This resulted in reference stimuli that had exactly the same peaky waveform from one presentation to the next and the waveform was only different if a test stimulus with a mistuned harmonic was presented [Figs. 1(b) and 1(c)]. For the third experiment, all components started at a random phase which resulted in reference stimuli [Fig. 1(d)] and test stimuli [Fig. 1(e)] with a different temporal pattern for every stimulus presentation. We determined FDLs for frequency shifts of the first (800 Hz), the second (1600 Hz), and the eighth (6400 Hz) harmonic.

C. Procedure

Animals were trained using operant conditioning with positive reinforcement (food pellets) in a Go/NoGo para-

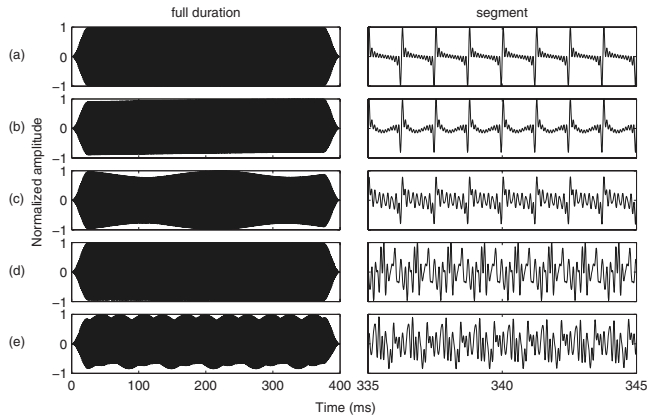


FIG. 1. Complex stimulus consisting of 12 harmonics and a fundamental frequency of 800 Hz. Left column shows complete stimulus length of 400 ms, right column a close up to compare the changes in the fine structure. [(a)–(c)] harmonic complex stimulus with all harmonics starting at sine phase with (a) no mistuning, (b) first harmonic mistuned by 0.07% (0.6 Hz threshold value), and (c) eighth harmonic mistuned by 0.07% (4.7 Hz); [(d)–(e)] harmonic complex stimulus with all harmonics starting at random phase with (d) no mistuning and (e) first harmonic mistuned by 2% (15.8 Hz threshold value). Note the differences in fine structure of every stimulus.

digm. During the session, we continuously repeated a reference stimulus every 1.3 s. The stimulus was either a pure tone (experiment 1) or a harmonic complex (experiments 2 and 3). To initialize a trial, the gerbil had to jump onto the pedestal and interrupt the light barrier. After a randomly chosen waiting interval between 1 and 7 s, the test stimulus was played instead of the reference stimulus. The test stimulus could either be the same as the reference stimulus during a catch trial or be different from the reference stimulus. For the first experiment, the test stimulus was a pure tone with a higher frequency than the reference stimulus. In the second and third experiment, the test stimulus was an inharmonic complex with one of the harmonics shifted upwards in frequency.

A correct response after a test stimulus ("Hit") was rewarded by switching on the feeder LED for 3 s and delivering a food pellet with a probability of 80%. Sessions consisted of a block of ten warm-up trials, followed by ten blocks of ten trials each. Within each block, three catch trials and seven different test trials were presented in a randomized order. Frequency increments in test trials consisted of fixed steps of a preselected Weber fraction. The step size was adjusted for the FDL measurements at the different harmonics. To provide at least some salient frequency differences to the animals, the step size for the two largest frequency increments that were not included in determining thresholds was double that of the step size used for determining thresholds. For example, a possible step size of 0.00005 Weber fraction around the threshold value (1.1 Hz in the 6400 Hz component) in the sine phase condition would result in frequency shifts of 0.32, 0.64, 0.96, 1.28, 1.60, 2.24, and 2.88 Hz for the seven test stimuli. In the block of warm-up trials, only the largest frequency increment was presented. The performance during these trials was not included in the data analysis.

D. Testing of human subjects

We obtained FDLs from four human subjects listening to the same harmonic complex stimuli that were generated by similar hardware as in the experiments with the gerbils. Subjects included the first author and three students, who took part in a practical course. All four subjects had normal hearing. Informed written consent was obtained from each subject. Except for the first author, the human subjects had no prior experience in psychoacoustic tasks. The students got course credits for completing the experiments. The experiments were conducted in a single-walled sound attenuating chamber (IAC Mini 250). The stimuli were presented via headphones (Sennheiser HDA 200). The headphones were calibrated using an artificial ear (Brüel&Kjær Type 4153) mounted on a microphone (Brüel&Kjær Type 4134) and a measuring amplifier (Brüel&Kjær Type 2610). All harmonics were presented at 60 dB SPL. The procedure being used in the gerbil experiments was adapted for humans, replacing the light barriers with a two-button response box. The human subjects had to press one of the buttons to initiate a trial and the other button to respond to a difference between reference and test stimulus. Feedback was given at correct responses by switching on a LED on the response box for 2 s.

E. Data analysis

A session was accepted as being valid if the subject responded correctly to at least 80% of the two test stimuli with the largest frequency difference and their false alarm rate did not exceed 20%. A psychometric function was constructed relating d' to the frequency shift in the test trials. The threshold was determined by linearly interpolating between adjacent points of the psychometric function as the amount of frequency shift resulting in a d' of 1.8 (Green and Swets, 1966). Two consecutive valid sessions in which thresholds differed no more than 1% Weber fraction for experiment 1 and experiment 3 and 0.1% Weber fraction for experiment 2 were combined to calculate the final FDL.

To exclude any training effects, we randomized the presentation order of every threshold for every animal within one experiment and repeated the first threshold after finishing each experiment. If the repeated threshold did not differ more than 3% (experiment 1), 1% (experiment 3), or 0.1% (experiment 2), respectively, from the threshold obtained the first time, then we assumed that no training effect affected the results. If it differed by more than that, the second threshold had to be repeated, then the third and so on until the repetition matched the original results. Data were always taken from the last measurement. The order in which the experiments were conducted was randomized and different for each animal. For a statistical analysis we used the statistical software package STATISTICA (StatSoft®, Version 7.0). Throughout the text, average results are reported as mean ± two times the standard error (mean ± 2SE). All p-values are two tailed unless stated otherwise.

F. Prior testing of the hearing sensitivity of gerbils

Prior to the frequency discrimination experiments, we tested the hearing sensitivity of all four gerbils at the fre-

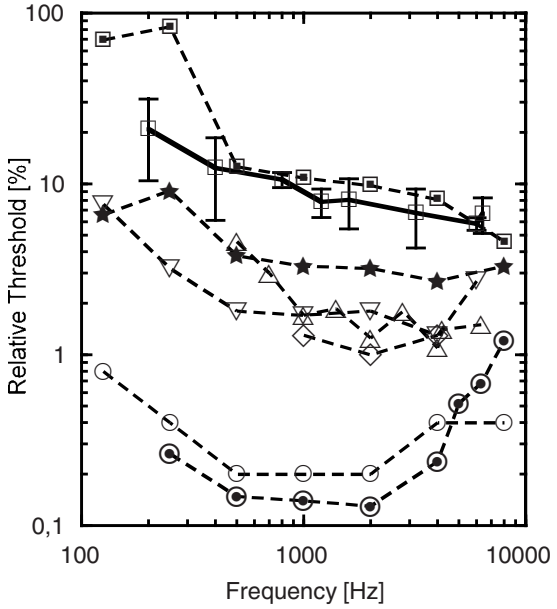


FIG. 2. Relative thresholds of pure tone frequency discrimination of the gerbils of the present study (open squares, continuous line, error bars =2SEs), in comparison to literature data: gerbil (filled squares), African monkey (*Cercopithecus mitis*, *Cercocebus albigena*, stars), and human (open circles): Sinnott *et al.* (1992); human (filled circles): Moore (1973); budgerigar (open diamond): Dent *et al.* (2000); starling (open standing triangles): Klump *et al.* (2000); and pigeon (open hanging triangles): Sinnott *et al.* (1980).

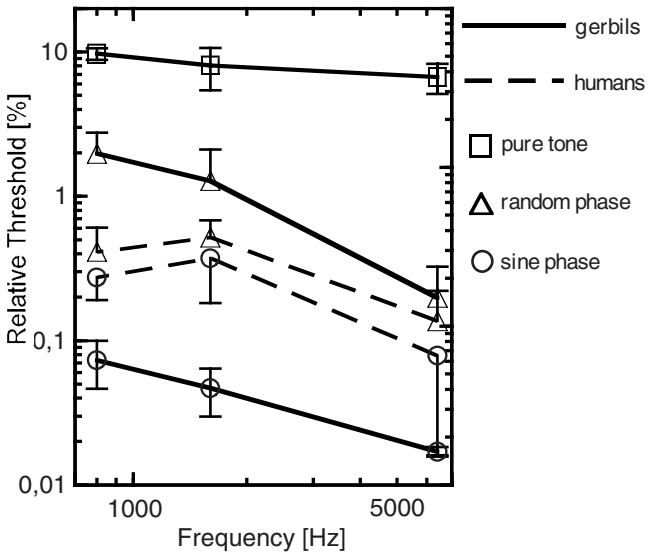


FIG. 3. Mean thresholds for gerbils and human subjects tested on the frequency discrimination ability of pure tones (only gerbils, squares) and on the frequency discrimination ability of the mistuned 800, 1600, and the 6400 Hz harmonic within a complex with all harmonics starting at random phase (triangles) and within a complex with all harmonics starting at sine phase (circles). Error bars represent 2 SF.

quencies of 100, 500, 1000, 2000, 4000, and 8000 Hz. Testing took place in the same chamber with the same procedure as was used for the frequency discrimination experiments. No stimulus was played except in test trials in which a single 800 ms pure tone was presented as the test signal. During a catch trial, no signal was presented. The method of constant stimuli was used with a step size in signal level of 3 dB. A session was accepted if at least 80% of responses to the two stimuli with the highest level were "Hits" and no more than 20% false alarms were reported. The hearing threshold was defined as the SPL at a d' of 1.8 (Green and Swets, 1966). Two consecutive valid sessions in which thresholds differed no more than 3 dB were combined to calculate the final hearing threshold based on 20 responses from 20 trials at each level and 60 catch trials. The presentation order of all tested frequencies was randomized for every animal and the first threshold was repeated after finishing the test series of the experiment to exclude any training effects. If the repeated threshold differed more than 3 dB from the threshold obtained the first time, then the next threshold of the series was repeated. The range of tested frequencies with the best hearing was between 1 kHz (-1.0 ± 4.1 dB SPL, mean $\pm$ standard deviation) and 8 kHz (-8.9 ± 1.5 dB SPL) with an average of -5.8 dB SPL. At the frequencies of 2 kHz and 4 kHz, we determined hearing sensitivity thresholds of -4.7 ± 2.7 dB SPL and -8.4 ± 1.3 dB SPL, respectively. Below 1 kHz, the hearing sensitivity declined and thresholds increased up to 16.2 ± 3.7 dB SPL at a frequency of 500 Hz and 38.0 ± 1.3 dB SPL at 100 Hz. In the low-frequency range up to 1 kHz, the gerbils' hearing did not differ much from the hearing sensitivity thresholds obtained by Ryan (1976). At frequencies from 1 kHz to 8 kHz we observed between 7 and 14 dB SPL lower hearing thresholds than Ryan (1976). All four gerbils could be considered as having a normal hearing.

TABLE I. FDLs of pure tones. Individual data of three gerbils and their mean with 2SE in Hz.

Frequency (Hz)	Lv	Sa	Th	Mean $\pm$ 2SE
200	45	22	58	42 ± 21
400	67	25	56	49 ± 25
800	84	77	72	78 ± 7
1200	84	112	86	94 ± 18
1600	87	147	153	129 ± 42
3200	158	196	295	216 ± 81
6000	377	327	345	350 ± 29
6400	383	528	374	428 ± 100

III. RESULTS

A. Experiment 1: Pure tone frequency discrimination

Figure 2 shows the mean pure tone FDLs of the gerbils in percent Weber fraction (solid line with open squares) in relation to the reference frequency. The corresponding FDLs in Hertz are listed in Table I. The relative frequency shift necessary for the detection decreased from a Weber fraction of 20.8% $\pm$ 10.4% (mean $\pm$ 2SE) at 200 Hz to 6.7% $\pm$ 1.6% at 6400 Hz. This corresponds to an absolute frequency difference of 41.7 ± 20.9 Hz at a reference frequency of 200 Hz. The absolute frequency shift necessary for detection increased gradually with increasing frequency of the reference signal up to 428.2 ± 100.2 Hz at 6400 Hz. A one-way repeated measure ANOVA revealed a significant effect of the reference frequency on the relative frequency shift at detec-

TABLE II. FDLs for three harmonics of an 800 Hz complex with all components starting at sine phase and for three harmonics of an 800 Hz complex with all components starting at random phase. Individual data of three and four gerbils, respectively, and of four human subjects and their mean with 2SE in Hz.

Gerbils						
Sine phase condition						
Harmonic number	Frequency of harmonic (Hz)	Lv	Sa	Th	Ls	mean ± 2SE
1	800	0.51	0.60	0.36	0.87	0.59 ± 0.21
2	1600	0.44	0.97	0.60	0.99	0.75 ± 0.27
8	6400	1.14	1.03	1.18	1.01	1.09 ± 0.08
Random phase condition						
Harmonic number	Frequency of harmonic (Hz)	Lv	Sa	Th		mean ± 2SE
1	800	19.9	9.2	18.4		15.8 ± 6.7
2	1600	33.4	16.6	11.4		20.5 ± 13.3
8	6400	5.8	19.8	12.5		12.7 ± 8.1
HUMANS						
Sine phase condition						
Harmonic number	Frequency of harmonic (Hz)	AK	AV	NG	AT	mean ± 2SE
1	800	1.9	3.1	2.2	1.5	2.2 ± 0.7
2	1600	7.6	6.5	1.5	8.2	6.0 ± 3.0
8	6400	2.9	10.5	5.5	1.3	5.1 ± 4.0
Random phase condition						
Harmonic number	Frequency of harmonic (Hz)	AK	AV	NG	AT	mean ± 2SE
1	800	1.9	3.5	2.5	5.4	3.3 ± 1.5
2	1600	9.8	4.7	8.9	10.5	8.3 ± 2.6
8	6400	2.4	15.1	7.1	5.5	8.8 ± 5.4

tion threshold ($p<0.05$; $\eta^2=0.56$) as well as on the absolute frequency shift at detection threshold ($p<0.05$; $\eta^2=0.95$). The *post hoc* pairwise comparisons revealed a significant difference between the mean thresholds for the absolute frequency discrimination determined for 800 and 6400 Hz reference tones (Tukey test, $p<0.01$). None of the other *post hoc* comparisons revealed significant differences.

B. Experiment 2: Frequency difference limens of mistuned components in a sine phase harmonic complex

In this experiment, we presented a harmonic complex to the gerbils in which every component started in sine phase [0°, Fig. 1(a)]. Applying a frequency shift on one of the harmonics produced a change in the fine structure of the waveform over the duration of the stimulus and a slight amplitude modulation occurred [Figs. 1(b) and 1(c)]. This stimulus condition resulted in the lowest FDLs of all three experiments. Figure 3 shows the amount of mistuning at threshold expressed as the Weber fraction relative to the frequency of the respective harmonic. The corresponding absolute threshold values in hertz are shown in Table II. The Weber fraction necessary for detecting the mistuning decreased from 0.073% ± 0.027% (mean ± 2SE) at the 800 Hz harmonic to 0.017% ± 0.001% at the 6400 Hz harmonic. Harmonic number significantly affected the Weber fraction necessary for detection of the mistuning (one-way repeated measure ANOVA, $p<0.05$, $\eta^2=0.81$). Pairwise *post hoc* tests revealed a significant decrease of the threshold Weber fraction between the data for the 800 and 6400 Hz harmonic (Tukey test, $p<0.01$) and a trend for the decrease comparing the 1600 and 6400 Hz harmonic (Tukey test, $p=0.08$). The corresponding absolute frequency shift significantly increased from 0.59 Hz ± 0.21 Hz at the 800 Hz harmonic to 1.09 Hz ± 0.08 Hz at the 6400 Hz harmonic [one-way repeated measure analysis of variance (ANOVA), $p<0.05$, $\eta^2=0.69$, a *post hoc* Tukey test revealed a significant difference between the data for the 800 Hz and the 6400 Hz harmonic, $p<0.05$].

C. Experiment 3: Frequency difference limens of mistuned components in a random phase harmonic complex

During the third experiment, we randomized the starting phase of every component of the 800 Hz complex for every stimulus we presented throughout the experiment in order to degrade temporal cues. Figure 3 shows the relative frequency shifts necessary for the detection of a mistuning. The corresponding absolute frequency shifts at threshold are listed in Table II. A one-way repeated measure ANOVA revealed a

significant difference in the relative frequency shifts at the detection threshold from 2.0% ± 0.8% (mean ± 2SE) at the 800 Hz harmonic to 0.22% ± 0.16% at the 6400 Hz harmonic ($p<0.05$; $\eta^2=0.79$). However, pairwise *post hoc* comparisons only revealed a significant decrease in the relative frequency shift necessary for the detection of a mistuning from the 800 Hz harmonic to the 6400 Hz harmonic (Tukey test, $p<0.05$). All other pairwise comparisons were not significant. A one-way repeated measure ANOVA revealed no significant effect of the harmonic on the absolute frequency shift in Hertz.

D. Comparison of all three experiments

Comparing the results of all three experiments with each other (Fig. 3), we observed a highly significant difference between the mean relative frequency shift at the detection threshold among the three different experimental conditions. A two-way repeated measure ANOVA with experimental conditions (stimulus type) and reference frequency as factors including the data from the three gerbils that completed all experiments resulted in values of $p<0.001$ and $p<0.05$, respectively. The stimulus type had a strong effect on the discrimination threshold (effect size $\eta^2=0.99$). The thresholds for the relative frequency shifts were the lowest for harmonics in sine phase stimuli, about one order of magnitude higher for harmonics in random phase stimuli, and about another order of magnitude higher for pure tones (all pairwise planned comparisons of the experimental condition within a reference frequency revealed significant differences, all $p<0.05$). The second main effect reference frequency also showed a high effect size η^2 of 0.78. There was no significant interaction between the factors experimental condition and reference frequency (two-way repeated measure ANOVA, $p=0.26$).

The performance of the gerbils during the sessions was different for the second and third experiments. Corresponding to the large difference in FDL, it was more difficult to obtain data in the random phase condition, than in the sine phase condition. In the sine phase condition, the gerbils needed on an average five sessions to obtain a threshold which was significantly less than in the random phase condition with an average of 14 sessions to obtain a threshold (paired t-test, $p<0.05$). When the gerbil switched from the sine phase condition to the random phase condition, their false alarm rate was significantly increased (from 98 to 160 false alarms, averaged over five sessions of two gerbils, paired t-test, $p<0.05$), whereas a slight decrease was observed when switching from random phase to sine phase (from 104 to 101 false alarms, averaged over five sessions of one gerbil, paired t-test, $p=0.92$).

E. Comparison with human data

Human FDLs for frequency shifts in the harmonic complex were obtained by using the same procedure and stimuli as for the gerbils to be able to compare FDLs directly (Table II, Fig. 3). A two-way repeated measure ANOVA with experimental condition (sine phase and random phase) and reference frequency (800, 1600, and 6400 Hz) as factors revealed no significant difference between the Weber fraction of the shift at threshold for sine phase and random phase stimuli ($p=0.129$). However, the reference frequency as a main effect had a significant effect on the Weber fraction at threshold ($p<0.01$, $\eta^2=0.79$). *Post hoc* tests revealed significant differences for the comparison of the data obtained for reference frequencies of 6400 and 1600 Hz (Tukey test, $p<0.01$) and for reference frequencies of 6400 and 800 Hz (Tukey test, $p<0.05$). No significant interactions between reference frequency and experimental condition were found. Neither experimental condition nor reference frequency had a significant effect on the absolute frequency differences necessary for the detection of the mistuning (two-way repeated measure ANOVA).

For sine phase stimuli, the Weber fraction of the shift at threshold was smaller in gerbils than in humans, i.e., gerbils were more sensitive for detecting the mistuning. This difference was significant for the 800 Hz harmonic (t-test, $p<0.005$) and the 1600 Hz harmonic (t-test, $p<0.02$), and there was a trend in the same direction for the 6400 Hz harmonic (t-test, $p=0.096$). For random phase stimuli, the Weber fraction of the shift at threshold was larger in gerbils than in humans. This difference was significant for the 800 Hz harmonic (t-test, $p<0.005$) and there was a trend for the 1600 Hz harmonic (t-test, $p=0.091$).

IV. DISCUSSION

A. Pure tone frequency difference limens

Compared to the discrimination ability of other animal species (Fig. 2), the gerbil exhibits a poor ability to discriminate the frequency of pure tones (Fig. 2, see also Sinnott *et al.*, 1992), especially in the low-frequency region. Human data are shown from Moore (1973) and Sinnott *et al.* (1992) for comparison. The highest frequency discrimination acuity in humans being as low as 0.2% shift in frequency is found between 500 Hz and 2 kHz. Bird species vary in their ability to discriminate between pure tones of different frequencies. Pigeons, for example, exhibit fairly high FDLs at frequencies higher than 1 kHz that are, however, within the range of nonhuman mammals (Sinnott *et al.*, 1980). Klump *et al.* (2000) reported pure tone FDLs in European starlings from 4.6% for a frequency of 500 Hz to 1.5% for 6300 Hz. Studying budgerigars, Dent *et al.* (2000) found FDLs of around 1.3% at frequencies of 1, 2, and 4 kHz. The SPLs used for the experiments in the different studies were all around 60–65 dB SPL which compares well to the sensation level used in the present study.

Pure tone frequency discrimination can be achieved based on two different mechanisms. The first mechanism applies to the low-frequency region where phase locking of auditory nerve (AN) fibers to the stimulus waveform occurs. For these frequencies, the distribution and change of interspike intervals in the AN fibers can be exploited to estimate the change in frequency (e.g., Wever, 1949). The second mechanism mainly applies to the high frequency region where phase locking is no longer possible. It evaluates the place of excitation in the cochlea for frequency discrimination. Ideas for a place mechanism in which frequencies are

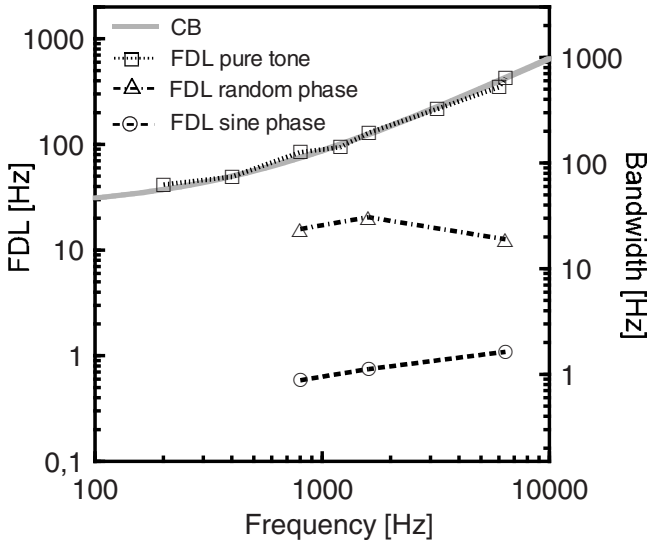


FIG. 4. Pure tone FDLs (black dotted line, squares), random phase FDLs (black dash-dotted line, triangles), and sine phase FDLs (black dashed line, circles) of gerbils are plotted in comparison to the cochlear place-frequency function (gray solid line). The CB function is derived from the general bandwidth function from Greenwood (1961) with parameters obtained from physiological data of gerbils (Müller, 1996).

represented tonotopically along the basilar membrane reach as far back as to the 17th century (Du Verney, 1683, c.f. Cheveigné, 2005). Later on, von Békésy (1960) and Zwicker (1970) proposed a place mechanism in which a frequency is represented as a certain excitation pattern along the basilar membrane. The detection of a change in frequency is based on changes in the excitation pattern.

Comparing a wide range of species, Fay (1992) favored a place mechanism in most animal species. He stated, based on analyses of cochlear place-frequency functions by Greenwood (1961, 1990), that perceptual units in the frequency domain such as the pure tone FDL, the critical bandwidth (CB), critical masking ratio bandwidth, and the psychophysical tuning curve can all be related to the spatial frequency map on the basilar membrane. In the Mongolian gerbil, a physiologically determined cochlear map (Müller, 1996) fitted by the cochlear place-frequency function (Greenwood, 1961) can be used to calculate the relation between the critical ratio bandwidth or the CB and the FDL (Kittel *et al.*, 2002). Figure 4 shows an estimate of the CB as a function of frequency (gray line) in comparison to the FDLs determined in this study. CBs were calculated from a cochlear place-frequency function that was fitted to the behavioral estimates of the CB (Kittel *et al.*, 2002). They range from 56 Hz at a center frequency of 200 Hz to 638 Hz at 6400 Hz. The function of the observed FDLs (black dotted line on top of the gray line) runs parallel to the CB function (i.e., it is offset by a scaling factor of 1.5) over the whole range of tested frequencies (Fig. 4, axes were adjusted accordingly). This parallel run of the CB function and the pure tone FDL function in the gerbil suggests that at least over the tested frequency range of 200–6400 Hz, a pure place mechanism might explain the observed FDLs.

What remains unexplained is the small scaling factor relating the FDL to the CB of the gerbil. Fay (1992) showed that the best FDLs in humans are related to the CB by a factor of 50. He proposed that this scaling factor applies to most mammals and birds at least for frequencies above 1 kHz. However, Sinnott *et al.* (1992) argued against this scaling factor by presenting data from African monkeys (*Cercopithecus mitis*, *Cercocebus albigena*) that show much larger pure tone FDLs than humans in spite of having a CB that is similar to the human CB. The FDLs observed in monkeys can only be matched to the CB function by a scaling factor of 3 (see Sinnott *et al.*, 1992) which is more similar to the scaling factor of 1.5 found for the gerbil in the present study. Studying humans, Sek and Moore (1995) related different measures of frequency discrimination to the auditory filter bandwidth. They observed that the scaling factor is reduced considerably at frequencies for which humans have to rely on a place mechanism for frequency analysis. It deteriorated from about 18 at 4 kHz to about 5 at 8 kHz. This is similar to the range of scaling factors from about 7 to 17 that was found by Buus *et al.* (1995) fitting the FDL data from European starlings to the starling's CB function. These scaling factors, however, are much larger than the value found for the gerbil.

What may determine the value of the scaling factor and the value of the FDL that can vary considerably between species? Fay (1992) proposed that the largest absolute frequency discrimination acuity will be found in mammals with a poor high frequency hearing (thus perceiving a small frequency range) and a long cochlea. Frequency discrimination ability will be poor when the cochlea is short but the frequency range of sensitive hearing is wide. The gerbil cochlea is only 11 mm long (Müller, 1996) which is short in comparison to 34 mm reported for the human cochlea (Miller, 2007). Corresponding to the difference in length, the number of hair cells along the basilar membrane is reduced from about 3500 inner hair cells (IHCs) in the human cochlea (Wright *et al.*, 1987) to about 1400 IHCs in the gerbil cochlea (Plassmann *et al.*, 1987). Moreover, the hearing range of the gerbil extending from approximately 0.1–60 kHz (Ryan, 1976) is much larger than the human hearing range extending from about 0.01 to 16 kHz (ISO 389-7, 1996). Furthermore, the gerbil's larger frequency range is mapped onto more critical bands than the frequency range of humans (estimated on the basis of data from Kittel *et al.*, 2002 for gerbils, and Greenwood, 1961 and Moore and Glasberg, 1983 for humans). The distance on the cochlear map covered by a critical band in the gerbil is about 0.21 mm which is small compared to the distance of 1.15 mm estimated for humans (Fay, 1992). Thus, a specific shift in frequency relative to the reference will result in a smaller change of the spatial pattern of excitation in the cochlea involving a much smaller number of hair cells in gerbils compared to humans (a difference by a factor of 5–8). The difference between human and gerbil FDLs may thus be related to the higher number of hair cells in humans and encoding of frequencies by a larger number of AN fibers. This might result in a more fine-scaled frequency representation in the cochlea.

B. Frequency difference limens of mistuned components in a harmonic complex

The results of the "mistuned harmonic" experiments appeared in stark contrast to the pure tone FDLs. While the performance of the gerbils in pure tone frequency discrimination was unusually poor and about 20 times worse than the performance of humans, gerbils were about four times better than humans in detecting a mistuned harmonic in the sine phase condition. When gerbils detected a mistuned harmonic in a harmonic complex with all components starting at a random phase, FDLs were about one order of magnitude worse than FDLs obtained in the sine phase condition. This makes gerbils about two to four times less sensitive in reporting the mistuning compared to humans in the random phase condition.

1. Models for pitch perception and frequency discrimination

Two classes of models have been applied to explain pitch perception and discrimination of complex tone stimuli: (1) spectral pattern recognition models and (2) autocorrelation models. The spectral pattern recognition models (early models by Goldstein, 1973, Wightman, 1973, and Terhardt, 1974) assume that the frequency of every harmonic is determined either by place or by temporal mechanisms and then the best fitting fundamental frequency is estimated. An internal central processor or pattern recognizer decides if the frequency of a component of the complex is belonging to the harmonic series or not and has therefore been called a "harmonic sieve" by Duifhuis *et al.* (1982). One constraint of this type of model is, however, that the harmonics of the complex have to be resolved. Furthermore, the phase of components of complex stimuli is not taken into account as the pitch is extracted from the peripheral excitation pattern. A spectral pattern recognition model thus implies that the frequency selectivity of the basilar membrane has to be sufficiently good in order to reach low FDLs. To evaluate the involvement of place mechanisms in the frequency discrimination ability of mistuned harmonics in a harmonic complex, we plotted the FDLs of the random phase condition and of the sine phase condition of our study in relation to the place-frequency map in Fig. 4. If a place model would explain the FDLs for mistuned harmonics in complex stimuli at both condition, then the FDLs would be parallel to the cochlear place-frequency function determined by Greenwood (1961). The FDLs of mistuned harmonics at the random phase and the sine phase condition cannot be fitted to the cochlear map function by the same scaling factor at least over the tested frequency range which makes place mechanisms more unlikely to operate over the whole frequency range tested. Furthermore, we would expect much higher FDLs for mistuned harmonics in a harmonic complex in gerbils than in humans due to the gerbil's poor pure tone FDLs, if we assume that the pattern recognition model applies. Our results do not support the prediction based on the spectral pattern recognition model.

Instead, the FDLs of mistuned harmonics in a complex with all harmonics starting at sine phase are two orders of magnitude smaller than pure tone FDLs. This applies not only to the 6400 Hz harmonic that is unresolved in the gerbil auditory system but also for the 800 and the 1600 Hz harmonic that are resolved. Detecting an inharmonicity in the sine phase complex stimuli by the Mongolian gerbil therefore seems not to involve the analysis of the spectral pattern of a complex stimulus or the evaluation of temporal patterns within a single auditory filter but rather points to processing of temporal patterns beyond the stage of the basilar membrane.

Most of the temporal models for the processing of harmonic tone complexes rely on autocorrelation which is computed on the basis of the temporal envelope or temporal fine structure in all frequency channels. Such models were first introduced by Licklider (1951) and further developed by Meddis and Hewitt (1991a, 1991b), Meddis and O'Mard (1997), Patterson *et al.* (1995), and Denham (2005) using a summary autocorrelation function over all channels. The autocorrelation function suggested by Licklider (1951) is calculated on the basis of spike occurrences from a single neuron or a group of neurons and depends on the waveform envelope of unresolved harmonics to extract the pitch of a complex. A summary autocorrelation function combines the periodicity estimates derived from all frequency channels and therefore they do not depend only on unresolved components of a harmonic complex.

2. Temporal cues available for detecting a mistuned harmonic in a complex

Temporal cues that might be used for detecting mistuning of a component in a complex with all harmonics in sine phase could be a beating in the mistuned harmonic complex or a gradual phase shift over the duration of the stimulus resulting in a change of the fine structure of the ongoing stimulus. A beating is produced when the frequency-shifted harmonic in the mistuned harmonic complex is unresolved and interacts with the adjacent harmonics in the complex. Human subjects in the study by Moore *et al.* (1985) identified the mistuning of unresolved harmonics by hearing a kind of "beat." This detection based on the amplitude modulation suggests that for humans temporal cues may play a greater role than spectral cues for detecting an unresolved mistuned harmonic in a complex stimulus. Temporal mechanisms for detecting a mistuned component in an otherwise harmonic complex in humans have also been suggested by Le Goff and Kohlrausch (2005). Their simulations showed that the excitation patterns for flat-spectrum and sloped-spectrum complex stimuli are less affected by mistuning the fundamental frequency component than the excitation pattern of the frequency-shifted pure tone. In a simulation of the envelope pattern of the output signal of the auditory filters, however, Le Goff and Kohlrausch (2005) showed that mistuning the fundamental frequency component in a sloped spectrum changed the temporal envelope of the signals in the auditory filters considerably. This points to the use of temporal mechanisms by humans even for resolved mistuned harmonics under certain circumstances.

Some authors proposed the evaluation of distortion products producing beating as a possible explanation for the

good performance in detecting a mistuned harmonic in a complex (Moore *et al.*, 1985; Pressnitzer and Patterson, 2001). Distortion products emerge at the point on the basilar membrane where the traveling waves of at least two nearby tones (with frequencies $f1$ and $f2$) overlap. Generally, the most prominent distortion product is the cubic $2f1$-$f2$ distortion which may under certain circumstances interact considerably with components of the presented stimulus on the basilar membrane. We estimated the possible interaction of the mistuned harmonic and the cubic distortion product which would develop at the place where the shifted harmonic has been. Taking into account measurements of distortion product otoacoustic emissions and neuronal responses by Faulstich and Kössl (1999) and Mom *et al.* (2001), we concluded that the SPL of the distortion product would be at least 20 dB lower than the level of the interacting mistuned harmonic. The estimated modulation depth due to the beating between the distortion product and the mistuned harmonic was around 1.1 dB. Since the gerbil's intensity discrimination threshold is at least 3 dB (Sinnott *et al.*, 1992), we conclude that the depth of modulation resulting from the beating of a distortion product with the mistuned component would not allow the gerbils to detect the beating and to use it as a detection cue.

A more likely temporal cue available to the gerbils for detecting the mistuning is an ongoing change in the fine structure over the duration of the mistuned harmonic complex. The change results from the gradual phase shift of the mistuned component compared to the other components. The peaky waveform of the sine phase stimulus may aid the gerbil in using this cue. In the random phase condition, however, only some of the randomly produced waveforms have a peaky temporal structure. Sensitivity to the phase or to changes of the phase in harmonic complex stimuli has been proposed in several physiological and psychophysical studies (e.g., Bilsen, 1973; Horst *et al.*, 1986, 1990; Moore and Glasberg, 1989; Simmons *et al.*, 1993). Phase relations might be detected on the basilar membrane when unresolved harmonics fall within one auditory filter. Phase relations for resolved harmonics might be preserved and detected in more central stages of the auditory system where they can be analyzed across channels. Hartmann (1988) showed that reversing the phase of the resolved fourth harmonic in comparison to the other components of a 200 Hz complex influenced the detection of a mistuning in relation to the duration of the stimulus. Since the mistuned component with the phase shift of 180° and the adjacent harmonics fall in different auditory filters, a processing of the phase relationship between the mistuned harmonic and the adjacent components is less likely to be analyzed on the basilar membrane. Thus, the phase effects observed by Hartmann (1988) suggest that the processing of these temporal cues in humans is not limited to cochlear within-channel mechanisms.

A further improvement for detecting a mistuned harmonic in the sine phase compared to the random phase stimulus could be explained by the stimulus statistics. In the sine phase condition, a constant temporal pattern is provided with every background stimulus that only changes during the presentation of a test stimulus. In the random phase stimulus

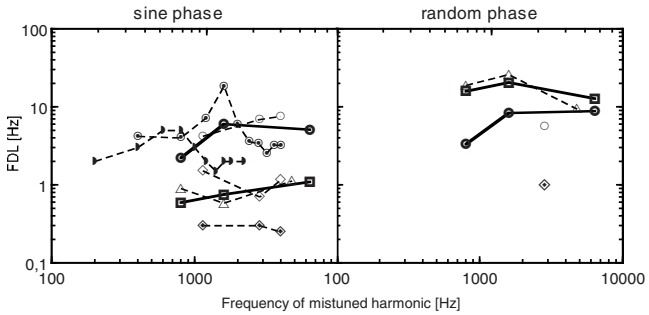


FIG. 5. Absolute thresholds in Hz of mistuned components of sine phase (left panel) and random phase (right panel) complex stimuli of gerbils (thick square, continuous line) and humans (thick circle, continuous line) determined in the present study in comparison to literature data. Zebra finch (filled diamond), budgerigar (open diamond) and human (open circle): Lohr and Dooling (1992); starling (open triangle): Klump *et al.* (2008); human (filled circle): Moore *et al.* (1985); human (half circle): Lee and Green (1994).

condition, we continuously varied the temporal fine structure and phase relations by randomizing the starting phase of every harmonic in every presented stimulus. Changing the temporal fine structure with every presented stimulus removed the possibility to compare temporal patterns of the harmonic complex sequentially. This high level of uncertainty in the stimuli might have resulted in the observed decline of FDLs in the random phase condition compared to the sine phase condition. The observation that gerbils needed more sessions to obtain a threshold in the random phase condition than in the sine phase condition and that they showed an increased false alarm rate when switching from the sine phase to the random phase condition also suggests that available cues in the random phase condition may have been less salient.

The amplitude modulation of the stimulus envelope resulting from mistuning could provide another possible cue for detecting a mistuned harmonic in a complex in the random phase condition. The depth of the amplitude modulation, however, has to be sufficiently large to be perceived by the gerbil. In the sine phase condition, the FDLs are too small to produce an exploitable modulation. Also in the random phase condition, the modulation spectrum of the mistuned complex stimulus at threshold value indicated that the modulation depth is not large enough to be detectable by the gerbil.

3. Comparison with other species

Figure 5 gives an overview of FDLs of mistuned components in harmonic complexes in several animal species and humans. Zebra finches and budgerigars exhibit very low FDLs for detecting mistuning of components in harmonic complexes in the sine phase condition (Lohr and Dooling, 1998) that were similar to the FDLs of gerbils determined in the present study. An increased FDL in the random phase condition compared to the sine phase condition could be shown for zebra finches in the study by Lohr and Dooling (1998). The European starling also exhibits discrimination thresholds that are similar to those found in gerbils (Klump *et al.*, 2008). The FDLs in the bird species were significantly lower compared to those of humans when all components of

the complex stimulus started in sine phase. The studies show that gerbils are not the only species that have such low FDLs for detecting a frequency shift in harmonic complexes in the sine phase condition. The difference in the salience of mistuning in the sine phase and the random phase conditions in gerbils and birds suggests that temporal fine structure of the stimuli may be of particular importance. Sinnott and Mosteller (2001) found gerbils to be equally sensitive in detecting frequency changes in speech formants compared to detecting a frequency shift of a harmonic in the random phase experiment as determined in the present study. Threshold values for detecting frequency changes in speech formants were similarly decreased compared to pure tone FDLs in gerbils as is described here for detecting a mistuning of a harmonic in a random phase complex. Furthermore, the ratio between the gerbil and human formant FDLs in the speech discrimination task and between the gerbil and human FDL in the random phase condition was found to be similar. These comparisons could be interpreted as a hint that humans and gerbils may employ similar mechanisms in processing frequency changes in both experiments. However, it may well be that both species use different mechanisms resulting in a similar sensitivity for detecting the frequency shift.

Human subjects were not as sensitive as gerbils or zebra finches (Lohr and Dooling, 1998) to the loss of distinctive temporal features of the stimulus during the presentation of random phase harmonic complexes. In humans, the FDLs for mistuned harmonics in harmonic complexes with all components starting at random phase were never significantly different to the FDLs for mistuned harmonics in the sine phase condition (see present study or Lohr and Dooling, 1998). These results are supported by studies from Dooling *et al.* (2002) and Patterson (1987) who showed that humans cannot discriminate cosine phase complex stimuli from random phase stimuli at a fundamental frequency of 800 Hz. This points to a reduced importance of temporal fine structure as a cue in humans compared to gerbils and birds.

Which constraints may bias gerbils and birds to rely more and humans to rely less on temporal cues when detecting mistuning of components in a harmonic complex? The length of the cochlea may hint at a possible explanation. The Mongolian gerbil has a short cochlea compared to the human cochlea with much fewer hair cells covering a CB. Thus, the same change in frequency in a gerbil and a human cochlea results in a much smaller change of the spatial pattern of excitation in the gerbil than in humans. In order to compensate for this, the gerbil and possibly the birds (also having a short cochlea) might have switched to exploiting temporal cues in complex stimuli to detect small amounts of mistuning. However, by proposing the use of temporal fine structure as a cue the question remains why gerbils seem not to exploit any temporal cues for detecting frequency differences in pure tones (at least in the frequency range below 2 kHz for which this is suggested in humans, e.g., Moore, 1973). Since humans have a much longer cochlea than birds or the gerbil, the evolutionary pressure to use changes in the temporal fine structure as a cue in the processing of frequency shifts in harmonic complexes may have been much lower than in the other species discussed here.

ACKNOWLEDGMENTS

We thank Susanne Groß and Anne Vellage for helping with the experiments, Antje Tjarks, Nantje Gelszat, and Anne Vellage for being motivated subjects in the human experiments, Holger Dierker for technical support, Stephan Ewert for assisting in the reverberation measurement, and Stefan Strahl for helping out in computational problems. This study was supported by the DFG within the Sonderforschungsbereich SFB TR31 "The Active Auditory System" and the International Graduate School GRK 591 "Neurosensory Science and Systems."

Bilsen, F. A. (**1973**). "On the influence of the number and phase of harmonics on the perceptibility of the pitch of complex signals," Acustica **28**, 60–65.

Buus, S., Klump, G. M., Gleich, O., and Langemann, U. (**1995**). "An excitation-pattern model for the starling (*Sturnus vulgaris*)," J. Acoust. Soc. Am. **98**, 112–124.

Cheveigné, A. (**2005**). "Pitch perception models," in *Pitch: Neural Coding and Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper, (Springer, New York), pp. 169–233.

Denham, S. L. (**2005**). "Pitch detection of dynamic iterated rippled noise by humans and a modified auditory model," BioSystems **79**, 199–206.

Dent, M. L., Dooling, R. J., and Pierce, A. L. (**2000**). "Frequency discrimination in budgerigars (*Melopsittacus undulatus*): Effects of tone duration and tonal context," J. Acoust. Soc. Am. **107**, 2657–2664.

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.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (**1982**). "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception," J. Acoust. Soc. Am. **71**, 1568–1580.

Faulstich, M., and Kössl, M. (**1999**). "Neuronal response to cochlear distortion products in the anteroventral cochlear nucleus of the gerbil," J. Acoust. Soc. Am. **105**, 491–502.

Fay, R. R. (**1992**). "Structure and function in sound discrimination among vertebrates," in *The Evolutionary Biology of Hearing*, edited by D. Webster, R. R. Fay, and A. Popper (Springer, New York), pp. 229–263.

Goldstein, J. L. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1516.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Greenwood, D. D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344–1356.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hartmann, W. M. (**1988**). "Pitch perception and the segregation and integration of auditory entities," in *Auditory Function*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 623–645.

Horst, J. W., Javel, E., and Farley, G. R. (**1986**). "Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms," J. Acoust. Soc. Am. **79**, 398–416.

Horst, J. W., Javel, E., and Farley, G. R. (**1990**). "Coding of spectral fine structure in the auditory nerve. II. Level-dependent nonlinear responses," J. Acoust. Soc. Am. **88**, 2656–2681.

ISO 389-7, (**1996**). "Acoustics—Reference zero for the calibration of audiometric equipment—Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions."

Kittel, M., Wagner, E., and Klump, G. M. (**2002**). "An estimate of the auditory-filter bandwidth in the Mongolian gerbil," Hear. Res. **164**, 69–76.

Klump, G. M., Groß, S., and Itatani, N. (**2008**). "Processing of a mistuned harmonic by the European starling auditory system," 31st MidWinter Research Meeting [J. Assoc. Res. Otolaryngol. **31**, 139].

Klump, G. M., Langemann, U., and Gleich, O. (**2000**). "The European starling as a model for understanding perceptual mechanisms," in *Auditory Worlds: Sensory Analysis and Perception in Animals and Man*, edited by G. A. Manley, H. Fastl, M. Kössl, H. Oeckinghaus, and G. M. Klump (Wiley-VCH, Weinheim), pp. 193–211.

Le Goff, N., and Kohlrausch, A. (**2005**). "The use of temporal cues for frequency discrimination of the fundamental component in a complex

tone," Proceedings of the Forum Acusticum, Budapest, pp. 1503–1507.
Lee, J., and Green, D. M. (**1994**). "Detection of a mistuned component in a harmonic complex," J. Acoust. Soc. Am. **96**, 716–725.
Licklider, J. C. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–134.
Lin, J.-Y., and Hartmann, W. M. (**1998**). "The pitch of a mistuned harmonic: Evidence for a template model," J. Acoust. Soc. Am. **103**, 2608–2617.
Lohr, B., and Dooling, R. J. (**1998**). "Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*)," J. Comp. Psychol. **112**, 36–47.
Meddis, R., and Hewitt, M. J. (**1991a**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., and Hewitt, M. J. (**1991b**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity," J. Acoust. Soc. Am. **89**, 2883–2894.
Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.
Miller, J. D. (**2007**). "Sex differences in the length of the organ of Corti in humans," J. Acoust. Soc. Am. **121**, EL151–EL155.
Mom, T., Bonfils, P., Gilain, L., and Avan, P. (**2001**). "Origin of cubic difference tones generated by high-intensity stimuli: Effect of ischemia and auditory fatigue on the gerbil cochlea," J. Acoust. Soc. Am. **110**, 1477–1488.
Moore, B. C. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.
Moore, B. C., 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., and Glasberg, B. R. (**1989**). "Difference limens for phase in normal and hearing-impaired subjects," J. Acoust. Soc. Am. **86**, 1351–1365.
Moore, B. C., Glasberg, B. R., and Peters, R. W. (**1984**). "Frequency and intensity difference limens for harmonics within complex tones," J. Acoust. Soc. Am. **75**, 550–561.
Moore, B. C., Peters, R. W., and B. R. Glasberg (**1985**): "Thresholds for the detection of inharmonicity in complex tones," J. Acoust. Soc. Am. **77**, 1861–1867.
Moore, B. C., Glasberg, B. R., and Peters, R. W. (**1986**). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.
Müller, M. (**1996**). "The cochlear place-frequency map of the adult and developing Mongolian gerbil," Hear. Res. **94**, 148–156.
Patterson, R. D. (**1987**). "A pulse ribbon model of monaural phase perception," J. Acoust. Soc. Am. **82**, 1560–1586.
Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). "Time domain modelling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.
Plassmann, W., Peetz, W., and Schmidt, M. (**1987**). "The cochlea in Gerbilline rodents," Brain Behav. Evol. **30**, 82–101.
Pressnitzer, D. and Patterson, R. D. (**2001**). "Distortion products and the perceived pitch of harmonic complex tones," in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonoven (Shaker, Maastricht), pp. 97–104.
Ryan, A. (**1976**). "Hearing sensitivity of the Mongolian gerbil, *Meriones unguiculatis*," J. Acoust. Soc. Am. **59**, 1222–1226.
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.
Simmons, A. M., Reese, G., and Ferragamo, M. (**1993**). "Periodicity extraction in the anuran auditory nerve II. Phase and temporal fine structure," J. Acoust. Soc. Am. **93**, 3374–3389.
Sinnott, J. M., and Mosteller, K. W. (**2001**). "Comparative assessment of speech sound discrimination in the Mongolian gerbil," J. Acoust. Soc. Am. **110**, 1729–1732.
Sinnott, J. M., Brown, C. H., and Brown, F. E. (**1992**). "Frequency and intensity discrimination in Mongolian gerbils, African monkeys and humans," Hear. Res. **59**, 205–212.
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.
Terhardt, E. (**1974**). "Pitch, consonance and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.
Von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).
Wever, E. G. (**1949**). *Theory of Hearing* (Wiley, New York).
Wightman, F. L. (**1973**). "The pattern-transformation model of pitch," J. Acoust. Soc. Am. **54**, 407–416.
Wright, A., Davis, A., Bredberg, G., Ulehlova, L., and Spencer, H. (**1987**). "Hair cell distributions in the normal human cochlea," Acta Oto-Laryngol., Suppl. **444**, 1–48.
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).

Critical bands and critical ratios in animal psychoacoustics: An example using chinchilla data

William A. Yost
Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

William P. Shofner
Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405-7000

(Received 11 August 2008; accepted 6 November 2008)

This paper suggests that critical ratios obtained in noise-masked tone studies are not good indicators of critical bandwidths obtained in both human and nonhuman animal subjects. A probe-tone detection study using chinchilla subjects suggests that they may be broadband processors in detection tasks as opposed to human subjects who use narrow-band, critical-band processing. If chinchilla and other nonhuman animal subjects are wideband processors, this can partially explain why their critical ratios are significantly greater than those measured in human subjects. Thus, large critical ratios obtained for nonhuman animals may indicate processing inefficiency rather than wide critical bands. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3037232]

PACS number(s): 43.66.Gf [ADP] Pages: 315–323

I. INTRODUCTION

A fundamental measure of auditory function is frequency selectivity, the ability of an animal to discern the frequency of one sound in the presence of sound(s) of different frequencies. Obtaining direct psychophysical measures of frequency selectivity [e.g., estimates of critical bandwidth (CBW)] can be time consuming, and, thus, such CBW measures of frequency selectivity have only been obtained for a few animals (see Fay, 1988). An indirect measure of frequency selectivity, the critical ratio (CR), is simple to obtain as it requires a single masked threshold measurement. The CR has been obtained for many animals (Fay, 1988). In this paper, we explain why the CR is probably a poor indicator of frequency selectivity in that the estimates of frequency selectivity derived from direct measurements (e.g., CBW) do not agree with those derived from the indirect CR method. For nonhuman animal subjects, estimates of CRs are considerably greater than those obtained for human subjects, and as such they imply that frequency selectivity in these nonhuman animal subjects would be much worse than in human subjects. The sharpness of tuning of auditory filters between human and nonhuman animals has recently been the subject of controversy in the literature (see Shera *et al.*, 2002; Ruggero and Temchin, 2005). We describe the outcome of an experiment, which might provide one reason why the CRs obtained for most nonhuman animal subjects are considerably larger than those obtained for human subjects.

The concept of the critical band and the use of the CR to estimate the CBW were developed during the 1930s and 1940s at Bell Laboratories directed by Fletcher.[1] Fletcher (1940) used a band-narrowing procedure and showed that only a "critical band" of frequencies centered on the tone to be detected affected a tone's noise-masked threshold. That is, the masked threshold of a tone remained unchanged as the bandwidth of a masking noise was decreased (holding the spectrum level of the masking noise constant) until the bandwidth was narrower than a critical value, and then the threshold of the tone decreased as the noise bandwidth was made narrower than this critical value. The masker bandwidth at which the tone's thresholds began to decrease as the bandwidth was narrowed was used as an estimate of the CBW. Fletcher (1940) showed that the CBW increased as the frequency of the signal tone increased, implying a widening of the CBW with increasing frequency. This and other observations were consistent with frequency selectivity measured at the auditory periphery, and it was assumed that the critical band reflected the biomechanical and neural processes occurring in the cochlea and auditory nerve.

Fletcher (1940) made a general assumption that a tone's masked threshold was proportional to the power of the noise within the critical band that spectrally surrounded the tone. He also observed that the power of the noise within the critical band was approximately equal to the power of the signal at the masked threshold. This second observation is often couched as a second assumption that the power of the signal at the masked threshold is equal to the power within the critical band. These two assumptions allowed Fletcher (1940) to derive the CR as an indirect estimate of the CBW. The derivation is as follows:

Assumption 1: Masked threshold is proportional to the power in the critical band:

(1) $P_t = K^* P_{cb}$, where P_t is the power of the tone at masked threshold, P_{cb} is the power within the critical band, and K is the proportionality constant.
(2) Or $P_t / P_{cb} = K$.
(3) P_{cb} can be rewritten as $P_{cb} = \mathrm{No}^* \mathrm{CBW}$, where No is the spectrum level (noise power per unit bandwidth) of the noise (which is held constant).
(4) Thus, from (2) $P_t / (\mathrm{No}^* \mathrm{CBW}) = K$.
(5) Or $\mathrm{CBW} = P_t / (K^* \mathrm{No})$.

Assumption 2: The power of the tone at the masked threshold equals the power of the noise within the critical band:

(6) $P_t = P_{cb}$.

(7) From (1) above $P_t/P_{cb}=K$.
(8) Thus, from (6) $K=1$.
(9) Then, from (5) above CBW$=P_t/$No.

The value of CBW that equals the ratio $P_t/$No, the signal-to-noise ratio at the masked threshold, is called the CR, and CR=CBW. Note that CR only requires a single estimate of the signal-to-noise ratio (expressed as the power of the tonal signal, P_t, at the masked threshold divided by the noise masker's spectrum level, No, in units of power).[2] However, for the CR to be equal to the CBW requires that both assumptions listed above are true. Using only Assumption 1, the CBW is a function of both the signal-to-noise ratio at the masked threshold and K, where K can be viewed as an efficiency value (see Patterson *et al.*, 1982). Let us examine the evidence that Assumption 2 is true. That is, the power of the signal at the masked threshold equals the power of the masking noise within the critical band, and thus $K=1$. Obtaining an estimate of K requires that the CBW is known.

Measures of the critical band in noise masking experiments have been studied extensively over the past 70 years. We will not attempt to review this history (see Moore, 1986). Scharf (1970) provided a thorough review of the critical band and CR, and he argued based on several data sets that the value of K measured for human subjects is about 0.4, and therefore the CR for human subjects is about 2.5 times the signal-to-noise ratio at the masked threshold, i.e., CR $=2.5^*P_t/$No. Perhaps Assumption 2 should be changed to CR$=2.5^*P_t/$No, and as such this revised CR ratio can be used to estimate CBW. However, Patterson *et al.* (1982) argued that the CR does not allow one to accurately estimate the CBW of elderly subjects in that the CRs obtained in this experiment were more a function of efficiency, the proportionality constant K, than they were of the CBW. Thus, the CR would not accurately estimate the CBW measured as a function of age.

We now deal with how well the CR accounts for critical bands in nonhuman animal subjects. First, almost all studies of noise masking of a tonal signal in nonhuman animal subjects indicate that the signal-to-noise ratio required for threshold performance (i.e., the CR when appropriately calculated) is substantially higher than that measured for human subjects. Figure 1 compares estimates of CRs for several mammalian species based on the summary provided by Fay (1988). Data from several mammals (Gourevitch, 1970; Green *et al.*, 1959; Miller, 1964; Nienhuys and Clark, 1979; Niemiec *et al.*, 1992; Pickels, 1975; Seaton and Trahiotis, 1975; Watson, 1963) are shown in Fig. 1. As can be seen, the CR measures for several nonhuman mammalian species are much larger than the CRs obtained for human subjects. Since the signal-to-noise ratio at threshold is used to calculate the CR as an indirect estimate of the CBW the large masked thresholds obtained for nonhuman animal subjects imply that they have wider critical bands than human subjects.

As stated previously, there are very few studies of direct measures of CBWs in nonhuman animal subjects. Niemiec *et al.* (1992) directly measured the critical band using several different methods along with the CR for the chinchilla. Table I summarizes some of these data [those using the notched-noise estimate of the equivalent rectangular bandwidth (ERB),[3] as defined in the procedures used by Glasberg and Moore (1990)], and they indicate that the value of K is greater than 1 for the chinchilla, sometimes far greater than 1. The work by Niemiec *et al.* (1992) showed that, on average, the estimates of the CBW for the chinchilla derived from direct estimates (e.g., notched-noise estimates of the ERB) were about 5%–10% larger than those measured for human subjects. On the other hand, the values of the CR obtained for the chinchilla were, on average, more than 300% larger than the CRs obtained for human subjects (Green *et al.*, 1959; see also Fig. 1). Evans *et al.* (1992) also measured critical bands for the guinea pig and found that the ERBs for the guinea pig were only about 12% larger than the ERB measured for human subjects. And, recently May *et al.* (2006) measured filter bandwidths in the mouse using a notched-noise masker and found narrower filter bandwidths than those reported by Ehret (1976) using CR estimates. Thus, it appears as if direct estimates of the critical band for some nonhuman animal subjects are much less than those based on CR measures and only somewhat larger than those estimated for human subjects, while the estimates of CRs for most mammals deviate greatly from those estimated for human subjects. Thus, the CR for nonhuman animal subjects

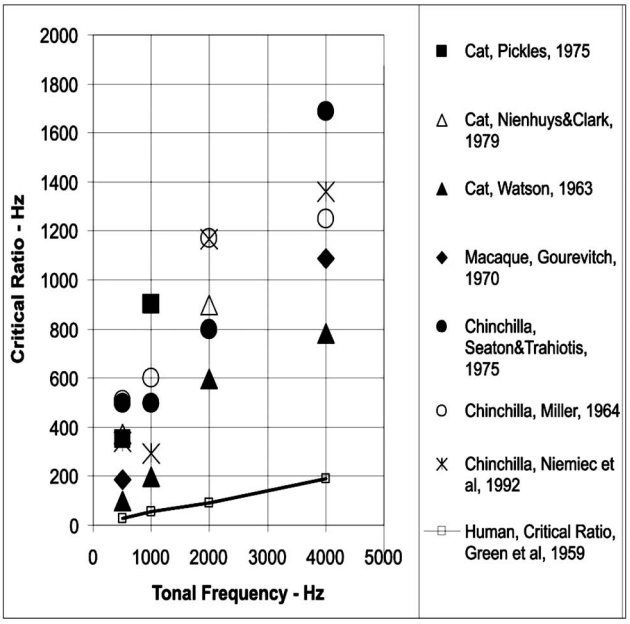


FIG. 1. CRs in Hz plotted as a function of signal (center) frequency (Hz) for several mammals. Adapted from Fay (1988).

TABLE I. CR measurements and critical band–ERB measurements for the chinchilla from Niemiec *et al.* (1992) along with the calculated value of K (see text also).

Signal frequency (Hz)	CR (Hz)	ERB (Hz)	K
500	341	59	5.8
1000	293	185	1.6
2000	1166	304	3.8
4000	1359	335	4.1

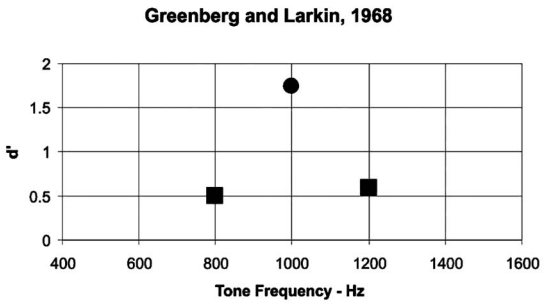


FIG. 2. Data from Greenberg and Larkin (1968) showing d' as a function of tone frequency in Hz. The circled point at 1000 Hz is the test signal, and the d' values for probe tones at 800 and 1200 Hz are shown as squares. Greenberg and Larkin (1968) plotted their data in terms of percent correct. Percent correct was converted to d' based on assuming a symmetric criterion and using the tables in Swets (1964).

may significantly overestimate the width of the actual critical band that could be directly measured in these animals.

The fact that the CR is considerably larger than the critical band in nonhuman animal subjects implies that these animals are inefficient in detecting a tonal signal in a background of noise, not necessarily that these animals have wide critical bands. In the next section we describe an experiment that suggests that chinchillas are inefficient in detecting tonal signals in a noise background because they operate like wideband[4] detectors, whereas human subjects use the narrow-band critical band for detecting noise-masked tonal signals.

A. Probe-tone experiment

In detecting a tonal signal in a background of wideband noise, a detector can maximize its sensitivity if it processes the noise in a narrow band centered on the frequency of the tonal signal (Green and Swets, 1974). That is, using a wideband process[4] is inefficient for detecting a tonal signal in a background of broadband noise. It has been shown in several experiments that human subjects appear to use the critical band for detecting tonal signals masked by wideband noise (e.g., Green, 1961).

Greenberg and Larkin (1968) measured the ability of human subjects to detect probe tones of different frequencies that were presented at levels of equal detectability but were presented infrequently and at frequencies that the subject was not trained to detect. The general paradigm is that the detectability[5] of noise-masked tones of different frequencies was obtained in a baseline experiment. In the test experiment, one of these tones with frequency (f_s) was presented as "the signal" during a practice session and on 75% of signal trials during the test condition. The remaining 25% of the signal trials contained probe tones with frequencies different from f_s but at levels that were previously determined to yield equal detectability. One example of a typical result is shown in Fig. 2. The detection of the probe tones was lower than that of the signal tone (f_s), and detection performance tended to follow a form that is consistent with a critical-band filter function. That is, the data suggested that human subjects use a filter centered on the frequency of the expected tonal signal for detection of that signal, and the level of the probe tones was attenuated by this detection filter. The filter has a bandwidth consistent with that measured for the critical band.[4] Thus, the Greenberg and Larkin (1968) probe-tone experiment reinforces the idea that human subjects process a tonal signal masked by a broadband noise using a narrow band characterized by the critical band. Several additional studies have been conducted using variations in the probe-tone technique (e.g., see Hafter and Saberi, 2001; Hafter *et al.*, 1993; Dai *et al.*, 1991; see Hafter *et al.*, 2007, for a review), and these studies consider the filter used by subjects in these experiments to be an "attention filter," reflecting the idea that a subject's attention is focused on the frequency of the tonal signal. The results from all of these experiments indicate that human subjects process masked tonal signals in a narrow band, even when the subjects are aware that several different tones may be presented during the probe-tone experiment (Hafter *et al.*, 1993).

The probe-tone technique is very similar to the stimulus generalization procedure often used in the animal behavior literature (Malott and Malott, 1970). In a stimulus generalization task, an animal subject is trained (e.g., reinforced) to respond to one value of a stimulus (e.g., one tonal frequency). During testing, sounds that differ from that used during training (e.g., test tones that differ in frequency from that used in training) are occasionally presented without reinforcement, and the responses of the animal to these infrequent test stimuli are recorded. If the animal responds to the test sounds with the same response magnitude that they responded to the training sound, the animal is described as generalizing to the test sounds. That is, the animals presumably perceive the test and training sounds as similar, leading to all sounds being responded to somewhat equally. The degree to which an animal does not generalize suggests the degree to which the animal does not perceive the test and training sounds as similar. The signal tone, f_s, in the Greenberg and Larkin (1968) paradigm is similar to a training tone in the generalization task, and the probe tones are similar to the test tones.

We used a task such as the Greenberg and Larkin (1968) procedure to test the generalization of the chinchilla to noise-masked tones of different frequencies. We intended this to be a baseline experiment for additional tests of generalization by the chinchilla as a means to explore aspects of pitch perception in the chinchilla (see, Shofner *et al.*, 2005). That is, we expected chinchilla subjects to show a lack of generalization when the test (f_s) and probe tones were considerably different in frequency, as indicated by the previous measures of CBW (Niemiec *et al.*, 1992) or even the CR. The fact that the animals showed strong generalization rather than a lack of generalization was surprising to us and led to our hypothesis that these data are evidence that the chinchilla is a wideband detector in experiments in which tones are masked by noise. Thus, we also conducted a probe-tone experiment with human subjects using the same procedure as was used for the chinchilla subjects to be able to compare performance in the probe-tone task for the two species.

II. METHOD

A. Chinchilla subjects

Participants. Four adult chinchillas (*Chinchilla lanigera*) were maintained at body weights of approximately 400–450 g. Animals were trained to detect a tone in the presence of a continuous low-level background noise. Animals were free to move around their cage that was placed inside of a single-walled sound-proof chamber, and they were not restrained in any manner. A pellet dispenser (Gerbrands G5120) with a reward chute attached to a response lever was located at one end of the cage.

Stimuli. Broadband noise was presented from a noise generator (Grayson-Stadler model 455C) and was low-pass filtered (Tucker-Davis Technologies, TDT, PF1) at a cutoff frequency of 10 kHz. Sine tones were synthesized digitally (TDT AP2) and presented through a 16-bit digital-analog converter (TDT DD1) at a conversion rate of 50 kHz. The output was low-pass filtered at 15 kHz. The 10-kHz noise was on continuously and was presented at a constant spectrum level of 10 dB [approximately 50 dB SPL (sound pressure level) over all levels], which was approximately 12 dB above the detection level of a wideband noise in the experimental setup (see Shofner *et al.*, 1993). The level of the tone was adjusted using a programmable attenuator (TDT PA4). The mixed tone and noise were played through a Realistic Minimus (3.5) loudspeaker placed near the pellet dispenser approximately 6 in. in front of the animal and approximately 30° to the right of the center. The sound level in the cage was monitored with an Ivie Technologies (1133) sound level meter.

Procedure. The behavioral task was a modified "yes/no" task. The animal initiated a trial by pressing down on the response lever. After the trial was initiated, there was a holdtime of 1.15–8.15 s, which varied randomly for each trial. At the end of the holdtime, either a tone trial or a catch trial was presented. If the animal released the lever before the random holdtime, then the trial was aborted and the countdown for the holdtime was halted. That trial began again with the next lever press using the same holdtime. For a tone trial, two bursts of the tone signal were presented in the presence of the background noise masker; the tones were 500 ms in duration shaped with 10-ms rise/fall times, presented with an interburst interval of 500 ms and a 500-ms interval after the second tone presentation. During a catch trial, only the continuous background noise was presented. The entire duration of each tone or catch trial was 2.0 s. The response window coincided with the duration of the trial but began 150 ms after the onset of the first tone burst. The resulting response window was therefore 1.85 s. When the animal released the lever during the response window, the response was recorded as a "yes." A "yes" response during a tone trial was treated as a hit, and that during a catch trial was treated as a false alarm. If the animal continued to hold down the lever throughout the response window, the response was treated as a "no." A 'no' response during a catch trial was treated as a correct rejection and as a miss during a tone trial. Chinchillas received food pellets (Noyes: Formula N) for hits and correct rejections only.

Psychometric functions for tones of 500, 1000, and 2000 Hz were obtained using a method of constant stimuli. Animals ran in a block of 40 trials and generally completed two to five blocks per day. Within a block of 40 trials, eight different tone levels in 5-dB steps were randomly presented. Each level was presented four times, and there were eight catch trials per block (20% of the trials). Each psychometric function was based on at least 50 blocks of data. The psychometric functions obtained at all three frequencies allowed us to ensure that all tones presented during the probe-tone test experiment were equally detectable.[5]

After all of the data had been collected for the psychometric functions, animals ran in the probe-tone paradigm. During testing using the probe-tone method, the broadband noise was again presented continuously at a 10-dB spectrum level (No), and the duration, shaping, and frequency of the tones as well as the response window and variable holdtime were the same as previously described. The signal tone (f_s) was 1000 Hz, and the probe tones were 500 and 2000 Hz. The sound level of all three tones was fixed to be equally detectable based on the psychometric functions. For animals C16, C18, and C24 the levels were chosen to give d' of 1.3, while for animal C12 the levels were set to yield a d' of 1.1. The value of d' was computed from the hit and false alarm rates obtained from each animal and each condition as the z-score associated with the hit rate minus that from the false alarm rate (see Green and Swets, 1974).

A block of 40 trials contained 32 tone trials (80%) and 8 catch trials (20%). Of the 32 tone trials, 24 (60% of 40 and 75% of 32) contained the signal tone (f_s=1000 Hz), 4 (10% of 40 and 12.5% of 32) contained the 500-Hz probe tone, and 4 (60% of 40 and 75% of 32) contained the 2000-Hz probe tone. All tones appeared randomly throughout a block. Any correct detection of the tone during tone trials was counted as a hit, but only hits for the signal tone (f_s=1000 Hz) were rewarded with a food pellet. Hits for the probe tones of 500 and 2000 Hz were not rewarded. Correct rejections for catch trials were also rewarded with a food pellet. Detection performance to signal and probe tones was measured using d', as described previously. The false alarms for the signal and probe trials were based on the same values obtained by Greenberg and Larkin (1968).

B. Human subjects

Participants. Three college students (one male and two females ranging in age from 21 to 25) served as subjects. They all reported that they had normal hearing and were experienced in performing in psychoacoustic tasks. They were paid $10/h for their participation in the experiment. Since it is not possible to have the human subjects run in the small animal sound-proof chambers, the human subjects ran in a single-walled sound-proof room that had about the same attenuation characteristics as the single-walled animal chambers.

Stimuli. The equipment, stimuli, and stimulus conditions used for the human subjects were very similar to those used for the chinchilla subjects, except that a TDT waveform generator (WG2) was used instead of the Grayson Stadler noise

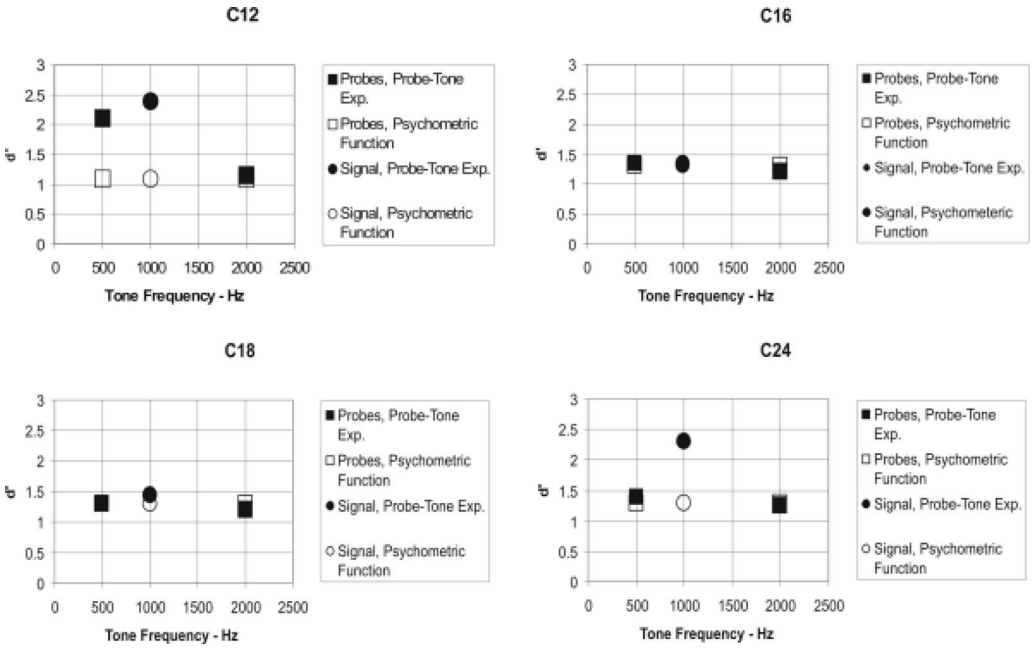


FIG. 3. Probe-tone experimental data for four chinchilla subjects (C12, C16, C18, and C24) showing d' as a function of tone frequency in Hz. Circled data points at 1000 Hz are for the signal frequency, and the squares are for the 500- and 2000-Hz probe frequencies. Open symbols are for estimates from the psychometric-function baseline experiment, and closed symbols are for the probe-tone test experiment.

generator. A Realistic Minimus (3.5) loudspeaker was placed near the subject's head approximately 22 in. in front of the subject and approximately 30° to the right of the center (similar to what was done for the chinchilla subjects). The sound levels at the place of the middle of the subject's head were measured with an Ivie Technologies (IE-33) sound level meter.

Procedure. The psychometric procedures and analyses used for the human subjects were the same as those used for the animal subjects with the following exceptions. A computer mouse was used by the subject to enter responses. As for the chinchilla subjects, the human subjects initiated a trial by clicking on a box on a computer terminal that said "Ready." Following the random holdtime and during the response window, a release of the mouse button was used to indicate that they perceived the presence of a signal sound. Keeping the mouse button down during the response window indicated that the signal was not presented. Eight point psychometric functions (varying from a signal-to-noise ratio of 8–22 dB in 2-dB steps; these values were determined from preliminary data) were first obtained for all subjects at the three frequencies, and then the level required for a d' of 1.3 (calculated in the same way as explained above) was used in the probe-tone test experiment. All other procedures for obtaining the psychometric functions for the human subjects were the same as for the chinchilla subjects.

During the probe-tone test experiment, each block of 40 trials began with three presentations of the 500-ms 1000-Hz signal tone presented at a 40-dB signal-to-noise ratio. Subjects were instructed that the object of the experiment was to be as correct as possible in detecting the presence of the signal when it was presented and in detecting its absence when it was not presented. Unlike in the chinchilla study, no feedback was provided, as was consistent with the procedures used by Greenberg and Larkin (1968). If feedback is viewed as similar to food reward, it is not possible to provide feedback only for hits during signal trials and for all correct rejections (as was done with food reinforcement for the chinchilla subjects) since this would mean that feedback would not be provided on some trials (probe-tone trials). The presentation of the 1000-Hz signal before each block is similar to a procedure used by Greenberg and Larkin (1968) and was intended to serve some of the functions of using food reward to "focus attention" on 1000 Hz. All other procedures were the same for the human subjects as they were for the chinchilla subjects.

III. RESULTS

The major results obtained from four chinchillas are shown in Fig. 3 as d' for the signal-tone (circles) and probe-tone (squares) frequencies. The open symbols are the values of d' obtained from the psychometric functions during the initial baseline experiments used to determine the signal-to-masker-noise ratio at each tonal frequency required for d' of 1.1 for C12 and of 1.3 for the other three animals. The closed symbols represent d' performance during the probe-tone test experiment. As can be seen, all four animals responded to all tones with values greater than $d'=1$, and they generalized to the probe tones with about the same level of performance (d') as they did to the test tone. The data also indicate that two animals (C12 and C24) performed with a higher d' during the test experiment than they did during the initial baseline experiment.

Figure 4 shows the psychometric functions for the four animals measured for f_s of 1000 Hz. All psychometric functions obtained at all frequencies were monotonically increasing functions, indicating improved detection with increasing

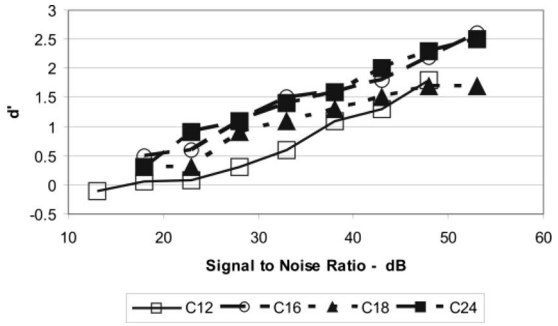


FIG. 4. Psychometric functions for the four chinchilla subjects (C12, C16, C18, and C24) showing d' as a function of signal-to-noise ratio (dB) for the 1000-Hz tonal signal.

signal-to-noise ratio. Psychometric functions obtained at the probe frequencies were very similar to those obtained at f_s, although different signal-to-noise ratios were required to obtain d' of 1.1 and 1.3. In all conditions the false alarm rate was low (maximally approximately 20%) and relatively consistent from animal to animal and over blocks of trials, and it decreased as the signal level increased. These data also indicate that the signal-to-noise ratio for a d' of 1.0 is between 25 and 33 dB, which is consistent with the results from other studies of the CR in chinchilla (e.g., Niemiec *et al.*, 1992; Miller, 1964; Seaton and Trahiotis, 1975; see also Fig. 1).

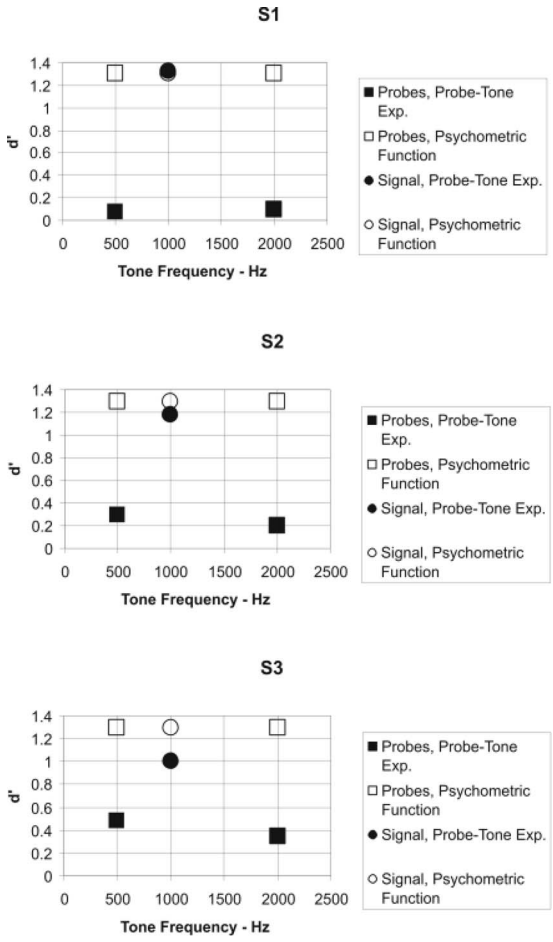


FIG. 5. Data from the three human subjects (S1, S2, and S3) using the same format as was used in Fig. 3.

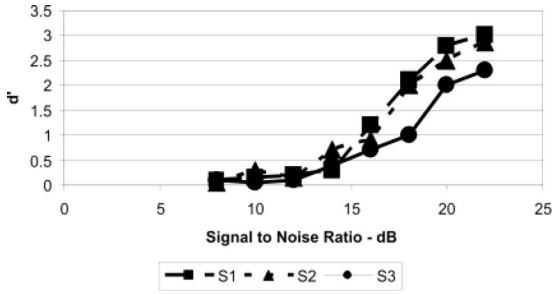


FIG. 6. Data from the three human subjects (S1, S2, and S3) using the same format as was used in Fig. 4.

Figures 5 and 6 display the data for the three human subjects using the same formant as was used for Figs. 3 and 4. As can be seen, the human subjects responded to the two probe tones (500 and 2000 Hz) with much lower values of d' during the probe-tone experiment than they did in the baseline experiment (Fig. 5), indicating a clear difference from the performance of the chinchilla subjects. The psychometric functions at 1000 Hz indicate that $d'=1.0$ was obtained at signal-to-noise ratios between 15 and 18 dB, which is higher than that often obtained for human subjects (e.g., Reed and Bilger, 1973). Also, the psychometric functions for the human subjects (Fig. 6) were considerably steeper than those for the chinchilla subjects (Fig. 4), and their slopes are consistent with those found in the literature (e.g., Green, 1961).

IV. DISCUSSION

The major results shown in Figs. 3 and 5 are that the chinchilla subjects generalized to the probe tones with approximately the same level of performance (d') as they did to the test tone, but the human subjects did not. We consider this evidence that the chinchilla is not a narrow-band detector, using something like the critical band, in detecting a tonal signal masked by a broadband noise. The fact that two animals responded at higher values of d' during testing than they did in the initial baseline experiment in which the psychometric functions were obtained might be due to an underestimation of the sensitivity of these animals to the signal tone during the baseline psychometric-function experiment. To test this hypothesis we ran all animals on an additional method of constant stimuli procedure to re-estimate the psychometric functions. Two animals (C12 and C24, but not C16 and C18) performed with higher values of d' during this second constant stimulus test experiment than they did during the initial baseline experiment, suggesting that indeed the original estimate of performance during the baseline experiments was too low for these two animals. The values of d' obtained during the retest of the psychometric-function experiment were very similar to those obtained in the probe-tone test experiment. We also compared the psychometric functions obtained during the first ten blocks of trials during the baseline psychometric-function experiment with those obtained during the last ten blocks. Again, only animals C12 and C24 showed improved performance between the first and last ten blocks, also suggesting that these animals were still improving in their performance. A similar comparison of the first ten blocks of trials to the last ten blocks during the

follow-up psychometric-function experiment did not reveal any differences in performance for all four animals. While all of the data for C12 and C24 shown in Figs. 3 and 4 may not accurately represent their absolute level of detection, the comparison of performance between probe tone and signal tones shown in Fig. 3 is similar to that from the other two animals in indicating broadband processing in a noise-masked tone detection task.

The retest of the psychometric functions indicated that at least for two of the chinchillas the estimate of thresholds is reliable. And even for the two animals whose initial threshold estimates appeared high, their thresholds appeared to be stable after the probe-tone test experiment. Since the thresholds obtained in this study are well within the range of thresholds measured for chinchilla subjects in other experiments (see Fig. 1), the present results would seem to be valid estimates of chinchilla subjects' performance in this masked threshold experiment.

Since the performance for the signal frequency (1000 Hz) for the three human subjects did not differ between the probe-tone experiment and the baseline psychometric-function experiment (Fig. 5), we did not obtain an additional estimate of the psychometric functions as we did for the chinchilla subjects. The higher signal-to-noise ratios obtained for the human subjects compared to those reported in other studies are probably due to the psychophysical procedure, which is substantially different from that typically used to estimate masked thresholds in human subjects. It is the case that the threshold signal-to-noise ratios are still, on average, approximately 12 dB lower for the human subjects than for the chinchilla subjects. This is consistent with the data in the literature (e.g., see Fig. 1).

Perhaps the chinchilla subjects generalized to the probe tones because of their prior experience with these tones during the baseline psychometric-function experiment. In all of the past probe-tone experiments with human subjects, including the one conducted for this paper, the human subjects have also had prior experience listening to the probe tones before being tested in the probe-tone test experiment.[4] In some studies the subjects are aware that information other than that provided by the signal may occur during an experimental run (Hafter *et al.* 1993). Despite these experiences, in none of the past experiments do human subjects generalize to the probe tones at anywhere near the level that the chinchilla subjects did in the chinchilla experiment reported in this paper. One explanation for these differences that we feel is worth pursuing is that the chinchilla subjects are wideband processors in these tasks and human subjects are narrowband processors, although we cannot completely rule out other reasons why chinchilla subjects generalize and human subjects do not.

If the data in Fig. 3 indicate broadband processing in noise-masked tone detection tasks, then this could explain why the CR is large in these animals. That is, chinchilla subjects require a larger signal-to-noise ratio for tonal detection than human subjects because chinchilla subjects are wideband processors, while human subjects are narrowband, critical band processors in noise-masked tone detection tasks. Green (1961) calculated that wideband detection could increase the signal-to-noise ratio by as much as 10 dB over that achieved with narrow-band processing. The results of the present study suggest that the chinchilla subjects required about 12-dB larger signal-to-noise ratio (CR) to detect the tonal signals than did the human subjects. However, if the critical bands for the chinchilla are 10%–12% larger than those for humans as the data of Niemiec *et al.* (1992) and Evans *et al.* (1992) suggest, then the difference due to wideband listening for the chinchilla subjects could be less than 10 dB. This difference is roughly consistent with Green's calculations of the decrement in signal-to-noise ratio that might occur if a wideband listening strategy was employed. While there may be other reasons why the chinchilla subjects perform poorer in detecting a noise-masked tonal signal than human subjects, the results of the present study appear consistent with the hypothesis that a contributing factor for the chinchilla subjects' poorer performance is that chinchillas are "wideband processors." Thus, it is possible that chinchilla subjects use, for whatever reason, wider bands to detect tonal signals masked by broadband noise than do human subjects. The same may be true for other nonhuman animal subjects whose CRs are much larger than those measured in human subjects. It is also interesting to note that human infants respond to probe tones with higher response rates than do adult humans, indicating that infants process tones-in-noise with a wideband listening strategy (see Bargones and Werner, 1994).

Greenwood (1990) argued that the larger CRs reflect wider critical bands in many nonhuman animals, and the differences in CR estimates of critical bands can be explained on the basis of the length of the cochlea in these animals and his assumptions regarding the relationship between the cochlear map of frequency and frequency selectivity. Greenwood (1990) produced a compelling series of detailed articles and arguments supporting these relationships. His work implies that frequency selectivity in many nonhuman animal subjects is somewhat accurately represented by the CR.

The analysis presented in this paper argues that the CR is not a good indicator of frequency selectivity. Thus, it is worth asking if the critical-band estimates of Niemiec *et al.* (1992), Evans *et al.* (1992), and May *et al.* (2006) are more indicative of frequency selectivity than the CR. Since several different procedures were used in several of these studies to directly estimate the critical band and since these procedures produced similar estimates of the critical band, which were much narrower than the CR estimates, these critical-band estimates appear to be reliable and valid.

One can use measures other than the critical band or ratio to infer an animal's frequency-selectivity ability. Measures of tonal frequency discrimination produce results that suggest that nonhuman animal subjects are poorer at tonal frequency discrimination than are human subjects (see Fay, 1988), which might imply that nonhuman animals are poorer at frequency resolution than human subjects. However, frequency discrimination is known to be variable across human subjects (Jesteadt and Bilger, 1974), so it is not clear how much difference there may be between a human subject who is poor at frequency discrimination and a nonhuman animal's

frequency discrimination ability. Moreover, frequency discrimination performance may reflect more than processing based on frequency selectivity (e.g., temporal processing; see Dye and Hafter, 1980). Tone-on-tone psychophysical simultaneous masking functions have been obtained for several nonhuman animal species. In general, the shapes of these psychophysical tonal masking tuning curves obtained for nonhuman animal subjects are similar to those that have been measured for human subjects (see Fay, 1988). This would imply that nonhuman animal and human subjects have similar frequency selectivities. Thus, it is our view that the direct measurements of frequency selectivity do suggest that frequency selectivity in the chinchilla is only slightly worse than that measured for human subjects, while the CR measurements differ greatly.

Why might the chinchilla (and perhaps other animals) use wider band processing in detecting a tonal signal masked by a noise masker? Both positive [e.g., food reinforcement used by Niemiec *et al.* (1992)] and negative reinforcers [e.g., shock avoidance used by Seaton and Trahiotis (1975)] have been used to estimate the CR and both types of motivation produce approximately the same CRs. Thus, it is unlikely that performance in these detection tasks is related to the type of motivation used in psychophysical tasks. In most studies the animal subjects appear to have reached asymptotic performance during training, so it is unlikely that additional training might improve performance. Thus, we do not have a compelling explanation as to why chinchilla subjects appear to be wideband detectors and human subjects appear to be narrow-band, critical-band detectors. It might be possible to induce narrow-band processing in the chinchilla by using some sort of cuing technique. Such cuing has been shown to improve performance in some detection tasks using human subjects (Richards and Neff, 2004). It is also possible that animals such as the chinchilla process signals as a wideband processor in order to not "miss" potential signals that might indicate a predator. That is, they may trade sensitivity for detecting a particular signal for an ability to process many different signals.

If the chinchilla and other nonhuman animal subjects are not using a narrow-band critical band to process a noise-masked tone, then the first assumption of Fletcher (1940) does not appear to hold for these animals. That is, the detection of a tone masked by a broadband noise may not appear to depend in any direct way on a critical band. While such nonhuman animal subjects show evidence of critical bands with bandwidths similar to those of human subjects, they may not use them in the same way as human subjects when they are asked to detect sinusoids in noisy environments.

V. CONCLUSIONS

Comparison of CR measures as an indirect indication of frequency selectivity to the direct measures of the critical band suggests that CRs are probably not a good indicator of frequency selectivity as measured in both human and nonhuman animal subjects.

The probe-tone masking experiment suggests that it is possible that the large CRs measured in nonhuman animal subjects, such as the chinchilla, are due to these animals using a broader band for signal processing in noise-masked detection tasks than human subjects use.

ACKNOWLEDGMENTS

The chinchilla experiment was conducted while both authors were at the Parmly Hearing Institute, Loyola University Chicago, Chicago, IL. The experiment with human subjects was conducted by W.A.Y. at Parmly after W.P.S. had left Parmly for Indiana University. The work was supported by a NIDCD Program Project Grant awarded to Loyola. The writing of this paper was supported by a NIDCD grant awarded to W.A.Y. at ASU.

[1]Rankovic, C. M., and Allen, J. B., Study of Speech and Hearing at Bell Telephone Laboratories, CD-ROM from the Acoustical Society of America (www.asa.aip.org), Melville, NY.

[2]Since No, spectrum level, is noise power per unit bandwidth, No actually has energylike units (power × time). In decibels, $10 \log \mathrm{CBW} = P_t$ in dB—No in dB.

[3]The term critical band is often used in a general sense to indicate a spectral region critical for determining masked threshold. Measures such as the ERB are specific calculations based on particular experimental and measurement procedures. Thus, the ERB is one of several ways to estimate the CBW and is not necessarily synonymous with the critical band.

[4]Wideband processing can mean one of two things in this paper. It could mean that the animal actually uses a wider band. Or it could mean that the animal uses several different narrow-band critical bands and sums or samples the outputs of these bands in making a detection decision. In either case, the signal level required for a constant level of performance would be greater than that required if the subject used a single narrow band (Green, 1961).

[5]Thresholds for all tones (signal and probes) are obtained in order to ensure that all tones are equally detectable in the probe-tone test experiment. Without such thresholds, performance for the probe tones may be based only (or partially) on their overall detectability rather than on narrow-band versus wideband processing.

Bargones, J. Y., and Werner, L. A. (**1994**). "Adults listen selectively: Infants Do Not," Psychol. Sci. **5**, 170–174.

Dai, H., Scharf, B., and Buus, S. (**1991**). "Effective attenuation of signals in noise under focused attention," J. Acoust. Soc. Am. **89**, 2837–2842.

Dye, R. H., Jr., and Hafter, E. R. (**1980**). "Just-noticeable differences of frequency for masked tones," J. Acoust. Soc. Am. **67**, 1746–1753.

Ehret, G. (**1976**). "Critical bands and filter characteristics of the ear of the house mouse (Mus musculus)," Biol. Cybern. **24**, 35–42.

Evans, E. F., Pratt, S. R., Spenner, H., and Copper, N. P. (**1992**). "Comparisons of physiological and behavioural properties: Auditory frequency selectivity," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horne (Pergamon, Oxford), pp. 159–170.

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay Associates, Winnetka, IL).

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gourevitch, G. A. (**1970**). "Detectability of tones in quiet and in noise by rats and monkeys," in *Animal Psychophysics: The Design and Conduct of Sensory Experiments*, edited by W. C. Stebbins (Appleton-Century-Crofts, New York), pp. 67–97.

Green, D. M. (**1961**). "Detection of auditory sinusoids of uncertain frequency," J. Acoust. Soc. Am. **33**, 897–903.

Green, D. M., Mckey, M. J., and Licklider, J. C. R. (**1959**). "Detection of a pulsed sinusoid in noise as a function of frequency," J. Acoust. Soc. Am. **31**, 1446–1452.

Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, New York).

Greenberg, G. Z., and Larkin, W. (**1968**). "Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method," J. Acoust. Soc. Am. **44**, 5513–5520.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for sev-

eral species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Hafter, E. R., Sarampalis, A., and Loui, P. (**2007**). "Auditory attention and filters," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 115–143.
Hafter, E. R., and Saberi, K. (**2001**). "A level of stimulus representation model for auditory detection and attention," J. Acoust. Soc. Am. **110**, 1489–1497.
Hafter, E. R., Schlauch, R. S., and Tang, J. (**1993**). "Attending to auditory filters that were not stimulated directly," J. Acoust. Soc. Am. **94**, 743–747.
Jesteadt, W., and Bilger, R. C. (**1974**). "Intensity and frequency discrimination in one- and two-interval paradigms," J. Acoust. Soc. Am. **55**, 1266–1276.
Malott, R. W., and Malott, M. K. (**1970**). "Perception and stimulus generalization," in *Animal Psychophysics: The Design and Conduct of Sensory Experiments*, edited by W. C. Stebbins (Appleton-Century-Crofts, New York), pp. 363–400.
May, B. J., Kimar, S., and Prosen, C. A. (**2006**). "Auditory filter shapes of CBA/CaJ mice: Behavioral assessments," J. Acoust. Soc. Am. **120**, 321–330.
Miller, J. D. (**1964**). "Auditory sensitivity of the chinchilla in quiet and in noise," J. Acoust. Soc. Am. **36**, 2010–2016.
Moore, B. C. J. (**1986**). *Frequency Selectivity in Hearing* (Academic, London).
Niemiec, A. J., Yost, W. A., and Shofner, W. P. (**1992**). "Behavioral measures of frequency selectivity in the chinchilla," J. Acoust. Soc. Am. **92**, 2636–2649.
Nienhuys, T. W., and Clark, G. M. (**1979**). "Critical bands following the selective destruction of the cochlear inner and outer hair cells," Acta Oto-Laryngol. **88**, 350–358.
Patterson, R. D., Nimo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.
Pickels, J. O. (**1975**). "Normal critical bands in the cat," Acta Oto-Laryngol. **80**, 245–254.
Reed, C. M., and Bilger, R. C. (**1973**). "A comparative study of S/No and E/No," J. Acoust. Soc. Am. **53**, 1039–1044.
Richards, V. M., and Neff, D. L. (**2004**). "Cuing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.
Ruggero, M. A., and Temchin, A. N. (**2005**). "Unexceptional sharpness of frequency tuning in the human cochlea," Proc. Natl. Acad. Sci. U.S.A. **102**, 18614–18619
Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. **1**, pp. 157–202.
Seaton, W. H., and Trahiotis, C. (**1975**). "Comparison of critical ratios and critical bands in the monaural chinchilla," J. Acoust. Soc. Am. **57**, 193–199.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Shofner, W. P., Whitmer, W. M., and Yost, W. A. (**2005**). "Listening experience with iterated rippled noise alters the perception of 'pitch' strength of complex sounds in the chinchilla," J. Acoust. Soc. Am. **118**, 3187–3197.
Shofner, W. P., Yost, W. A., and Sheft, S. (**1993**). "Increment detection of bandlimited noises in the chinchilla," Hear. Res. **66**, 67–80.
Swets, J. A. (**1964**). *Signal Detection and Recognition by Human Observers* (Wiley, New York).
Watson, C. S. (**1963**). "Masking of tones by noise for the cat," J. Acoust. Soc. Am. **35**, 167–172.

Pitch discrimination interference: The role of ear of entry and of octave similarity

Hedwig E. Gockel[a)]
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Ervin R. Hafter
Department of Psychology, University of California at Berkeley, 3210 Tolman Hall, Berkeley, California 94720

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 20 June 2008; revised 22 September 2008; accepted 13 October 2008)

Gockel *et al.* [(2004). J. Acoust. Soc. Am. **116**, 1092–1104] reported that discrimination of the fundamental frequency (F0) of two sequentially presented complex tones (the target) was impaired when an additional complex tone (the interferer) was presented simultaneously with and to the same ear as the target, even though the target and interferer were filtered into separate frequency regions. This pitch discrimination interference (PDI) was greatest when the target and interferer had similar F0s. The current study examined the role of relative ear of entry of the target and interferer and whether the dependence of the PDI effect on the relative F0 of target and interferer is based on pitch height (F0 as such) or pitch chroma (the musical note). Sensitivity (d') was measured for discrimination of the F0 of a target with a nominal F0 of 88 Hz, bandpass filtered from 1375 to 1875 Hz. The interferer was bandpass filtered from 125 to 625 Hz. The contralateral interferer produced marked PDI, but smaller than for ipsilateral presentation. PDI was not larger when the interferer's F0 was twice the nominal target F0 than when it was a factor of 1.9 or 2.1 higher.
 [DOI: 10.1121/1.3021308]

PACS number(s): 43.66.Hg, 43.66.Fe, 43.66.Rq [RLF] Pages: 324–327

I. INTRODUCTION

In the auditory system, complex sounds are filtered into frequency channels or bands by the action of the basilar membrane. It is commonly assumed that it is possible to attend selectively to single channels or combinations of channels so as to achieve good performance in detection and discrimination tasks (Fletcher, 1940; Moore *et al.*, 1997). However, under some conditions, it appears to be difficult or impossible for human listeners to process the outputs of some channels while ignoring others. This is manifested in effects such as across-channel masking (Schooneveldt and Moore, 1987; Moore *et al.*, 1990), modulation detection interference for amplitude and/or frequency modulated tones (Yost and Sheft, 1989, 1990; Moore *et al.*, 1991), pitch discrimination interference (PDI) (Gockel *et al.*, 2004, 2005), interference with detection of an increment in level of a tone by amplitude modulation of a spectrally distant tone (Gallun and Hafter, 2006), and disruption of intensity discrimination by a spectrally remote sound that is roved in level (Buss, 2008).

In the case of PDI, discrimination of the fundamental frequency (F0) of a target complex tone that is filtered into a restricted spectral region can be impaired by the presence of an interfering complex tone with fixed F0 that is filtered into a different spectral region (Gockel *et al.*, 2004, 2005). The effect is greatest when the F0 of the interferer is close to the (mean) F0 of the target. Furthermore, the effect is greater when the interferer has a high pitch salience than when it has a low pitch salience. For example, a strong PDI effect is obtained when the target contains only unresolved components, while the interferer contains resolved components (Gockel *et al.*, 2004). Part of the PDI effect may be explicable in terms of perceptual grouping of the target and interferer, which would be stronger when their F0s were similar (Broadbent and Ladefoged, 1957; Brokx and Nooteboom, 1982). This could explain the importance of the relative pitch salience of the target and interfering sounds. If the two sounds are perceptually fused, the higher the pitch salience of the interferer with fixed F0 the more it will "dilute" the perceived pitch change produced by the change in F0 of the target, leading to impaired F0 discrimination of the target. However, perceptual grouping probably cannot explain the whole of the PDI effect since PDI occurs even under conditions where strong cues are available to promote perceptual segregation of the target and interferer. For example, PDI is reduced, but not eliminated, by gating the interferer on before and off after the target, or by presenting the interferer continuously (Gockel *et al.*, 2004).

In the present paper, two factors that might affect PDI are explored. The first factor is the ear of entry of the target

[a)]Author to whom correspondence should be addressed. Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk. Data were collected while this author was working in the Department of Psychology, University of California at Berkeley, 3210 Tolman Hall, Berkeley, CA 94720.

and interferer. Presentation of two sounds, one to each ear, can provide a powerful cue for perceptual segregation of the sounds (Cherry, 1953; Hartmann and Johnson, 1991), especially when the sounds do not overlap spectrally (Darwin and Hukin, 2004). The target and interferer used here were filtered into two spectral regions separated by one octave, and their F0 was never identical. When these stimuli were presented dichotically (target to one ear and interferer to the other), subjects reported hearing two independent sound sources, one at each ear. In contrast, when two tone complexes have the same F0 and the complexes are filtered into two separate spectral regions with overlapping filter slopes, listeners report a single sound source coming from the center of the head (Broadbent and Ladefoged, 1957). If perceptual grouping is important for PDI, then presentation of the target and interferer to opposite ears (called the dichotic condition) should lead to a strong reduction of PDI relative to the case where the target and interferer are presented to the same ear. However, if some PDI persisted in the dichotic condition, this would provide evidence to support the idea that PDI does not depend entirely on perceptual grouping. Furthermore, the presence of PDI in the dichotic condition would indicate that at least part of the PDI effect occurs at a level in the auditory system where there is significant binaural interaction.

The second factor explored in the present paper is concerned with the relative F0s of the target and interferer. Gockel *et al.* (2004) argued that similarity in pitch of the target and interferer is an important factor influencing the amount of PDI. However, pitch has two dimensions, pitch height and pitch chroma (Bachem, 1950). Two tones whose F0s are separated by an octave have different pitch heights but the same pitch chroma. The question addressed here was: is the similarity of pitch that governs PDI based on pitch height, pitch chroma, or a combination of the two? To address this question, PDI was measured for a wide range of F0 relationships between the target and interferer, including an octave relationship.

II. METHOD AND STIMULI

A. Stimuli

In a two-interval two-alternative forced choice (2AFC) task, subjects had to discriminate between the F0s of two sequentially presented target complex tones with a nominal F0 of 88 Hz and a fixed difference, $\Delta F0$, between the F0s of the two tones within a trial. In one, randomly chosen, interval, the target complex had an F0 equal to $F0-\Delta F0/2$, while in the other interval its F0 was $F0+\Delta F0/2$. The value of $\Delta F0$ was selected and fixed for each subject based on pilot data, so that in the easiest condition, i.e., in the absence of an interferer, the sensitivity index, d', was between 1.5 and 2.0. The following values for $\Delta F0$ were used: 3.5% for two subjects, 5.0% for three subjects, and 7.0% for one subject. The target was presented either alone or with a synchronously gated harmonic complex (the interferer), which had an F0 that was the same in the two intervals of a trial. The F0 of the interferer was either equal to the arithmetic mean of the actual F0s of the two target complexes in the two intervals or increased by various amounts (15%, 45%, 90%, 100%, or 110%). Thus, the interferer's F0 was never identical to the F0 of the simultaneously presented target complex. The target was always presented to the left ear, while the interferer was presented either ipsilaterally or contralaterally to the target. Each harmonic complex was bandpass filtered (slopes of 48 dB/octave) into one of two frequency regions. The target was bandpass filtered into a "MID" region with nominal cutoff frequencies (3-dB down points) of 1375 and 1875 Hz, while the interferer was filtered into a "LOW" region with nominal cutoff frequencies of 125 and 625 Hz. Thus, the interferer would contain resolved harmonics (Plomp, 1964; Plomp and Mimpen, 1968; Moore and Ohgushi, 1993; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003), while the target would contain only unresolved harmonics. These frequency regions were chosen following Carlyon and Shackleton (1994), and the stimuli were similar to those used in experiment 1 of Gockel *et al.* (2004), except that in the previous study the target complex was bandpass filtered between 1375 and 15 000 Hz. In the present study, the bandwidth of the target was chosen to be identical to that of the interferer because we wanted to produce a large PDI effect (when the target and interferer had equal F0s) to maximize the chance of observing differences between PDIs for interferers whose F0s were widely separated from the nominal target F0.

The nominal F0 was randomly varied between trials over the range of $\pm 10\%$. The filter cutoffs for both the LOW and the MID region were varied by the same factor in order to keep fixed the degree of resolvability of the components. This F0 randomization encouraged subjects to compare the pitch of the two target complexes in each trial and discouraged them from basing their decision on a long-term memory representation of the sound. All complex tones had a root-mean-square (rms) level of 53 dB sound pressure level (SPL), and components were added in sine phase. The stimulus duration was 400 ms, including 5-ms raised-cosine onset and offset ramps. The silent interval between the two stimuli within a trial was 500 ms. To mask possible distortion products, a continuous pink background noise, spectrally limited between 2 and 1512 Hz (corresponding to 1.1×1375 Hz), was presented diotically in all conditions. Its overall rms level was 49.2 dB SPL. Its rms level in the region from 125 to 625 Hz, the nominal frequency band covered by the interferer, was 10 dB below that of the interferer.

Tones were generated and bandpass filtered digitally. They were played using a Turtle Beach Santa Cruz sound card (20-bit digital-to-analog conversion) with a sampling rate of 44.1 kHz. Stimuli were presented using Sennheiser HD580 headphones. Subjects were seated individually in a sound-attenuating booth.

B. Procedure

Subjects had to indicate the interval in which the target sound had the higher F0. They were requested to focus attention on the target and to ignore the interferer as much as possible when it was present. Each interval was marked by a light, and visual feedback was provided after each trial. If the

interferer was present, it was present in both intervals of the 2AFC task. The amount by which the interferer's F0 was increased above that of the nominal target F0 was fixed within a block of 105 trials; the first five trials in each block were considered as "warm-up" trials and were discarded. The upward shift of the interferer's F0 was pseudorandomized between blocks. One block was run for each condition in turn before additional blocks were run in any other condition, with one exception: usually a block of "target alone" trials was run first, so that subjects were familiar with the timbre of the target. This was followed by a block with the interferer present. The relative ear of presentation of the interferer was fixed within a session; ipsilateral and contralateral conditions were run in alternating repeated sessions. The total duration of a single session was about 2 h, including rest times. For all subjects, at least 500 trials were run per condition. The data collection proper (after 6–10 h of practice) took between 15 and 22 h. To familiarize subjects with the procedure and equipment, they participated in at least three practice sessions, and up to two more if practice effects were seen. The practice covered conditions both with and without the interferer.

C. Subjects

Six subjects with normal hearing participated, one of whom was the first author. They ranged in age from 18 to 46 years. All of the subjects except one had some musical training. One had participated in earlier studies on PDI. All subjects participated in all conditions.

III. RESULTS AND DISCUSSION

The filled symbols in Fig. 1 show the mean d' values and the corresponding standard errors averaged across five subjects as a function of the ratio between the interferer's F0 and the nominal target F0. The data of one "aberrant" subject are plotted separately (open symbols) and will be described below. The upward and downward pointing triangles show results for contralateral and ipsilateral presentation of the interferer, respectively. For clarity, data for the contralateral interferer are plotted toward the left of their true values on the x-axis. Performance was best, with a d' value of 1.86, in the absence of an interferer (filled circle in the left upper corner, condition None). Performance was worst, with a d' value of 0.32, for the ipsilateral interferer with an F0 corresponding to the nominal target F0. The size of the PDI effect (d' without interferer minus d' in the presence of the interferer) for this condition was 1.54, which is substantially larger than the PDI of about 1.0 observed by Gockel *et al.* (2004) for the corresponding condition of their experiment 1. For the ipsilateral interferer whose F0 was 1.45 times the nominal target F0, PDI was about 0.7, which is substantially larger than the PDI of about 0.2 observed by Gockel *et al.* (2004) in their experiment 1 for an F0 ratio of 1.3 (the largest they employed). The larger PDI effects in the present study are probably due to the fact that here the target and interferer had equal bandwidths and overall levels, while in experiment 1 of the earlier study the level per component was the same for the target and interferer, and the target's bandwidth was a factor of 27 larger than that of the interferer.

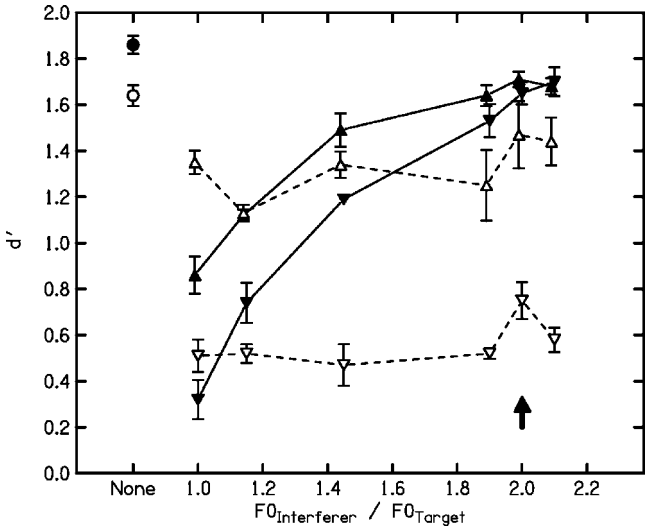


FIG. 1. Performance (d') in an F0-discrimination task, plotted as a function of the ratio between the interferer's F0 and the nominal target F0, which was 88 Hz. The target was bandpass filtered from 1375 to 1875 Hz, while the interferer was filtered from 125 to 625 Hz. The filled symbols connected by solid lines show the mean d' values for five subjects who produced very similar results and the associated standard errors across subjects. The open symbols connected by dashed lines show the mean d' values for one aberrant subject, who showed no tuning effect, and the associated standard errors across five blocks with 100 trials each. The downward and the upward pointing triangles show performance in the presence of ipsilaterally and contralaterally presented interferers, respectively. The filled and the open circles in the upper-left corner show performance in the absence of any interferer (condition None) for the five subjects and the single aberrant subject.

In the presence of a contralateral interferer, performance was clearly impaired relative to that without an interferer. However, the PDI effect was smaller than for an ipsilateral interferer. For both ipsi- and contralateral interferers, performance improved as the interferer's F0 increased. Performance leveled out when the interferer's F0 was more than 1.9 times that of the target, and there was no sign of a local increase in PDI (decrease in performance) when the interferer's F0 was twice that of the target.

One subject showed a different pattern of results from all the others. The open symbols show the mean d' and the corresponding standard errors across five blocks with 100 trials each, for this subject. While his performance was better than that of the other subjects when the interferer's F0 was equal to the nominal target F0, he showed no tuning effect for either side of presentation; i.e., his performance was independent of the F0 of the interferer relative to that of the target. This lack of tuning of the PDI effect has not been observed among the 12 subjects tested in this and previous studies.

To examine the statistical significance of the results, a repeated measures two-way analysis of variance (ANOVA) with factors side of presentation of the interferer and interferer's F0 was calculated, using the mean PDI value for each subject and condition as input. Note that the results of the ANOVA were basically unchanged when the data from the

aberrant subject were excluded. The ANOVA showed a highly significant main effect of the F0 of the interferer [$F(5,25)=23.57$, $p<0.001$], a significant main effect of side of presentation of the interferer [$F(1,5)=11.95$, $p<0.05$], and a significant interaction between side and F0 [$F(5,25)=5.31$, $p<0.05$]. To check for an "octave effect," two separate ANOVAs were calculated, one for the contralateral and one for the ipsilateral interferer, using only the PDI values for the three highest F0 values (F0 ratios of 1.9, 2.0, and 2.1). The results of both ANOVAs showed no significant main effect of the interferer's F0. Thus, PDI did not locally increase when the interferer's F0 was an octave above the nominal target F0. Planned t-tests for the three highest F0 values showed that for all cases but one (ipsilateral interferer with frequency ratio of 2.1), the observed PDI was significantly larger than zero (one-tailed, 5% error level).

In the present study, the relationship between the interferer's F0 and the nominal target F0 was varied by *increasing* (rather than *decreasing*) the interferer's F0 above that of the target. This was done to ensure that all components of the interferer were resolved in all conditions. Therefore, the current findings apply only to the situation in which the interferer's F0 is one octave above that of the target. However, there is no obvious reason why an interferer with an F0 one octave below the nominal target F0 should have a special status, differing from that of an interferer with an F0 above the nominal target F0.

IV. SUMMARY AND CONCLUSIONS

F0 discrimination between two target tones containing only unresolved harmonics was clearly impaired by the presence of an additional tone complex containing resolved harmonics. The impairment was less strong when the interferer was presented to the ear opposite to the ear receiving the target. However, in spite of the fact that ear of entry provides a strong segregation cue, a marked PDI effect was observed for the contralateral interferer. Furthermore, a small but significant PDI effect was observed even when the interferer was presented contralaterally to the target and had an F0 that was as much as a factor of 2.1 above the nominal target F0. These findings give further support to the idea that PDI can occur under conditions that are likely to lead to perceptual segregation of the target and interferer. The finding of significant PDI when the interferer was presented to the opposite ear to the target suggests that at least part of the PDI effect originates at a level of the auditory system where there is binaural interaction. No local increase in PDI was observed when the interferer's F0 was an octave above that of the target complex. Thus, the similarity in pitch which has been found to influence PDI seems to be based on pitch height rather than on pitch chroma.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. DCD 00087. We thank Anastasios Sarampalis for help with calibration and generous "sharing" of his booth and office. We would also like to thank Richard Freyman and two anonymous reviewers for helpful comments on an earlier version of this paper.

Bachem, A. (**1950**). "Tone height and tone chroma as two different pitch qualities," Acta Psychol. **7**, 80–88.

Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," J. Acoust. Soc. Am. **113**, 3323–3334.

Broadbent, D. E., and Ladefoged, P. (**1957**). "On the fusion of sounds reaching different sense organs," J. Acoust. Soc. Am. **29**, 708–710.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Buss, E. (**2008**). "Across-channel interference in intensity discrimination: The role of practice and listening strategy," J. Acoust. Soc. Am. **123**, 265–272.

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?," J. Acoust. Soc. Am. **95**, 3541–3554.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech with one and two ears," J. Acoust. Soc. Am. **25**, 975–979.

Darwin, C. J., and Hukin, R. W. (**2004**). "Limits to the role of a common fundamental frequency in the fusion of two sounds with different spatial cues," J. Acoust. Soc. Am. **116**, 502–506.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Gallun, F. J., and Hafter, E. R. (**2006**). "Amplitude modulation sensitivity as a mechanism for increment detection," J. Acoust. Soc. Am. **119**, 3919–3930.

Gockel, H., Carlyon, R. P., and Moore, B. C. J. (**2005**). "Pitch discrimination interference: The role of pitch pulse asynchrony," J. Acoust. Soc. Am. **117**, 3860–3866.

Gockel, H., Carlyon, R. P., and Plack, C. J. (**2004**). "Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms," J. Acoust. Soc. Am. **116**, 1092–1104.

Hartmann, W. M., and Johnson, D. (**1991**). "Stream segregation and peripheral channeling," Music Percept. **9**, 155–184.

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., Glasberg, B. R., Gaunt, T., and Child, T. (**1991**). "Across-channel masking of changes in modulation depth for amplitude- and frequency-modulated signals," Q. J. Exp. Psychol. **43A**, 327–347.

Moore, B. C. J., Glasberg, B. R., and Schooneveldt, G. P. (**1990**). "Across-channel masking and comodulation masking release," J. Acoust. Soc. Am. **87**, 1683–1694.

Moore, B. C. J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.

Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.

Plomp, R., and Mimpen, A. M. (**1968**). "The ear as a frequency analyzer II," J. Acoust. Soc. Am. **43**, 764–767.

Schooneveldt, G. P., and Moore, B. C. J. (**1987**). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. **82**, 1944–1956.

Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.

Yost, W. A., and Sheft, S. (**1989**). "Across-critical-band processing of amplitude-modulated tones," J. Acoust. Soc. Am. **85**, 848–857.

Yost, W. A., and Sheft, S. (**1990**). "A comparison among three measures of cross-spectral processing of amplitude modulation with tonal signals," J. Acoust. Soc. Am. **87**, 897–900.

The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians

Dee A. Nikjeh,[a] Jennifer J. Lister, and Stefan A. Frisch
Department of Communication Sciences and Disorders, University of South Florida, 4202 East Fowler Avenue, PCD 1017, Tampa, Florida 33620

(Received 26 March 2008; revised 13 October 2008; accepted 14 October 2008)

Auditory pitch discrimination and vocal pitch accuracy are fundamental abilities and essential skills of a professional singer; yet, the relationship between these abilities, particularly in trained vocal musicians, has not been the subject of much research. Difference limens for frequency (DLFs) and pitch production accuracy (PPA) were examined among 20 vocalists, 21 instrumentalists, and 21 nonmusicians. All were right-handed young adult females with normal hearing. Stimuli were harmonic tone complexes simulating piano tones and represented the mid-frequency of the untrained female vocal range, $F0=261.63–392$ Hz (C4–G4). DLFs were obtained by an adaptive psychophysical paradigm. Vocal pitch recordings were analyzed to determine PPA. Musicians demonstrated superior pitch discrimination and production accuracy compared to nonmusicians. These abilities did not distinguish instrumentalists and vocalists. DLF and PPA were significantly correlated with each other only for musicians with instrumental training; however, PPA was most consistent with minimal variance for vocalists. It would appear that a relationship between DLF and PPA develops with musical training, and these abilities can be differentially influenced by the type of specialty training. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3021309]

PACS number(s): 43.66.Hg, 43.75.Zz, 43.70.Mn, 43.75.St [DD] Pages: 328–338

I. INTRODUCTION

Auditory pitch discrimination and vocal pitch control reflect abilities necessary for accurate integration of sensory perception, motor planning, and vocal production skills. These processes are regarded by researchers and music educators as fundamental abilities for musical talent and essential skills for vocal musicians (Bentley, 1966; Geringer, 1983; Seashore, 1919; Titze, 1994; Watts *et al.*, 2003). Intuitively, it seems that pitch perception and production should be directly related. Auditory pitch discrimination and vocal pitch control have been proposed to be independent abilities between which a relationship develops with training or maturation (Geringer, 1983; Goetze *et al.*, 1990; Yarbrough *et al.*, 1991). Vocal pitch production is unequivocally affected by acoustic cues (Higgins *et al.*, 2005; Seifert *et al.*, 2002; Ward and Burns, 1978) and may be manipulated by alterations in auditory feedback (Elman, 1981; Hain *et al.*, 2000; Jones and Munhall, 2000; Leydon *et al.*, 2003). A convergence of psychoacoustic evidence acknowledges a relationship between these skills; however, there is an inadequate understanding of the nature of this relationship particularly in formally trained vocal musicians (Amir *et al.*, 2003; Kishon-Rabin *et al.*, 2001; Watts *et al.*, 2003).

While pitch extraction is basic to the perception of speech intonation, rapid and precise pitch perception is critical to the processing of music. The typical contours of pitch change to convey relevant information in the English language are larger than one-half an octave; whereas, melodies of Western culture are generally written with smaller pitch intervals of approximately 1/6th to 1/12th of an octave (Ayotte *et al.*, 2002; Sundberg, 1994). This perhaps explains why Seashore (1919) thought that the auditory perception of minute pitch differences "is a fundamental capacity in musical talent, and upon it rests most of the powers of appreciation and expression in music" [p. 42, as cited in Pedersen and Pedersen, (1970)].

Pitch is the perceptual correlate of frequency and perception of pitch corresponds to the fundamental frequency ($F0$), or the frequency of the lowest partial in the frequency spectrum of a periodic sound (Gelfand, 1998; Sundberg, 1994). The just noticeable difference between two pitches is referred to as difference limen for frequency (DLF) (Gelfand, 1998). The consensus among published research is that formally trained instrumental musicians have more precise frequency discrimination (smaller DLFs) than nonmusicians (Kishon-Rabin *et al.*, 2001; Spiegel and Watson, 1984). Spiegel and Watson (1984) described a "relative acuteness of the ears" and a "mystique" associated with the listening abilities of performers, conductors, and composers of classical music (p. 1690). Kishon-Rabin *et al.* (2001) reported that instrumental musicians compared to nonmusicians had approximately 50% smaller DLF values for pure tones and suggested that extensive music training influences auditory pitch discrimination.

In a subsequent study, Amir *et al.* (2003) compared the relationship between auditory pitch discrimination and vocal pitch matching of pure-tone stimuli among instrumental musicians and nonmusicians. None had previous vocal training. A positive correlation between vocal production and auditory perception was reported ($r=0.67$, $p<0.001$); however, this statistic was based on the combination of all individual data. The relationship between these skills for instrumental musicians versus nonmusicians was not reported. Both auditory

[a]Electronic mail: dnikjeh@aol.com

discrimination and vocal pitch matching were found to be more accurate for instrumentalists than nonmusicians suggesting that instrumental music training may enhance perception of acoustic parameters otherwise ignored by nonmusicians.

Psychoacoustic research suggests that auditory skills may differ between musicians of distinct musical genres (Kishon-Rabin *et al.*, 2001; Spiegel and Watson, 1984). Spiegel and Watson (1984) suggested that musicians who tune their own instrument (e.g., wind instruments) have better frequency discrimination than those who do not (e.g., piano). Kishon-Rabin *et al.* (2001) did not replicate this particular finding; however, they reported that classically trained instrumental musicians (i.e., those who received musical training on the viola, violin, French horn, and bassoon) have significantly better frequency discrimination than those who have a contemporary musical background (i.e., those who received music training on the keyboard, guitar, or percussion instrument). Electrophysiologic measures of auditory-neural differences among musically trained and untrained individuals illustrate that different types of auditory experiences distinctly influence auditory perception. Data from cortical auditory evoked potentials (CAEPs) suggest that conductors have enhanced attention to spatially located sounds compared to other musicians (Nager *et al.*, 2003). Accordingly, instrumental musicians appear to selectively encode acoustic parameters (e.g., pitch, timbre, and interval) most relevant to their musical genre (e.g., jazz and rock) (Tervaniemi *et al.*, 2006), instrumental training (Pantev *et al.*, 2003), or practice strategies (Seppänen *et al.*, 2007).

Although the term "musician" includes instrumentalists and vocalists, psychoacoustic (Kishon-Rabin *et al.*, 2001; Spiegel and Watson, 1984) and neurophysiological research (Koelsch *et al.*, 2002; Koelsch, *et al.*, 1999; Münte *et al.*, 2003; Shahin *et al.*, 2003) has focused predominantly on instrumental musicians (e.g., violinists and keyboard players) rather than vocal musicians. Comparative research on the effects of long-term music training for vocal musicians is lacking. Recently Nikjeh *et al.* (2008) reported that vocal and instrumental musicians have slightly different CAEPs (i.e., mismatch negativity, P3a) to small unattended pitch deviances. Further, auditory-neural responses to small changes of pitch occurred earlier for those vocal musicians with a combination of voice and instrumental training than musicians with instrumental only or vocal only training, suggesting that a combination of music training may provide auditory-neural advantages.

Compared to instrumental musicians, vocalists are unique since their musical "instrument" is endogenous to the body (i.e., the larynx and vocal tract). Integration of auditory and kinesthetic feedback from the voice begins with vocalizations as neonates long before any actual music training. The vocal production of music requires the integration of multiple neural systems including the sensorimotor, auditory, limbic, and executive (Jones and Munhall, 2000; Münte *et al.*, 2003; Mürbe *et al.*, 2004; Titze, 1994). Based on the early work of Kirchner and Wyke (1965), recent researchers have argued that the receptor system in the larynx is continuously influenced by the auditory system and that auditory feedback is crucial for controlling vocal fundamental frequency (*F*0) (Leydon *et al.*, 2003; Perry *et al.*, 1999; Titze, 1994). Previous studies have argued in favor of an interactive closed-loop auditory-governance system (Davidson, 1959; Elman, 1981; Lane and Tranel, 1971; Perry *et al.*, 1999); that is, an auditory perceptual monitoring system that is sensitive to errors and uses feedback to make adjustments whenever error-performance signals are detected. According to Titze (1994), "… pitch is one of those dimensions that if correctly adjusted draws little attention to itself, but if incorrect it can reduce the acceptability and intelligibility of the human voice" (p. 214). For vocal musicians, accurate control of pitch and intonation patterns is of paramount importance.

For speech production, auditory feedback may be the primary control for variation in pitch that constitutes intonation patterns. For singers, auditory feedback has the same function, but by its very nature, real-time auditory feedback does not provide a reliable means for controlling the precision of vocal characteristics for singing. Moreover, auditory feedback alone cannot explain a singer's ability to accurately control pitch when they have difficulty hearing their own voices, such as in choir singing or with loud orchestral accompaniment (DiCarlo, 1994; Mürbe *et al.*, 2002).

A number of music educators and researchers report that vocal instruction and training facilitate kinesthetic feedback or "internal models" of pitch that assist in controlling fundamental frequency and maintaining targeted pitches (Jones and Munhall, 2000; Leydon *et al.*, 2003; Mürbe *et al.*, 2004; Sapir *et al.*, 1983). In other words, the professional singer associates an auditory image of pitch with an internal sensation (DiCarlo, 1994). Similarly, Jones and Munhall (2000) proposed that auditory feedback is necessary for the acquisition and maintenance of these internal models so that singers develop a neural representation or an internal model with the intention that the perceived pitch is matched to a reference pitch in the brain. Furthermore, Leydon *et al.* (2003) described a pitch-shift reflex within the auditory system that is triggered in response to discrepancies between the intended and perceived pitches and propose that this compensatory reflexive response contributes to the production of vocal vibrato in professional singers.

Traditionally, psychoacoustic studies of pitch discrimination have used sinusoidal tones (pure tones) consisting of only a single frequency partial to measure DLFs (Amir *et al.*, 2003; Kishon-Rabin *et al.*, 2001; Leydon *et al.*, 2003; Ward and Burns, 1978; Watts *et al.*, 2003). CAEPs derived from electroencephalography (EEG) consistently demonstrate greater auditory sensitivity for harmonic complexes compared to pure tones (Novitski *et al.*, 2004; Pantev *et al.*, 1998; Shahin *et al.*, 2003). All naturally occurring periodic sounds have a sound spectrum consisting of a time-varying pattern of multiple harmonic partials across a wide frequency range (Novitski *et al.*, 2004). Musical sounds, in particular, are temporally, spectrally, and structurally complex. Thus, it may be argued that in order to establish the most precise measure of DLFs in musicians, selected stimuli should be musically relevant harmonic tone complexes. Further, so that correlations of auditory pitch perception and vocal pitch production may be more accurately assessed, measurements of

task performance should use the same stimuli for both perception and production tasks and represent a gender-appropriate vocal range. By contrast, Amir *et al.*, (2003) chose pure-tone frequencies from the midmale vocal range (131, 165, and 196 Hz) for the vocal task, and correlated this task to an auditory discrimination task for pure-tone frequencies of 250 Hz, 1 kHz, and 1.5 kHz. Measurements of pure tones and unrelated frequencies for pitch production and perception may not accurately reflect the relationship between auditory pitch perception and vocal pitch production for musicians. The present study proposes that the chosen frequencies should be equivalent for both tasks and should also be representative of a musically relevant frequency range.

Auditory pitch discrimination (perception) and vocal pitch accuracy (production) have been identified as related processes that demonstrate the integration of sensory perception with motor planning, and as essential skills for vocal musicians. Although many researchers and music educators acknowledge this relationship between auditory pitch discrimination and vocal pitch control, the nature and development of this relationship remain elusive. Furthermore, scientific research on musicians with vocal training is sparse. Findings for instrumental musicians are often generalized to include *all* musicians. While vocal musicians adhere to the same rigorous training as other musicians, little is specifically known about this particular class of musicians. This investigation sought to examine the pitch perception-production relationship by psychoacoustic analysis of auditory discrimination and vocal production among formally trained vocal and instrumental musicians using musically relevant and gender-appropriate stimuli. Based on previous outcomes, it was predicted that measures of auditory pitch discrimination and vocal pitch production would be more accurate for musicians relative to nonmusicians. Further, it was hypothesized that since vocalists receive auditory and kinesthetic feedback from their instrument as neonates, prior to any actual music training, that the same measures of pitch perception and vocal production would yield superior results for vocal musicians compared to instrumental musicians.

II. METHODS

A. Participants

Sixty-one female students (21 instrumental musicians, 20 vocal musicians, and 20 nonmusicians) were recruited from the University of South Florida (USF). All musicians were formally trained. For the purpose of this investigation, formal music training refers to a minimum of five years of professionally directed and implemented music instruction and technical exercises provided by a professional musician and/or music educator. Table I provides details of musicians' musical training. Instrumental and vocal musicians had a mean age of 22 years (range of 18–33) and an average of 10.5 years of formal music training. On average, musicians received approximately nine years of instruction in their respective area (instrumentalist=9.8 yr; vocalists=9.0 yr). In addition to formal vocal training, almost half of the singers (9 of 19) also received training on a musical instrument (string or wind instrument). The instrumental musicians were all trained on either wind or string instruments. None had training on percussion instruments, nor had any received any formal vocal training. By comparison, mean age for nonmusicians was 23 years (range of 20–34). All had less than 12 months of music training. All participants were right handed, native speakers of English, and had pure-tone thresholds less than 25 dB HL bilaterally at octave frequencies from 250 to 4 kHz. To control for factors that may influence auditory pitch perception, there was no history of neurological impairment, absolute pitch ability, exposure to tone languages, or previous participation in psychoacoustic experiments. Documented informed consent was obtained from all subjects in accordance with the ethical guidelines established by the USF Institutional Review Board.

B. Stimuli

Auditory stimuli consisted of harmonic tone complexes that approximated the physical characteristics of piano tones. Since the assumption underlying this investigation was that accuracy of vocal pitch production is related to auditory frequency perception, the stimuli were the same for each task and chosen from the midfrequency range of the untrained female vocal pitch range. The stimuli extended from music tones C4 to G4 ($F0$=261.63 Hz to $F0$=392 Hz) (Hirano, 1981). In physical terms, the interval between adjacent whole tones is a 12% difference between the fundamental frequencies of each tone. The interval between a reference tone and its semitones is a fundamental frequency difference of roughly 6%, while the interval between a reference tone and its quarter tones is a 3% difference between fundamentals. Stimuli for the pitch production task included C4 ($F0$ =261.63 Hz), E4 ($F0$=329.63 Hz), and G4 ($F0$=392 Hz). In addition, there were three semitones including C4# ($F0$ =277.32 Hz, 6% increase from C4), D4# ($F0$=310.97 Hz, 6% decrease from E4), and F4# ($F0$=369.81 Hz, 6% decrease from G4), and three additional quarter tones (3% increase from C4, $F0$=269.48 Hz; 3% decrease from E4, $F0$ =320.03 Hz; and 3% decrease from G4, $F0$=380.58 Hz) for a total of nine tones. Stimuli for the auditory discrimination task centered around C4 ($F0$=261.63 Hz), E4 ($F0$ =329.63 Hz), and G4 ($F0$=392 Hz), and the stimuli were generated dynamically by algorithm during the perception task as described in the procedures. Each stimulus contained a fundamental ($F0$) and three harmonics. The amplitude of each harmonic was divided by its harmonic number to create a natural amplitude contour in the frequency spectrum. Stimuli were digitally generated (sampling rate=50 kHz) and presented using Tucker-Davis Technologies (TDT) RP2 Real-Time Processor and locally written software. Stimuli were routed bilaterally through a TDT HB7 headphone buffer and Etymotic Research (ER2) insert earphones at an overall level of 75 dB sound pressure level.

C. Procedures

1. Pitch production accuracy (PPA)

To assess vocal pitch matching accuracy, participants were instructed to listen to each of the nine tones until it ended and then to reproduce the tone on the vowel /a/ at the

TABLE I. Instrumental musician (IN) and vocal musician (V) musical history.

Musician	Total years music training	Age onset voice	Total years voice	Age onset instrument	Total years instrument	Instrument category
IN01	10	N/A	N/A	11	10	Wind
IN02	8	N/A	N/A	10	8	Wind
IN03	15	N/A	N/A	3	15	String
IN04	8	N/A	N/A	7	8	String
IN05	15	N/A	N/A	6	15	Wind
IN06	8	N/A	N/A	11	8	String
IN07	11	N/A	N/A	11	11	Wind
IN08	9	N/A	N/A	10	9	Wind
IN09	8	N/A	N/A	7	8	String
IN10	13	N/A	N/A	5	13	String
IN11	8	N/A	N/A	4	8	Srting
IN12	7	N/A	N/A	15	7	Wind
IN13	7	N/A	N/A	11	7	Wind
IN14	9	N/A	N/A	11	9	String
IN15	7	N/A	N/A	11	7	Wind
IN16	15	N/A	N/A	10	15	Srting
IN17	9	N/A	N/A	11	9	Wind
IN18	10	N/A	N/A	11	10	String
IN19	8	N/A	N/A	4	8	String
IN20	13	N/A	N/A	5	13	String
IN21	8	N/A	N/A	10	8	Wind
Mean	9.8 yr			8.8 yr	9.8 yr	
V01	13	5	13	N/A	N/A	
V02	13	17	7	11	9	String
V03	17	9	9	6	9	String
V04	13	13	5	5	9	String
V05	6	8	6	N/A	N/A	
V06	8	11	8	N/A	N/A	
V07	14	8	14	N/A	N/A	
V08	9	11	9	11	5	Wind
V09	11	8	6	8	6	String
V10	10	9	10	N/A	N/A	
V11	11	12	6	7	4	String
V12	10	9	7	N/A	N/A	
V13	8	8	8	6	5	Wind
V14	8	16	5	12	6	String
V15	10	14	7	N/A	N/A	
V16	11	12	11	N/A	N/A	
V17	19	8	19	8	13	String
V18	13	12	9	N/A	N/A	
V19	13	5	13	N/A	N/A	
V20	8	10	8	N/A	N/A	
Mean	11.3 yr	10.3 yr	9.0 yr	8.2 yr	7.3 yr	

same pitch as accurately as possible for 3 s. In addition to insert earphones, subjects wore a head-set microphone positioned at a constant microphone-to-mouth distance of 2.5 centimeters and placed off center at the right corner of the mouth. Each stimulus tone was 1 s in duration and randomly presented three times, totaling 27 stimuli. Interstimulus interval (ISI) was manually controlled by the examiner allowing time for the subject to respond prior to the next presentation. To verify comprehension of the task, each subject had two practice opportunities using a synthesized piano tone other than the test stimuli (e.g., B3, $F0=246.94$ Hz). The productions were directly recorded into a computer using a sampling rate of 22 050 Hz and 16 bits/sample. An autocorrelation analysis of $F0$ for the middle 50% of each production was performed using PRAAT software (Version 4.4). The $F0$ of the three vocal trials were averaged to determine a mean fundamental production of each targeted tone for each participant.

2. Difference limen for frequency (DLF)

To establish DLFs, an adaptive three-interval, three-alternative forced-choice (3I/3AFC) paradigm was used in conjunction with a three-down, one-up stepping rule to estimate the frequency difference corresponding to 79.4% correct point on the psychometric function for each participant and condition (Levitt, 1971; Moore and Peters, 1992). For each single incorrect response, the frequency difference be-

tween standard and target stimuli increased, and three correct responses were required before the frequency difference decreased between standard and target stimuli. A change in the direction of frequency difference between standard and target stimuli constituted a reversal. The 79% 3I/3AFC paradigm has been shown to be more efficient with less threshold bias than the two-interval, two-alternative forced-choice (2I/2AFC) method (Leek, 2001). DLF may be underestimated using a 2I/2AFC procedure due to chance performance (i.e., 50% for 2I/2AFC compared to 33.3% for 3I/3AFC). DLFs were determined for three randomly ordered harmonic complexes: $F0=261.63$ Hz (C4), $F0=329.63$ Hz (E4), and $F0=392$ Hz (G4). Participants were seated in front of a computer monitor in a sound treated booth, listened to the three harmonic complexes represented visually on the monitor with gray boxes, and selected the box (via computer mouse) representing the harmonic complex that was different from the other two. Each harmonic complex was 200 ms in duration (10 ms rise and fall times); ISI was 300 ms.

The paradigm began with a 6% fundamental frequency difference between the target and comparison stimuli. This frequency difference approximates the interval between a reference piano tone and its semitone. For the first three reversals, the frequency difference changed by 3% (i.e., a quarter tone difference). Thereafter, the step-size was.375%. Following ten reversals, the frequency difference of the final six reversals was averaged and reported as the DL for that run. A minimum of three runs of each condition was completed. When necessary, additional runs were completed until three consistent threshold estimates were obtained. Runs were considered consistent when DLs were within a factor of 2. Final DLF for each condition was determined by averaging the three most consistent runs.

III. RESULTS

A. Pitch production accuracy

Measures of pitch production are reported in terms of *relative* accuracy in percent (relPPA%) and were calculated as the absolute difference between the produced F0 and the targeted frequency relative to the targeted frequency in percent ($\text{relPPA}\% = \Delta f/f \times 100$); that is, the relPPA% value decreases as the difference between the produced $F0$ and the targeted $F0$ decreases. Group data for vocal pitch production were also converted to semitones to compare to the Western music scale. PPA data are summarized in Table II.

On average, musicians vocally replicated the targeted pitch to within one-eighth of the simulated piano tone $F0$ (relPPA=1.28%, 0.22 semitone). By comparison, nonmusicians deviated from the reference pitch by more than one semitone (relPPA=7.83%, 1.3 semitones). Vocalists replicated pitch more accurately than instrumental musicians (relPPA=1.05% and 1.50%, respectively); however, differences between the musician groups were not significant. Vocalists appeared to have minimal variability in their production accuracy (SD=2.6 Hz) compared to instrumental musicians (SD=7 Hz).

To evaluate for the effect of group and stimulus condition, a two-way repeated-measure analysis of variance (ANOVA) was conducted on individual PPA measurements with Tone (nine tones) as the within-subjects factor and Group (nonmusicians, instrumental musicians, and vocal musicians) as the between-subjects factor. *Post hoc* analyses were conducted using the least significant difference (LSD) test. Alpha level was set at 0.05.

Overall, PPA differed significantly by Group, $F(2,59)=13.67$, $p \leqslant 0.0001$. Tone effect was not significant, $F(8,472)=1.23$, $p=0.28$, nor was there an interaction between Group and Tone, $F(16,472)=0.83$, $p=0.65$. *Post hoc* analyses indicated that both groups of musicians vocally matched a reference tone with much greater pitch precision than nonmusicians ($p \leqslant 0.0001$). For this particular vocal task, PPA did not significantly differ between vocal and instrumental musicians ($p=0.76$); however, the standard deviation was significantly smaller for vocal musicians than instrumental musicians, indicating more consistent vocal performance on a t-test over standard deviations, $t(16)=3.38$, $p=0.004$.

B. Auditory pitch discrimination

Similar to measurements of PPA, the minimal detectable changes in frequency (Δf) were transformed to *relative* DLF thresholds in percent ($\text{relDLF}\% = \Delta f/f \times 100$). The minimal detectable change in frequency was also converted to music semitones. DLF data are summarized in Table III. Musicians discriminated between two harmonic tones with only one-eighth of a difference between tone fundamental frequencies (relDLF=1.35%, 0.23 semitone). By contrast, the just noticeable difference between two harmonic tones for nonmusicians was greater than one-quarter of a musical tone (relDLF=3.19%, 0.53 semitone). The vocalists and instrumentalists perceived all harmonic complexes equally well with minimal response variability; however, nonmusicians showed a smaller DLF for G4 than for either C4 or E4 and much larger response deviation for all tones.

To evaluate for the effect of group and stimulus condition, a two-way repeated-measure ANOVA was conducted on individual DLF measurements with Tone (C4=261.63 Hz, E4=329.63 Hz, and G4=392 Hz) as the within-subjects factor and Group (nonmusicians, instrumental musicians, and vocal musicians) as the between-subject factor. *Post hoc* analyses were conducted using LSD test. Alpha level was set at 0.05.

Overall, the average relDLF% differed significantly by Group, $F(2,59)=5.45$, $p=0.01$ and by Tone $F(2,118)=3.77$, $p=0.026$. The interaction between Group and Tone did not reach statistical significance $F(4,118)=2.23$, $p=0.07$. *Post hoc* analyses indicated that the nonmusicians had significantly larger DLFs than the two musician groups ($p \leqslant 0.006$); the two musician groups did not differ significantly ($p=0.97$). Also, nonmusicians had significantly smaller DLFs for G4 than for the other two music tones ($p < 0.03$).

TABLE II. Pitch production group data for nonmusicians, instrumental musicians, and vocal musicians for the nine vocally produced tones.

Group	Tone F0	Diff (Hz)	SD	relPPA%	Semitone
Nonmusician	261.63/C4	19.20	20.97	7.34%	1.22
	269.48/C4+qtr	21.45	22.68	7.96%	1.33
	277.32/C4#	16.33	22.58	5.89%	0.98
	329.63/E4	25.48	29.80	7.73%	1.29
	320.03/E4−qtr	27.44	32.16	8.57%	1.43
	310.97/D4#	25.89	25.79	8.33%	1.39
	392.00/G4	32.56	41.88	8.31%	1.38
	380.58/G4−qtr	36.78	37.92	9.66%	1.61
	369.81/F4#	24.64	40.53	6.66%	1.11
Means		25.53	31.33	7.83%	1.31
Instrumental	261.63/C4	4.44	5.02	1.70%	0.28
	269.48/C4+qtr	3.47	4.71	1.29%	0.22
	277.32/C4#	3.63	5.74	1.31%	0.21
	329.63/E4	7.97	14.69	2.42%	0.40
	320.03/E4−qtr	3.11	2.97	0.97%	0.16
	310.97/D4#	4.34	5.58	1.40%	0.23
	392.00/G4	5.08	6.25	1.30%	0.22
	380.58/G4−qtr	6.64	5.36	1.75%	0.29
	369.81/F4#	4.93	5.40	1.33%	0.22
Means		4.85	6.95	1.50%	0.25
Vocal	261.63/C4	3.68	4.33	1.41%	0.24
	269.48/C4+qtr	3.28	2.13	1.22%	0.20
	277.32/C4#	2.30	1.76	0.83%	0.14
	329.63/E4	2.95	1.74	0.90%	0.15
	320.03/E4−qtr	2.92	2.03	0.91%	0.15
	310.97/D4#	2.21	1.32	0.71%	0.12
	392.00/G4	3.29	2.20	0.84%	0.14
	380.58/G4−qtr	6.36	3.05	1.67%	0.28
	369.81/F4#	3.55	2.28	0.96%	0.16
Means		3.39	2.67	1.05%	0.18
Nonmusician means	Combined stimuli	25.53	31.33	7.83%	1.30
Musician means	Combined stimuli	4.12	4.81	1.28%	0.22

C. Relationship between pitch production accuracy and difference limen for frequency

Relationships between pitch production accuracy (relPPA%) and auditory pitch discrimination (relDLF%) were examined using the Pearson correlation coefficient. Since there was no effect of Tone for PPA data or any interaction between Group and Tone for PPA or DLF data, the individual data for PPA and DLF were collapsed across Tone to establish indices of overall PPA and DLF ability for each individual. Individual PPA averages as a function of individual DLF averages are displayed in a scatter plot in Fig. 1. A positive correlation of all individuals was found between the two measures ($r=0.23$, $p=0.035$); however, to report that PPA and DLF are correlated skills would be an inaccurate interpretation of the data. Visual inspection of the scatter plot reveals that the performance of the vocalists is tightly clustered between 1% and 2% for PPA and DLF illustrating minimal response deviation and thus, no correlation between the variables for the vocal musicians ($r=0.18$, $p=0.235$). By comparison, nonmusicians have tremendous performance variability with no clear relational pattern between PPA and DLF ($r=0.003$, $p=0.494$). Only instrumental musicians have a significant positive correlation between PPA and DLF ($r=0.39$, $p=0.04$).

D. Comparison of PPA and DLF

A comparison of PPA and DLF group means are illustrated in Fig. 2. Nonmusicians demonstrated significantly poorer pitch production ability than auditory discrimination, $t(20)=2.46$, $p=0.02$. On average, vocal pitch matching deviated 7.83% (1.3 semitones) from the targeted harmonic tone while auditory discrimination between harmonic complexes averaged 3.19% (0.53 semitone). In contrast, PPA (1.5%) and DLF (1.4%) were comparable skills for instrumental musicians, $t(20)=0.69$, $p=0.50$. Yet, for vocal musicians, PPA (1.05%) was superior to DLF (1.3%), $t(19)=3.17$, $p=0.005$, reflecting an influence of vocal training.

TABLE III. DLF group data for nonmusicians, instrumental musicians, and vocal musicians for each of the three harmonic tones.

Group	Stimulus	DLF (Hz)	SD	relDLF%	Semitone
Nonmusician	261.63/C4	9.05	8.80	3.45%	0.58
	329.63/E4	11.52	13.79	3.49%	0.59
	392.00/G4	10.24	12.29	2.61%	0.44
Means		10.3	11.7	3.19%	0.53
Instrumental	261.63/C4	3.75	1.25	1.43%	0.24
	329.63/E4	4.88	2.21	1.48%	0.25
	392.00/G4	5.02	0.96	1.28%	0.21
Means		4.5	1.6	1.40%	0.23
Vocal	261.63/C4	3.57	1.03	1.35%	0.23
	329.63/E4	4.53	1.84	1.25%	0.21
	392.00/G4	5.39	1.80	1.30%	0.22
Means		4.2	1.1	1.30%	0.22
Nonmusician means	Combined stimuli	10.3	11.7	3.19%	0.53
Musician means	Combined stimuli	4.4	1.4	1.35%	0.23

E. Influence of instrumental training

1. Vocalists

Since almost half (9 of 19) of the vocal musicians had also received instrumental training, it was questioned whether those musicians who received more than one type of training would perform differently than those who did not have this additional musical experience. To evaluate for an influence of instrumental training, the vocal musician group was split and repeated-measures ANOVAs were completed for each behavioral measure comparing nonmusicians, instrumental musicians, vocal only musicians (VO), and vocal musicians with instrumental training (VI). relDLF% continued to differ significantly by Group, $F(3,58)=3.62$, $p=0.02$. *Post hoc* analysis revealed no significant difference between those singers with instrumental training and those without ($p=0.78$). Each musician group had significantly smaller DLFs than nonmusicians (instrumentalists, $p=0.01$; VO, $p=0.03$; VI, $p=0.02$). Differences between instrumental musicians and either group of vocal musicians were not significant (VO, $p=0.91$; VI, $p=0.81$). relPPA% also continued to differ significantly by Group, $F(3,57)=3.28$, $p=0.03$. *Post hoc* measures indicated that PPA was significantly more accurate for each musician group than nonmusicians (instrumentalists, $p=0.01$; VO, $p=0.03$; VI, $p=0.03$). There was no significant difference between VI and VO vocalist groups ($p=0.87$), nor were differences significant between instrumentalists and either group of vocalists (VO, $p=0.98$; VI, $p=0.83$).

An influence of instrumental training on the relationship between DLF and PPA was examined using the Pearson correlation coefficient. Measures of DLF and PPA were significantly correlated for those VI ($r=-0.867$, $p=0.001$) but not for those without instrumental instruction ($r=0.43$, $p=0.105$).

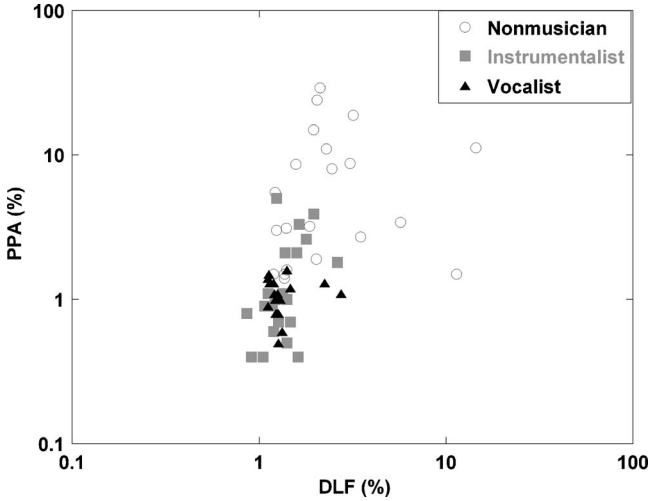


FIG. 1. Scatter plot of PPA and DLF illustrating individual pitch production data (relPPA%) as a function of individual discrimination data (relDLF%) plotted on a logarithmic scale.

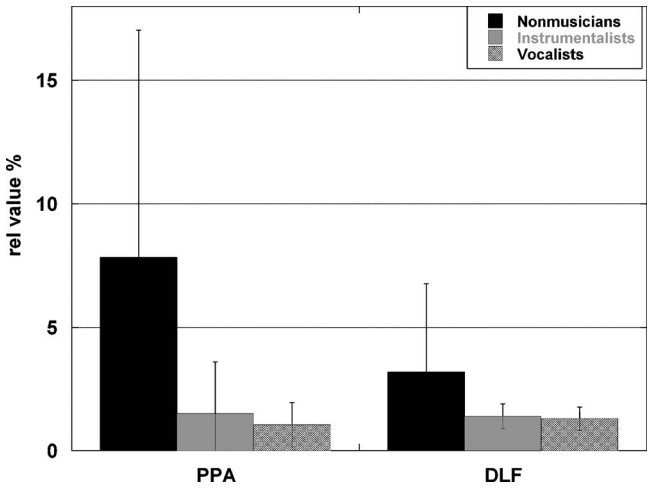


FIG. 2. PPA and DLF means across tones for nonmusicians, instrumental musicians, and vocal musicians. Error bars indicate one standard deviation from the mean.

2. String musicians versus wind musicians

Because previous studies have questioned whether auditory perception differences exist between musicians who play different types of instruments, the instrumental musician group was divided into those who were trained on wind instruments ($n=10$) and those trained on string instruments ($n=11$). Repeated measures ANOVAs were completed for each behavioral measure comparing nonmusicians, instrumental musicians trained on string instruments (IS), instrumental musicians trained on wind instruments (IW), and vocal musicians. relDLF% continued to differ significantly by Group, $F(3,58)=3.59$, $p=0.02$. *Post hoc* analysis revealed no significant differences between the two groups of instrumental musicians ($p=0.99$). Each musician group had significantly smaller DLFs than nonmusicians (IS, $p=0.02$; IW, $p=0.03$; vocalists, $p=0.006$). Differences between vocalists and either group of instrumental musicians were not significant (IS, $p=0.96$; IW, $p=0.98$). relPPA% differed significantly by Group, $F(3,57)=3.30$, $p=0.03$. *Post hoc* measures indicated that PPA was more accurate for all musicians than nonmusicians (vocalists, $p=0.008$; IS, $p=0.04$; IW, $p=0.03$). There was no significant difference between string and wind musicians ($p=0.81$) nor between vocalists and either group of instrumental musicians (IS, $p=0.80$; IW, $p=0.97$).

The Pearson correlation coefficient was used to examine whether the type of instrument on which a musician was trained would influence the relationship between auditory pitch discrimination and PPA. Measures of DLF and PPA were significantly correlated for wind musicians ($r=0.75$, $p=0.006$) but not for string musicians ($r=0.29$, $p=0.19$).

F. Years of music training and psychoacoustic variables

On average, the musicians had 10.5 years of music training ranging from 6 to 19 years. Evidence was lacking to support any significant correlation between the number of years of music training and a musician's frequency discrimination (DLF, $r=0.11$, $p=0.25$) or pitch production accuracy (PPA, $r=-0.07$, $p=0.33$).

G. Age training initiated and psychoacoustic variables

The average age at which these musicians began training was 8.5 years (range=3–15). There was no evidence to support significant correlations between the age that music training was initiated and a musician's frequency discrimination (DLF, $r=-0.23$, $p=0.08$) or pitch production accuracy (PPA, $r=-0.21$, $p=0.09$).

IV. DISCUSSION

A relationship between auditory pitch discrimination and PPA is readily acknowledged; however, the nature of the relationship is unclear. Instrumentalists have been shown to have superior auditory discrimination (Kishon-Rabin *et al.*, 2001) and vocal pitch matching abilities (Amir *et al.*, 2003); yet, there is a paucity of investigative research that examines and describes these behavioral processes in adult vocal musicians. Recent neurophysiologic findings show differences between instrumental and vocal musicians in cortical auditory evoked responses (Nikjeh *et al.*, 2008). Since a vocal musician's instrument is endogenous to the body and given that a vocalist's professional performance relies on precise pitch discrimination and laryngeal neuromuscular control, it was anticipated that measures of pitch perception and production would be more accurate for vocal musicians than instrumental musicians.

As a group, musicians vocally replicated pitches of musical tones with six times the accuracy of nonmusicians (relPPA=1.28% compared to 7.83%). Surprisingly, vocalists (relPPA=1.05%) and instrumentalists (1.5%) did not significantly differ on this task. Both groups vocally replicated the music tone's $F0$ within or below a 0.25 semitone. One plausible explanation for this lack of distinction between the two musician groups is the simplicity of the vocal task. The single pitch imitation task is representative of those used in previous studies and was designed to also investigate differences between musicians and nonmusicians. While this task was simplistic for musicians, it was actually challenging for some of the nonmusicians. It is likely that a more complex pitch production task (e.g., sequence of tones, variable pitch range, or faster stimulus presentation) would delineate differences between the vocal and instrumental musicians, but may be impossible for nonmusicians to perform. Nevertheless, it is interesting to note that instrumentalists, with no previous vocal training, were still five times more accurate than nonmusicians in pitch production. These results replicate and extend the findings of Amir *et al.* (2003) who used single pure tones for vocal pitch production. They found that instrumental musicians with no previous training in voice had both superior pitch discrimination and vocal pitch production abilities compared to nonmusicians. Collectively, these results suggest that music training, whether for an instrument or voice, nurtures a musician's sensitivity to acoustic parameters, such as pitch, rhythm, and timbre, as well as the proprioceptive memories of pitch production using the voice. However, the exact nature and time course of the development of this relationship remains unclear. In the present study, there were no significant correlations between psychoacoustic measures and years of music training or the age that training was initiated. That said, music training as a whole does appear to integrate and strengthen a sensory perception-motor production relationship for all types of musicians. This provides evidence to support an interactive relationship between the neural mechanisms connected to the auditory and laryngeal systems.

Though PPA was similar between vocalists and instrumentalists, vocal musicians had minimal response variability within their group. Nearly every vocalist replicated each pitch within 0%–3% of the targeted music tone, while responses of instrumental musicians varied between 0% and 6% (with one individual varying up to 18%). The precision and uniformity of the vocalists' performance reflect their vocal training and imply that intense vocal training has a positive effect on laryngeal neuromuscular control such that pitch production is not only precise but consistent. Findings support the initial premise that vocal musicians may be more

sensitive to small acoustic changes and minor laryngeal adjustments due to the endogenous nature of their instrument and the neonatal development and kinesthetic integration of their auditory-vocal feedback loop.

Formally trained musicians demonstrated superior auditory pitch discrimination compared to nonmusicians; DLF for music tones was 50% smaller (i.e., relDLF=1.35% compared to 3.19%). Musicians detected pitch changes within one-quarter semitone difference (0.23) between two fundamental frequencies, while nonmusicians required a fundamental frequency difference greater than one-half of a semitone (0.53) to discriminate between musical tones. These results support neurophysiologic findings that musicians have superior auditory-neural sensory memories for acoustic parameters (Nikjeh *et al.*, 2008) and are compatible with the approximate 2:1 ratio in performance between musicians and nonmusicians found by Kishon-Rabin *et al.* (2001) and Spiegel and Watson (1984). The relDLF values for musicians reported by Spiegel and Watson (1984) and Kishon-Rabin *et al.* (2001) were smaller than those in this study (e.g., approximately 1% for instrumental musicians in both studies). This difference may be explained by choices of stimuli and paradigm procedure. Although Spiegel and Watson (1984) used both pure tones and complex (square wave) sounds, component frequencies were above 440 Hz and presented in a 2I/2AFC paradigm that targeted 71% correct performance. For the 2I/2AFC paradigm, the listener heard two tones and was asked to indicate whether the tones were the same or different. Kishon-Rabin *et al.* (2001) also measured 71% correct performance but used a 3I/3AFC paradigm and stimuli consisted of pure tones at 250 Hz, 1 kHz and 2 kHz. For the 3I/3AFC paradigm, listeners heard three tones and were asked to indicate which tone was different from the other two. Each of the selected conditions in the present study (harmonic complexes, low fundamental frequencies, 79% correct, 3I/3AFC) has been shown to elicit larger DLFs in the normal population compared to pure-tone stimuli at frequencies between 500 and 2 kHz and a 2I/2AFC paradigm (Leek, 2001; Moore, 1997; Moore and Peters, 1992). Consequently, a combination of all three of these conditions is expected to elicit higher DLFs than the previous studies. Consistent with these conclusions, Kishon-Rabin *et al.* (2001) reported that for all subjects the largest relDLF% occurred for 250 Hz. Examination of their data reveals that if the DLFs of the three runs for the 250 Hz condition are averaged together, relDLF% for musicians and nonmusicians are remarkably similar to those in the present study. Namely, average relDLF would be 1.42% for musicians and 2.86% for nonmusicians, compared to 1.35% and 3.19%, respectively, in the present study.

DLF measures did not significantly differ between vocal (relDLF=1.3%) and instrumental musicians (1.4%). All musicians had excellent pitch discrimination with minimal response variability (1–2 Hz) reflecting their extensive music training and similar instruction in aural theory. While DLF differences were not significant between these two classes of musicians, this is the first study to show that formally trained vocal musicians, like instrumental musicians, have superior pitch discrimination abilities.

Auditory processing skills have been shown to vary among musicians subsequent to the type of instrument played, practice strategies, and performance venues. For this reason, vocal and instrumental musician groups were parsed into smaller defined groups (vocalists with instrumental training, vocalists without instrumental training, string instrumentalists, and wind instrumentalists). Neither the type of instrument played nor the addition of instrumental training distinguished group measures of DLF or PPA. The relationship between pitch discrimination and production measures was positively correlated for musicians who played wind instruments and negatively correlated for VI. These particular findings were unexpected and may be accidental since after dividing the musician groups, the number of participants in each was 10 or less and the range of performance was very small. However, our preliminary electrophysiological study of preattentive cortical auditory evoked responses found that VI appear to have an auditory-neural advantage over instrumental or VO (Nikjeh *et al.*, 2008). Future investigations may wish to explore whether training in a combination of music areas (e.g., voice and instrument, more than one type of instrument) facilitates further auditory-neural advantages.

The findings of the present study are consistent with a consensus among music educators and researchers that a degree of auditory pitch discrimination may serve as a prerequisite to PPA (Geringer, 1983; Goetze *et al.*, 1990; Yarbrough *et al.*, 1991). Auditory discrimination was almost three times more accurate than pitch production for those with no music training. By comparison, pitch discrimination and vocal production were equivalent skills for instrumental musicians, reinforcing a theory that instrumental training facilitates the integration of the body's sensory and motor systems. In contrast to the other two groups, pitch production was more accurate than pitch perception for vocal musicians. PPA was within one-sixth of a semitone (1.05%), while DLF was approximately one-quarter of a semitone (1.3%). This implies that vocalists internally discriminate between a reference and a target tone with greater precision than they distinguish between two externally presented tones. This implication supports previous findings (DiCarlo, 1994; Kirchner and Wyke, 1965; Mürbe *et al.*, 2002, 2004) that formally trained vocal musicians develop explicit sensory memory representations and enhanced laryngeal neuromuscular reflexes secondary to extensive music training and perhaps due to the endogenous nature of their instrument. Thus, while PPA and DLF develop together in musicians, the relationship between PPA and DLF abilities is complex. For those with no musical training PPA was poorer than DLF, suggesting that the perception of small pitch differences is inherently easier than production. For instrumental musicians, PPA was equal to DLF, indicating that both perceptual and production skills improved together for instrumental musicians. For vocal musicians who are specifically trained for pitch production precision, PPA actually surpasses DLF.

It follows then that any correlation between pitch production precision and auditory discrimination abilities would differ based on group membership. Similar to the results of Amir *et al.* (2003), a combination of *all* individual data yielded a significant positive correlation between PPA and

DLF ($r=0.23$, $p=0.035$). However, a breakdown of data by group revealed that nonmusicians had considerable response variability and no relational pattern between PPA and DLF. By contrast, vocal musicians had consistent responses and minimal variability for both conditions reflecting an impact of music training in auditory perception and vocal control. Only instrumental musicians and VI had significant correlations between auditory discrimination and PPA, reinforcing the implication that instrumental music training facilitates a sensory perception-motor production relationship. This finding has both translational education and therapeutic implications. Perhaps auditory training techniques used by music educators may be incorporated into learning strategies for populations facing sensory perception deficits and motor production challenges, including those with hearing impairment, dyslexia, and/or neurological deficits.

V. CONCLUSIONS

In summary, formally trained musicians, regardless of specialty training, demonstrated superior auditory pitch discrimination and vocal pitch production for harmonic tone complexes (simulated piano tones) compared to nonmusicians. Overall, behavioral measures of PPA and DLF were similar between instrumental and vocal musicians suggesting that music training nurtures a musician's sensitivity to acoustic parameters, thereby facilitating both auditory perception and vocal production regardless of musician category or type of instrument played. PPA was most consistent for vocalists reflecting the influence of extensive vocal training on laryngeal reflexes and sensory-motor memory representations of pitch. Auditory discrimination and laryngeal pitch control appear to develop on a continuum. Previous researchers suggested that these skills may be independent abilities between which a relationship develops with training. Present findings provide support for the argument that auditory pitch discrimination and PPA may be inherently related systems, rather than independent abilities, between which a relationship strengthens with music training.

Whether these exceptional abilities reflect an inherent talent that guides the decision to seek music training, or these abilities develop from intense training, or a combination of these factors, is yet unknown. Regardless, findings indicate that auditory perception and laryngeal control are facilitated by music training. Perhaps the "link" between these two intricate systems lies within the neural substrates of auditory sensory memories and laryngeal motor reflexes that underlie cognitive behavior. At present, the exact nature of this complex relationship continues to remain elusive, but converging evidence from psychoacoustic and neurophysiologic measures brings it closer to our grasp.

Amir, O., Amir, N., and Kishon-Rabin, L. (**2003**). "The effect of superior auditory skills on vocal accuracy," J. Acoust. Soc. Am. **133**, 1102–1108.

Ayotte, J., Peretz, I., and Hyde, K. (**2002**). "Congenital amusia: A group study of adults afflicted with a music-specific disorder," Brain **125**, 238–251.

Bentley, A. (**1966**). *Musical Abilities in Children* (George G. Harrap and Co. Ltd., London).

DiCarlo, N. (**1994**). "Internal voice sensitivities in opera singers," Folia Phoniatr Logop **46**, 79–85.

Davidson, I. (**1959**). "Sidetone delay, reading rate, articulation, and pitch," J. Speech Hear. Res. **2**, 266–270.

Elman, J. (**1981**). "Effects of frequency-shifted feedback on the pitch of vocal production," J. Acoust. Soc. Am. **70**, 45–50.

Gelfand, S. (**1998**). *Hearing: An Introduction to Psychological and Physiological Acoustics* (Dekker, New York).

Geringer, J. (**1983**). "The relationship of pitch-matching and pitch discrimination abilities of pre-school and fourth grade students," J. Res. Music Educ. **31**, 93–99.

Goetze, M., Cooper, W., and Brown, C. (**1990**). "Recent research in singing in the general music classroom," Bull. Council Res. Music Ed. **104**, 16–37.

Hain, T., Burnett, T., Kiran, S., Larson, C., Singh, S., and Kenney, M. (**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.

Higgins, M., McCleary, E., Ide-Helvie, D., and Carney, A. (**2005**). "Speech and voice physiology of children who are hard of hearing," Ear Hear. **26**, 546–558.

Hirano, M. (**1981**). *Clinical Examination of Voice* (Springer-Verlag, New York).

Jones, J., and Munhall, K. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

Koelsch, S., Schmidt, B., and Kansok, J. (**2002**). "Effects of musical expertise on the early right anterior negativity: An event-related brain potential study," Psychophysiology **39**, 657–663.

Koelsch, S., Schröger, E., and Tervaniemi, M. (**1999**). "Superior preattentive auditory processing in musicians," NeuroReport **10**, 1309–1313.

Kirchner, J., and Wyke, B. (**1965**). "Articular reflex mechanisms in the larynx," Ann. Otol. Rhinol. Laryngol. **74**, 748–768.

Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (**2001**). "Pitch discrimination: Are professional musicians better than non-musicians?," J. Basic Clin. Physiol. Pharmacol. **12**, 125–144.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Leek, M. (**2001**). "Adaptive procedures in psychophysical research," Percept. Psychophys. **63**, 1279–1292.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Leydon, C., Bauer, J., and Larson, C. (**2003**). "The role of auditory feedback in sustaining vocal vibrato," J. Acoust. Soc. Am. **114**, 1575–1581.

Moore, B. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, San Diego), pp. 159–167 and 313.

Moore, B., and Peters, R. (**1992**). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," J. Acoust. Soc. Am. **91**, 2881–2893.

Münte, T., Nager, W., Beiss, R., Schroeder, C., and Altenmüller, E. (**2003**). "Specialization of the specialized: Electrophysiological investigations in professional musicians," Ann. N.Y. Acad. Sci. **999**, 131–139.

Mürbe, D., Pabst, F., Hofman, G., and Sundberg, J. (**2002**). "Significance of auditory and kinesthetic feedback to singers' pitch control," J. Voice **16**, 44–51.

Mürbe, D., Pabst, F., Hofman, G., and Sundberg, J. (**2004**). "Effects of a professional solo singer education on auditory and kinesthetic feedback—A longitudinal study of singers' pitch control," J. Voice **18**, 236–241.

Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornells, A., and Münte, T. (**2003**). "The fate of sounds in conductor's brains: An ERP study," Brain Res. Cognit. Brain Res. **17**, 83–93.

Nikjeh, D. A., Lister, J., and Frisch, S. A. (**2008**). "Hearing of note: An electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians," Psychophysiology, **45**, 994–1007.

Novitski, N., Tervaniemi, M., Huotilainen, M., and Näätänen, R. (**2004**). "Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures," Brain Res. Cognit. Brain Res. **20**, 26–36.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L., and Hoke, M. (**1998**). "Increased auditory cortical representation in musicians," Nature (London) **23**, 811–814.

Pantev, C., Ross, B., Fujioka, T., Trainor, L., Schulte, M., and Schulz, M. (**2003**). "Music and learning-induced cortical plasticity," Ann. N.Y. Acad. Sci. **999**, 438–450.

Pedersen, D., and Pedersen, N. (**1970**). "The relationship between pitch recognition and vocal pitch production in sixth grade students," J. Res. Music Educ. **18**, 265–272.
Perry, D., Zatorre, R., Petrides, M., Alivisatos, B., Meyer, E., and Evans, A. (**1999**). "Localization of cerebral activity during simple singing," NeuroReport **10**(18), 3979–3984.
Sapir, S., McClean, M., and Larson, C. (**1983**). "Human laryngeal responses to auditory stimulation," J. Acoust. Soc. Am. **73**, 315–321.
Seashore, C. (**1919**). *The Psychology of Musical Talent* (Silver Burdett Company, New York).
Seifert, E., Oswald, M., Bruns, U., Vischer, M., Kompis, M., and Haeusler, R. (**2002**). "Changes of voice and articulation in children with cochlear implants," Int. J. Pediatr. Otorhinolaryngol. **66**, 115–123.
Seppänen, M., Brattico, E., and Tervaniemi, M. (**2007**). "Practice strategies of musicians modulate neural processing and the learning of sound-patterns," Neurobiol. Learn Mem. **87**, 236–247.
Shahin, A., Bosnyak, D., Trainor, L., and Roberts, L. (**2003**). "Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians," J. Neurosci. **23**, 5545–5552.
Spiegel, M., and Watson, C. (**1984**). "Performance on frequency discrimination tasks by musicians and non-musicians," J. Acoust. Soc. Am. **76**, 1690–1696.
Sundberg, J. (**1994**). "Perceptual aspects of singing," J. Voice **8**, 106–122.
Tervaniemi, M., Castaneda, A., Knoll M., and Uther, M. (**2006**). "Sound processing in amateur musicians and nonmusicians: ERP and behavioral indices," NeuroReport **17**, 1125–1128.
Titze, I. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).
Ward, D., and Burns, E. (**1978**). "Singing without auditory feedback," J. Res. Singing, **1**, 24–44.
Watts, C., Barnes-Burroughs, K., Adrianopoulos, M., and Carr, M. (**2003**). "Potential factors related to untrained singing talent: A survey of singing pedagogues," J. Voice **17**, 298–307.
Watts, C., Murphy, J., and Barnes-Burroughs, K. (**2003**). "Pitch matching accuracy of trained singers, untrained subjects with talented singing voices, and untrained subjects with non-talented singing voices in conditions of varying feedback," J. Voice **17**, 185–194.
Yarbrough, C., Green, G., Benson, W., and Bowers, J. (**1991**). "Inaccurate singers: An exploratory study of variables affecting pitch-matching," Bull. Council Res. Music Ed. **107**, 23–33.

Spectral weights for sample discrimination as a function of overall level[a)]

Lori J. Leibold,[b)] Hongyang Tan, and Walt Jesteadt
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 28 September 2007; revised 23 October 2008; accepted 27 October 2008)

Doherty and Lutfi [(1996). "Spectral weights for overall level discrimination in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **99**, 1053–1058] examined the weights assigned to individual components of a six-tone complex during a sample discrimination task and reported that hearing-impaired subjects gave the most weight to components in the region of their high-frequency hearing loss. In contrast, weighting patterns varied for normal-hearing subjects. In the current study, the same six-tone complex, comprised of the octave frequencies from 0.25 to 8 kHz, was presented to three subjects with normal hearing in high-pass noise, in low-pass noise, and in quiet at two overall levels. Consistent with Doherty and Lutfi, subjects assigned more weight to the 4-kHz component in the high-pass noise condition, but roughly equal weight to all components in the lower-level quiet condition. Weights in the low-pass noise and higher-level quiet conditions, however, were similar to those in the high-pass noise condition. A second experiment compared weights for seven subjects in quiet at four different mean levels. Weights for the highest-frequency components increased as the overall level of the complexes was increased. These results suggest that overall level, rather than degree of hearing loss or sensation level, was the primary cause of the effect that Doherty and Lutfi observed.
 [DOI: 10.1121/1.3033741]

PACS number(s): 43.66.Lj, 43.66.Fe [MW] Pages: 339–346

I. INTRODUCTION

An analysis of trial-by-trial responses can be used to determine the weight subjects assign to individual elements of complex sounds when they are making decisions (e.g., Berg, 1989; Lutfi, 1989, 1995; Richards and Zhu, 1994). In some cases, weights are assessed by computing the correlation between the random variations in the level of each frequency component in a complex tone and the subject's responses. High magnitude relative weights indicate spectral components that contribute more to the subject's decision. Measures of spectral weights are appealing because they can provide information about how individual subjects use information across multiple frequency regions, including conditions in which components are widely spaced.

Doherty and Lutfi (1996) examined the spectral weights that normal-hearing and hearing-impaired subjects assign to the individual frequency components of a multitone complex during a sample discrimination task. For their sample discrimination task, levels of each component of a six-tone complex containing the octave frequencies from 0.25 to 8 kHz were randomly drawn from two different distributions that differed only in the mean level. The subject's decision was to identify the sample that was drawn from the distribution with the higher mean level. Because the difference in the mean level and the standard deviation of the distributions were uniform across the six tones, an optimum decision strategy would weight all tones equally. Considerable individual differences were observed in the weighting strategies of normal-hearing subjects performing the sample discrimination task. However, the average weight varied little across components for normal-hearing subjects. In contrast, hearing-impaired subjects consistently gave the most weight to frequency components in the region of their high-frequency hearing loss. Doherty and Lutfi (1996) suggested that these subjects might have been attempting to compensate for their hearing loss by placing relatively more weight on frequency components that were in the region of their hearing loss.

The finding that hearing-impaired subjects place the greatest weight on frequency components located in the region of their hearing loss during sample discrimination is surprising because those components are presented at a relatively lower sensation level (SL). Mounting evidence suggests that the relative level of individual components plays an important role in determining weights (e.g., Berg, 1990; Lutfi and Jesteadt, 2006). For example, Lutfi and Jesteadt (2006) recently demonstrated that the highest-level components in a multitone sequence are typically assigned the greatest weight, even when the lower-level tones contain more information. There is some indication, however, that patterns of spectral weight are independent of SL. Alexander and Lutfi (2004) measured perceptual weights from normal-hearing and hearing-impaired subjects in an informational masking task and found that, while SL influenced informational masking, weights were not related to SL.

[a)]Portions of these results were presented at the 29th MidWinter Meeting of the Association for Research in Otoloaryngology in Baltimore, MD in February 2006.

[b)]Present address: Department of Allied Health Sciences, University of North Carolina School of Medicine, Chapel Hill, NC 27599. Electronic mail: leibold@med.unc.edu

In the current study, the first experiment examined the relation between level and spectral weight by measuring weights from normal-hearing subjects in quiet, in the presence of high-pass noise, and in the presence of low-pass noise. Subjects were also tested in quiet at the more intense levels used in the noise conditions. The second experiment examined the relation between overall level and weights in greater detail and without noise by comparing the pattern of weights for sample discrimination in quiet at four different mean levels.

II. EXPERIMENT 1: RELATION BETWEEN AUDIBILITY AND SPECTRAL WEIGHTS IN SUBJECTS WITH NORMAL HEARING

The goal of the first experiment was to compare the pattern of weights for sample discrimination of a six-tone complex in normal-hearing subjects across four different conditions. Two quiet conditions replicated the stimuli and conditions used by Doherty and Lutfi (1996) to test normal-hearing and hearing-impaired subjects. Two additional conditions manipulated the audibility of individual components by introducing masking noise to approximate either a sloping high-frequency or a reverse-sloping low-frequency hearing loss.

A. Method

1. Subjects

Three subjects participated in experiment 1. Their ages ranged from 19 to 33 years. All subjects had normal-hearing sensitivity, with thresholds for 200-ms tones in quiet of 20 dB SPL or better at 0.5, 1, 2, and 4 kHz in both ears, when measured using a two-interval, forced-choice (2IFC) adaptive procedure. Subjects were paid for their participation.

2. Stimuli and conditions

The stimuli were tones or multitone complexes presented for 300 ms, including 10-ms $\cos^2$ onset/offset ramps. Following Doherty and Lutfi (1996), the complexes consisted of the six octave frequencies from 0.25 to 8 kHz. On each trial, the levels of the individual frequency components were selected independently and at random from a normal distribution with a standard deviation of 5 dB around one of two mean values. The multitone complex drawn from the distribution with the higher mean level is referred to as the signal. The complex drawn from the distribution with the lower mean level is referred to as the nonsignal.

Stimuli were digitally generated at a 50-kHz sampling rate using an array processor (TDT AP2) and were played through 16-bit digital-to-analog converters (TDT DD1). The stimuli were then low-pass filtered at 20 kHz (TDT FT6) and attenuated (TDT PA4). The output of the attenuator was fed to a headphone buffer (TDT HB6), then to a remote passive attenuator in the sound-treated room, and finally to the left ear of the subject through a Sennheiser HD 250 Linear II headphone.

The four conditions are summarized in Table I. Components for the signal and nonsignal complexes in the quiet 65/62.5 condition were drawn from Gaussian distributions with mean levels of 65 and 62.5 dB SPL, respectively, and standard deviations of 5 dB. The level of the increment used in this condition follows the level used by Doherty and Lutfi (1996) to determine spectral weights in their normal-hearing subjects and was expected to result in performance close to $d'=1.0$ for normal-hearing listeners. In the quiet 80/75 condition, the mean levels of each component for the signal and nonsignal complexes were 80 and 75 dB SPL, respectively. This condition was included to correspond to the levels of components used by Doherty and Lutfi (1996) to test hearing-impaired subjects because it resulted in a performance close to $d'=1.0$ for those subjects. In the high-frequency noise and low-frequency noise conditions, weights were measured in the presence of noise designed to simulate a high-frequency sloping and a low-frequency reverse-sloping hearing loss, respectively. The filtered noises were generated digitally prior to testing at a sampling rate of 50 kHz (MATLAB, Mathworks). The high-frequency noise was high-pass filtered with a 3-dB down cut-off frequency of 2500 Hz (6 dB/octave). The low-frequency noise was low-pass filtered with a 3-dB down cut-off frequency of 750 Hz (8 dB/octave). The mean rms level of each noise was 65 dB SPL. The spectrum level of the high-frequency noise was therefore 23 dB SPL while the spectrum level of the low-frequency noise was 36 dB SPL. Consistent with the quiet 80/75 condition and with the levels used by Doherty and Lutfi (1996), the mean levels of each component for the signal and nonsignal complexes were 80 and 75 dB SPL, respectively. Because the noise was more than 30 dB lower than the lowest level of any of the tonal components, variability in the noise did not contribute to variability in the level of individual tones.

TABLE I. Summary of conditions for experiment 1.

Conditions	Masker	Mean signal/nonsignal level (dB SPL)
Quiet 65/62.5	None	65/62.5
Quiet 80/75	None	80/75
High-freq noise	High-pass noise	80/75
Low-freq noise	Low-pass noise	80/75

3. Threshold procedure

Prior to the measurement of perceptual weights, the audibility of individual components was determined for each of the six components in the complex (octave frequencies from 0.25 to 8 kHz) alone in quiet and in the presence of the high-frequency and low-frequency noise. Thresholds were estimated using a 2IFC adaptive procedure that estimated 71% on the psychometric function (Levitt, 1971). The starting level of the adaptive track was 30 dB SPL for tones presented in quiet and 60 dB SPL for tones presented in either the high-frequency or low-frequency noise. The initial step size was 4 dB, followed by a step size of 2 dB after the fourth reversal. Each trial consisted of a 300-ms warning interval and two 300-ms observation intervals. The interstimulus interval was 500 ms. A 300-ms feedback interval followed the two observation intervals, visually indicating the interval that contained the signal. Thresholds for each

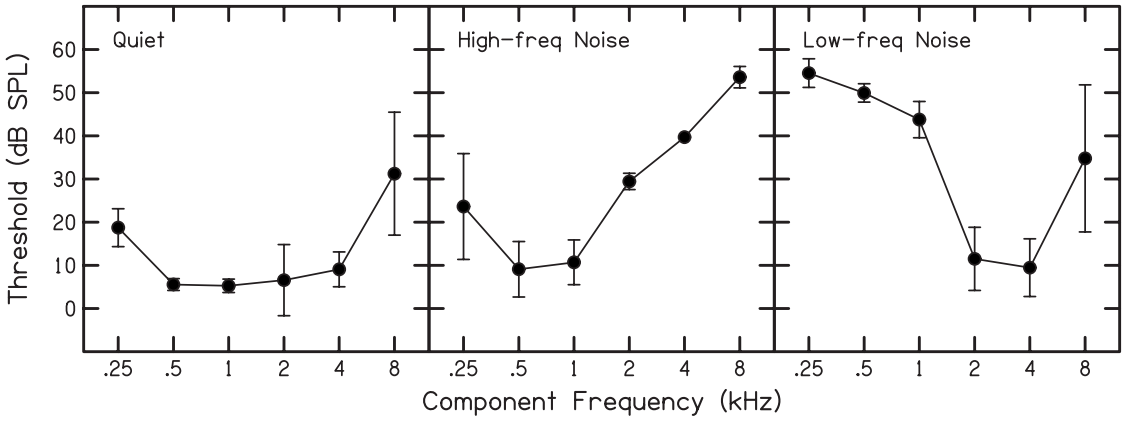


FIG. 1. Mean thresholds for the individual components of the six-tone complex plotted as a function of condition. Error bars are ±1 SD of the mean threshold across subjects.

50-trial block were computed by calculating the average of all reversals points for the 2-dB step size. Two threshold estimates were completed for each of the six components for each of the four conditions.

Subjects were tested as a group. Each subject was tested in a separate section of a double-walled Industrial Acoustics sound-treated room that had single walls between sections. The group was run in 2-h sessions, including regular breaks.

4. Sample discrimination procedure

Weights were measured for each condition using a 2IFC procedure. The signal and nonsignal complexes were presented in random order on each trial. Subjects were instructed to indicate the interval that contained the complex drawn from the more intense distribution. The warning, observation, interstimulus, and feedback intervals were identical to those used for the threshold procedure. Response feedback was provided and indicated the interval that contained the signal complex. Test order across the four conditions was identical for all three subjects: quiet 65/62.5, high-frequency noise, low-frequency noise, and quiet 80/75. Subjects completed ten consecutive blocks of 100 trials for each condition (1000 total trials).

B. Results

Figure 1 presents mean thresholds as a function of component frequency, with data shown in separate panels for each tone presented in quiet, in the presence of the high-pass noise, and in the presence of the low-pass noise. Error bars are ±1 SD of the mean threshold across subjects. As expected, the presence of the high-pass noise increased thresholds for frequency components at and above 2 kHz. The presence of the low-pass noise increased threshold for frequency components at and below 1 kHz. Note that the patterns of thresholds observed in the presence of the high-frequency and low-frequency noise are consistent with thresholds for subjects with sloping high-frequency and reverse-sloping low-frequency hearing loss, respectively.

Following Lutfi (1995), spectral weights were estimated for the sample discrimination task from the following linear regression model:

$$D = \Sigma w_i x_i + C,$$

where D is the subject's response, x_i is the difference in the level of the ith component across the two intervals, w_i is the weight applied to the ith component, and C is a constant. The regression coefficients or "beta" weights were then normalized so that the sum of the absolute weights was equal to 1.[1]

The normalized weights for individual components were averaged across subjects and are shown in Fig. 2, plotted as a function of component frequency. Error bars represent ±1 SD of the mean weight across subjects. Data for each condition are provided in separate panels. Consistent with Doherty and Lutfi (1996), the average weighting functions indicated that each component in the complex was assigned approximately equal weight in the quiet 65/62.5 condition, but subjects assigned more weight to the 4-kHz component in the high-frequency noise condition. Of particular interest to the current study, weights in the quiet 80/75 and low-frequency noise conditions resembled those obtained in the high-frequency noise condition. That is, subjects tended to

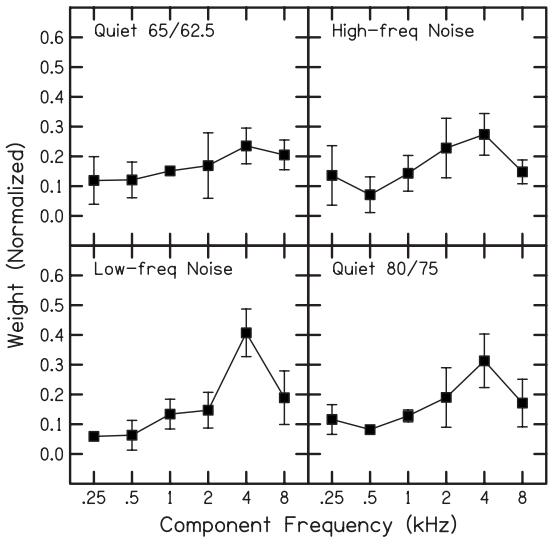


FIG. 2. Mean normalized weights for the individual components of the six-tone complex plotted as a function of condition. Error bars are ±1 SD of the mean weight across subjects.

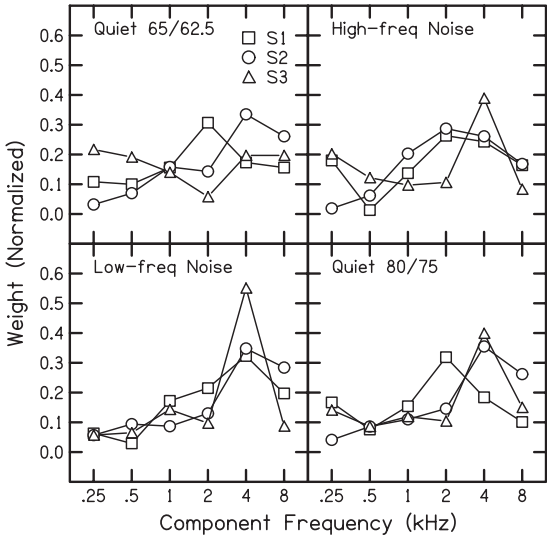


FIG. 3. Weights for the individual components of the six-tone complex are shown for each subject. Weights for each condition are plotted in separate panels.

place the most weight on the high-frequency components when the overall level of the complexes was high, regardless of the SL of individual components.

A repeated-measures analysis of variance (ANOVA) performed on the normalized weights yielded results consistent with the pattern observed in Fig. 2. No significant main effect of condition was observed $[F(3,6)=0.9;p=0.5]$. The main effect of component frequency was significant $[F(5,10)=3.9;p<0.05]$, indicating weights differed across frequency. The component frequency × condition interaction was not significant $[F(15,30)=1.5;p=0.2]$.

Individual differences in weight were evident for each of the four conditions. Figure 3 shows the normalized weights for individual subjects. Data for each condition are provided in separate panels. For the quiet 65/62.5 condition, differences in weighting functions were observed across the three subjects. S1 gave the most weight to the 2-kHz component, S2 gave the most weight to the 4- and 8-kHz components, and S3 gave approximately equal weight across all components, except for the 2-kHz component, which was assigned relatively less weight. Consistent with the mean weights, however, subjects tended to assign the greatest weight to the 4-kHz component in the high-frequency noise, low-frequency noise, and quiet 80/75 conditions.

To examine differences in performance across the four conditions, estimates of the observed sensitivity (d'_{obs}) were computed for each subject from the trial-by-trial data. Average estimates of d'_{obs} for each condition are shown in Table II. For the condition with the lowest overall level and the smallest level increment (quiet 65/62.5), d'_{obs} was 0.74. In contrast, d'_{obs} was consistently higher for three conditions with the higher overall level and larger level increment.

Following Berg (1990), estimates of efficiency were also computed for each condition. Overall performance efficiency (η_{obs}) provides an indication of performance relative to ideal performance and was determined by computing $\eta_{obs}=(d'_{obs}/d'_{ideal})^2$. Weighting efficiency ($\eta_{wgt}$) provides an estimate of the loss in efficiency that results from using nonoptimal weights and was determined by computing $\eta_{wgt}=(d'_{wgt}/d'_{ideal})^2$. Finally, noise efficiency (η_{noise}) provides an estimate of the loss in efficiency from other factors such as internal noise. Noise efficiency was determined by computing $\eta_{noise}=(d'_{obs}/d'_{wgt})^2$. Average estimates of the three measures of weighting efficiency for each condition are shown in Table II. Overall performance efficiency (η_{obs}) is uniform across conditions, making it clear that the low value of d'_{obs} in the quiet 65/62.5 condition was a result of the smaller mean difference. The estimate of weighting efficiency (η_{wgt}) appears higher in the quiet 65/62.5 condition than in the other conditions, as would be expected from the more uniform weights for that condition observed in Figs. 2 and 3. A repeated-measures ANOVA with two within-subjects factors (condition and component frequency) was performed on the estimates of weighting efficiency. This analysis revealed no significant main effect of condition $[F(3,6)=2.9;p=0.13]$, however, even though there was a significant interaction of component frequency and condition in the individual weights.

C. Discussion

The purpose of experiment 1 was to examine the relation between audibility and weight by measuring weights from normal-hearing subjects in quiet, in the presence of high-pass noise, and in the presence of low-pass noise. The six components were not at equal SL even in the quiet conditions, but there was no relation between either the mean threshold (Fig. 1) or individual subject's thresholds and the weight assigned to individual components in the quiet 65/62.5 condition. Consistent with weighting functions for subjects with high-frequency hearing loss (Doherty and Lutfi, 1996), normal-hearing subjects tended to place the most weight on the highest-frequency components when the audibility of high-frequency, high-level components was reduced via high-frequency noise. When the audibility of low-frequency, high-level components was reduced via low-frequency noise or when the complexes were presented in quiet at the higher overall levels used by Doherty and Lutfi (1996) to test

TABLE II. Mean sensitivity (d') and efficiency (η) estimates for experiment 1. For each mean estimate, ±1 SD is also provided.

	d'_{obs}	d'_{wgt}	d'_{ideal}	η_{obs}	η_{wgt}	η_{noise}
Quiet 65/62.5	0.74 ± 0.25	1.14 ± 0.07	1.22	0.39 ± 0.27	0.86 ± 0.10	0.45 ± 0.27
Quiet 80/75	1.45 ± 0.24	2.19 ± 0.09	2.45	0.35 ± 0.11	0.80 ± 0.07	0.44 ± 0.13
High-freq noise	1.50 ± 0.40	2.13 ± 0.13	2.45	0.39 ± 0.21	0.76 ± 0.09	0.51 ± 0.30
Low-freq noise	1.26 ± 0.33	2.06 ± 0.23	2.45	0.28 ± 0.15	0.71 ± 0.14	0.41 ± 0.38

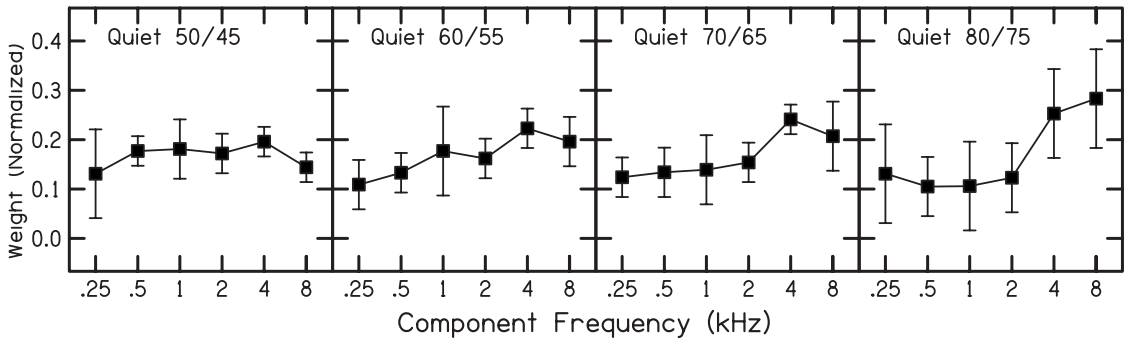


FIG. 4. Mean weights for the individual components of the six-tone complex are shown across subjects. Weights for each condition are plotted in separate panels. Error bars are ±1 SD of the mean weight across subjects.

hearing-impaired subjects, however, subjects continued to assign greatest weight to the highest-frequency components. These results demonstrate a relation between overall level and spectral weighting patterns in sample discrimination, at least in normal-hearing subjects.

Although Doherty and Lutfi (1996) did not obtain weights for normal-hearing subjects at the same higher overall levels used to test their hearing-impaired subjects, Kortekaas *et al.* (2003) reported that normal-hearing subjects assigned the greatest weight to higher-frequency components with increasing level for a level discrimination task. In addition, Lentz (2007) recently demonstrated that weights for spectral-shape discrimination for some subjects were also influenced by the overall level of the stimulus. Both studies will be discussed in greater detail below.

These findings suggest that overall level, rather than SL of the tones or compensation for hearing loss, was the primary cause of the effect that Doherty and Lutfi (1996) observed for hearing-impaired subjects. However, interpretation of these results is complicated by the difference in the level of the average increment for the quiet 65/62.5 condition (2.5 dB) compared to the average increment for the three conditions presented at a higher overall level (5 dB), by the use of noise in some conditions and not others, and by the lack of a significant difference in weighting efficiency (η_{wgt}) across conditions.

III. EXPERIMENT 2: RELATION BETWEEN OVERALL LEVEL AND SPECTRAL WEIGHTS

In experiment 1, subjects tended to place the most weight on the highest-frequency components when the complexes were presented in quiet at the higher overall levels used by Doherty and Lutfi (1996) to test hearing-impaired subjects. This pattern of weights was observed in quiet or when audibility of components was reduced via high-frequency or low-frequency noise. These results indicate a relation between overall level and spectral weighting patterns in sample discrimination when the samples differ in intensity. However, the level of increment between the signal and nonsignal complexes was smaller for less intense quiet 65/62.5 condition compared to the other three conditions. The difference in signal strength, along with the corresponding difference in sensitivity (d'_{obs}), complicates the interpretation of differences in weighting patterns across conditions. The goal of the second experiment was to determine the effect of overall level on perceptual weights in more detail by comparing the pattern of weights for sample discrimination of the six-tone complex in quiet at four different mean levels. The same level increment was used across the four conditions.

A. Method

1. Subjects

Seven additional subjects participated in all conditions. All subjects (19–33 years) had normal-hearing sensitivity, with thresholds for 200 ms tones in quiet of 20 dB SPL or better at 0.5, 1, 2, and 4 kHz in both ears, when measured using a 2IFC adaptive procedure. Subjects were paid for their participation. Subjects were tested in three groups, with the groups run as described in experiment 1.

2. Stimuli, Procedure, and Conditions

All aspects of the stimuli and procedure used in experiment 2 were the same as in experiment 1. Subjects were tested in each of the four conditions. The quiet 80/75 condition was identical to the quiet 80/75 condition described in experiment 1. The mean levels of each component for the signal and nonsignal complexes were 80 and 75 dB SPL, respectively. In the quiet 70/65, quiet 60/55, and quiet 50/45 conditions, the mean levels of each component for the signal and nonsignal complexes were 70 and 65, 60 and 55, and 50 and 45 dB, respectively. Subjects completed ten blocks of 100 trials for each condition. Four subjects were tested in the following order: quiet 50/45, quiet 60/55, quiet 70/65, and quiet 80/75. The remaining three subjects were tested in the reverse order. Subjects completed ten consecutive blocks of 100 trials, resulting in 1000 trials for each condition, prior to testing for the next condition.

B. Results

As described in experiment 1, regression coefficients were normalized so that the sum of the absolute weights was equal to 1. Mean normalized weights for individual components across the seven subjects are shown in Fig. 4, plotted as a function of component frequency. Error bars represent ±1 SD of the mean weight across subjects. Data for each condition are plotted in separate panels. For the quiet 50/45 condition, mean weights varied little across components. The

TABLE III. Mean sensitivity (d') and efficiency (η) estimates for experiment 2. For each mean estimate, ± 1 SD is also provided.

	d'_{obs}	d'_{wgt}	d'_{ideal}	η_{obs}	η_{wgt}	η_{noise}
Quiet 50/45	1.70 ± 0.21	2.30 ± 0.08	2.45	0.49 ± 0.11	0.88 ± 0.06	0.55 ± 0.11
Quiet 60/55	1.46 ± 0.16	2.32 ± 0.14	2.45	0.36 ± 0.08	0.90 ± 0.10	0.40 ± 0.08
Quiet 70/65	1.75 ± 0.22	2.28 ± 0.12	2.45	0.52 ± 0.12	0.87 ± 0.09	0.61 ± 0.19
Quiet 80/75	1.66 ± 0.28	2.05 ± 0.26	2.45	0.48 ± 0.14	0.71 ± 0.17	0.66 ± 0.13

pattern of weights changed, however, as the mean level of the complexes was increased. For the quiet 60/55, quiet 70/65, and quiet 80/75 conditions, more weight was assigned to the two highest components (4 and 8 kHz).

A repeated-measures ANOVA with two within-subjects factors (component frequency and mean level) was performed on the normalized weights. The analysis was consistent with the trends observed in Fig. 4. No significant main effect of component frequency $[F(5,30)=2.2;p=0.10]$ or mean level $[F(3,18)=0.7;p=0.5]$ was observed. However, the component frequency $\times$ mean level interaction was significant $[F(15,90)=8.4;p<0.001]$. The interaction suggests a difference in the pattern of weights with changes in the mean level.

As shown in Table III, systematic changes in d'_{obs} were not observed across conditions. However, a one-way repeated-measures ANOVA found that subjects' estimates of weighting efficiency (η_{wgt}) were significantly different as the mean level changed $[F(3,18)=4.6;p=0.01)$, consistent with the component frequency $\times$ mean level interaction in the weights. No systematic differences in noise efficiency (η_{noise}) were observed with changes in the mean level.

C. Discussion

The purpose of experiment 2 was to examine the relation between overall level and spectral weight without differences in the overall difficulty of the task or possible effects of noise. The pattern of weights changed with the overall level, with more weight assigned to the highest two components (4 and 8 kHz) as the mean level of the complexes was increased. Consistent with experiment 1, there does not appear to be a relation between weights and SL *per se*. Considered together, the results of experiments 1 and 2 suggest that the overall level, not the SL or degree of hearing loss, may have been the primary cause of the effect that Doherty and Lutfi (1996) observed.

It is not clear why more weight was given to high-frequency components with increasing level. Perceptual weights reported by Kortekaas *et al.* (2003) are consistent with the idea that subjects assign greater weight to higher-frequency components of a complex as the overall level of the complex is increased. Kortekaas *et al.* (2003) varied the range of their stimuli by varying the number of components from 3 to 24 while holding the midpoint constant at 1.6 kHz and component spacing constant at 1 Bark per component. They found that subjects continued to assign more weight to the highest-frequency components at higher levels, regardless of the absolute frequencies involved. This result is clearest in their Fig. 3, which shows perceptual weights for seven-component intensity discrimination obtained with a procedure similar to that used in the current study, but with a narrower frequency spacing. In addition, Kortekaas *et al.* (2003) examined whether the changing pattern of weights reflected increased spread of excitation with increasing level by introducing flanking bands of noise expected to reduce spread of excitation. The pattern of weights did not change with the flanking noise.

Lentz and Leek (2003) compared spectral-shape discrimination in hearing-impaired and normal-hearing subjects using six tones spanning the range from 0.2 to 3 kHz with equal spacing in log frequency. For standard stimuli, the tones were uniform in level, but both equal-SPL and equal-SL conditions were included in the study. The comparison or signal stimuli consisted of the same six tones, but with an average level difference between the lowest three and the highest three. In the conditions of most interest to the current study, Lentz and Leek (2003) varied each tone around its mean level with a standard deviation of 1.5 dB and a total range of ± 4.5 dB. Subjects with normal hearing tended to put more than optimum weight on the middle two components, while those with hearing impairment tended to assign more weight to the highest and lowest tones. This result was most apparent when the spectral shape of the six-tone complex was shifted to produce a higher level for the three lowest-frequency tones relative to the three highest-frequency tones. Overall performance and weighting efficiency were comparable for the two groups of subjects. In a subsequent study, Lentz (2007) evaluated the effect of level in a similar spectral-shape discrimination task in subjects with normal hearing, using six tones spanning the range from 0.7 to 4 kHz. In the standard stimulus, the average level of the six tones was 35, 50, or 85 dB SPL. The subjects placed greater weight on the higher-frequency components and this effect increased with overall stimulus level while weighting efficiency decreased with level.

Both Kortekaas *et al.* (2003) and Lentz (2007) considered whether peripheral interactions at the level of the basilar membrane could account for the increased weight assigned to higher-frequency components at higher levels. Kortekaas *et al.* (2003) noted that the lowest and highest components would be expected to receive more weight based on spread of excitation, but rejected that account because only the highest-frequency components received more weight in their study. In addition, the use of flanking noise bands to restrict spread of excitation had no effect in most cases. Thus, Kortekaas *et al.* (2003) attributed the increased weight assigned to higher frequencies to an edge effect due to lack of suppression of the highest-frequency components. In contrast,

Lentz (2007) concluded that the increased weight assigned to higher frequencies at higher levels could be accounted for by upward spread of excitation. Using output from a level-dependent filter bank model (Glasberg and Moore, 1990), Lentz (2007) estimated the level of excitation at the output of each of 22 filters for standard and signal stimuli on a trial-by-trial basis. She then computed the correlation between the difference in excitation at each filter output and each subject's responses to determine a weighting function based on the excitation pattern rather than on the physical level of the stimuli. An excitation pattern at each overall level (i.e., 35, 50, and 85 dB SPL per component) averaged across five individual signal levels was also computed. Lentz (2007) found that weights based on the correlation between subjects' responses and the excitation patterns were in good agreement with weights based on correlations between subjects' responses and the physical stimuli. This analysis demonstrated that the change in the excitation patterns between standard and signal stimuli at filter frequencies above the highest of the six components increased with stimulus level, while the change in excitation at frequencies below the lowest of the six components remained constant or decreased with level.

As part of an earlier study related to the current experimental conditions (Leibold *et al.*, 2007), two approaches were used to test the hypothesis that spectral weights could be determined by the contributions of individual components of a five-tone complex to overall loudness as predicted by the loudness model proposed by Moore *et al.* (1997). The analyses used by Leibold *et al.* (2007) were similar in some respects to those used by Lentz (2007) in that loudness estimates generated by the model were based on excitation patterns generated by a level-dependent filter bank. In the first analysis performed by Leibold *et al.* (2007), the predicted loudness of the two stimuli presented on every trial was computed and the model "voted" for the louder stimulus. The correlation of those votes with the level of each of the five tones was computed to determine the perceptual weights predicted by the loudness model. Note that the approach taken by Leibold *et al.* (2007) differs from the one taken by Lentz (2007) who computed the correlation between the filter bank output and subjects' responses. In a second analysis, Leibold *et al.* (2007) varied the level of each tone individually, holding the other tones constant at the mean level. The relative change in estimated loudness resulting from varying each tone was used as a measure of the spectral weight that should be assigned to that tone by the loudness model. The two approaches used by Leibold *et al.* (2007) provided comparable results, but only the correlational analysis was reported because it yielded a direct estimate of spectral weights and provided a more rigorous test of the loudness model.

Because Leibold *et al.* (2007) were concerned with the relation between component spacing and loudness, component spacing was varied in both analyses, but overall level was held constant at 60 dB SPL (53 dB per component). As shown in Fig. 2 of Leibold *et al.*, 2007 the correlational analysis predicts a greater contribution from tones at the edge of the pattern for all but the widest bandwidth. The frequency ratio between adjacent components in that condition was 1.59. For the next widest bandwidth, the ratio was 1.26. The loudness model could not predict results for the widest bandwidth where there was little or no predicted peripheral interaction between adjacent components. In the patterns used by Kortekaas *et al.* (2003), the spacing was at intervals of 1 Bark, but the approximate ratio between adjacent components was 1.16, well within the range where a peripheral effect would be predicted. For the patterns used by Lentz (2007), the ratio between adjacent components was 1.42, near the upper limit of the range where peripheral interaction is observed in the loudness model. In the current study, the ratio was 2.0.

In the current study, the relative contribution of each component to the overall loudness of the six-tone complex was estimated in experiment 2 for levels ranging from 50 to 80 dB SPL using the second analysis performed by Leibold *et al.* (2007) and described above. The results of this analysis suggested relatively flat functions that did not vary with level. A similar analysis of the stimuli used by Lentz (2007) showed a greater contribution to overall loudness by the highest-frequency component and an increase in this effect at higher levels. This effect was not observed, however, if the component spacing was increased to the octave spacing used by Doherty and Lutfi (1996) and the current study. The current results suggest that the increased spectral weight assigned to the highest-frequency components cannot be accounted for using a peripheral model, when the component spacing is broad. It is assumed that the increase reflects a central process. It would be parsimonious to assume that the same central effect is responsible for the effects observed by Lentz (2007) and by Kortekaas *et al.* (2003), but peripheral interactions cannot be ruled out for the more closely spaced frequencies used in those studies.

A similar increase in the perceptual weight assigned to higher-frequency information with increasing level has been reported in at least one measure of speech perception. Calandruccio and Doherty (2007) recently examined spectral weighting strategies for normal-hearing and hearing-impaired subjects performing a sentence recognition task. Consistent with the results of Doherty and Lutfi (1996) for nonspeech stimuli, hearing-impaired subjects tended to place the most weight on higher-frequency speech information. Of particular interest to the current study, normal-hearing subjects also tended to weight the higher-frequency sentence information the most, but only when sentences were presented at the same high overall presentation level used for hearing-impaired subjects.

IV. SUMMARY AND CONCLUSIONS

Normal-hearing subjects tended to place the most weight on the highest-frequency components of a multitone complex during a sample discrimination task when the audibility of the high-frequency components was reduced via high-frequency noise. This result is consistent with the pattern of weights observed for hearing-impaired subjects by Doherty and Lutfi (1996).

Subjects also assigned more weight to the highest-frequency components of a multitone complex when the au-

dibility of low-frequency components was reduced via low-frequency noise and when the complexes were presented in quiet at the higher overall levels used by Doherty and Lutfi (1996) for their hearing-impaired subjects.

For sample discrimination of the complexes in quiet, no relation between SL and perceptual weight was observed. However, weights for the highest-frequency components increased as the overall level of the complexes was increased.

Existing models of peripheral excitation cannot account for the increased weight assigned to the highest-frequency components at high levels for the widely spaced components used in the present study and by Doherty and Lutfi (1996). The effect is therefore assumed to reflect central auditory processes.

ACKNOWLEDGMENTS

Research supported by NIH R01 DC006648 and T32 DC00013. Subject recruitment supported by NIH P30 DC04662.

[1]Negative weights would not be expected in this task and, indeed, the only negative weight observed prior to normalization had a value of −0.001.

Alexander, J. M., and Lutfi, R. A. (**2004**). "Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights," J. Acoust. Soc. Am. **116**, 2234–2247.

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

Berg, B. G. (**1990**). "Observer efficiency and weights in a multiple observation task," J. Acoust. Soc. Am. **88**, 149–158.

Calandruccio, L., and Doherty, K. A. (**2007**). "Effect of overall level on spectral weighting strategies for speech," Bulletin Am. Aud. Soc. **32**, 40.

Doherty, K. A., and Lutfi, R. A. (**1996**). "Spectral weights for overall level discrimination in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **99**, 1053–1058.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Kortekaas, R., Buus, S., and Florentine, M. (**2003**). "Perceptual weights in auditory level discrimination," J. Acoust. Soc. Am. **113**, 3306–3322.

Leibold, L. J., Tan, H., Khaddam, S., and Jesteadt, W. (**2007**). "Contributions of individual components to the overall loudness of a multitone complex," J. Acoust. Soc. Am. **121**, 2822–2831.

Lentz, J. L. (**2007**). "Variation in spectral-shape discrimination weighting functions at different stimulus levels and signal strengths," J. Acoust. Soc. Am. **122**, 1702–1712.

Lentz, J. L., and Leek, M. R. (**2003**). "Spectral shape discrimination by hearing-impaired and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 1604–1616.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lutfi, R. A. (**1989**). "Informational processing of complex sound. Intensity discrimination," J. Acoust. Soc. Am. **86**, 934–944.

Lutfi, R. A. (**1995**). "Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks," J. Acoust. Soc. Am. **97**, 1333–1334.

Lutfi, R. A., and Jesteadt, W. (**2006**). "Molecular analysis of the effect of relative tone level on multitone pattern discrimination," J. Acoust. Soc. Am. **120**, 3853–3860.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "Model for the prediction of thresholds, loudness, and partial loudness," J. Audio Eng. Soc. **45**, 224–237.

Richards, V. M., and Zhu, S. (**1994**). "Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients," J. Acoust. Soc. Am. **95**, 423–434.

Acoustic, psychophysical, and neuroimaging measurements of the effectiveness of active cancellation during auditory functional magnetic resonance imaging

Deborah A. Hall[a] and John Chambers[b]
MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom

Michael A. Akeroyd[c]
MRC Institute of Hearing Research (Scottish Section), Glasgow Royal Infirmary, 16 Alexandra Parade, Glasgow G31 2ER, United Kingdom

John R. Foster[d]
MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom

Ron Coxon[e]
Sir Peter Mansfield Magnetic Resonance Centre, School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

Alan R. Palmer[f]
MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom

(Received 10 April 2008; revised 17 October 2008; accepted 20 October 2008)

Functional magnetic resonance imaging (fMRI) is one of the principal neuroimaging techniques for studying human audition, but it generates an intense background sound which hinders listening performance and confounds measures of the auditory response. This paper reports the perceptual effects of an active noise control (ANC) system that operates in the electromagnetically hostile and physically compact neuroimaging environment to provide significant noise reduction, without interfering with image quality. Cancellation was first evaluated at 600 Hz, corresponding to the dominant peak in the power spectrum of the background sound and at which cancellation is maximally effective. Microphone measurements at the ear demonstrated 35 dB of acoustic attenuation [from 93 to 58 dB sound pressure level (SPL)], while masked detection thresholds improved by 20 dB (from 74 to 54 dB SPL). Considerable perceptual benefits were also obtained across other frequencies, including those corresponding to dips in the spectrum of the background sound. Cancellation also improved the statistical detection of sound-related cortical activation, especially for sounds presented at low intensities. These results confirm that ANC offers substantial benefits for fMRI research. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3021437]

PACS number(s): 43.66.Lj, 43.66.Vt, 43.50.Ki [BCM] Pages: 347–359

I. INTRODUCTION

In recent years functional magnetic resonance imaging (fMRI) has become a popular technique for noninvasive measurements of central auditory function in humans (e.g., Hall *et al.*, 2003; Hall, 2006; Johnsrude *et al.*, 2002). Its use, however, is complicated by the problem that the particular magnetic resonance (MR) pulse sequence—echo-planar imaging (EPI)—used to acquire the images of functional brain activity also generates an intense acoustic noise[1]. This noise results from the switching of currents in the gradient coil of the MR scanner. Because this occurs in a large static magnetic field, it effectively turns the MR scanner into a giant loudspeaker. The noise level typically exceeds 100 dB SPL at the position of a listener (Foster *et al.*, 2000; More *et al.*, 2006; Ravicz *et al.*, 2000). To protect listeners against the damaging noise exposure during scanning, passive noise reduction is routinely achieved by ensuring that listeners wear earplugs and/or circumaural ear defenders (Ravicz and Melcher, 2001). When these methods are used together, a considerable amount of attenuation can be achieved. For example, the ear defenders used in the present set of experiments provide 25 dB of attenuation at 600 Hz rising to 40 dB at 4 kHz. However, passive attenuation alone is still insufficient for acceptably quiet conditions, and even with it, the distracting, loud, repetitive noise is known to elevate feelings of anxiety, emotional distress (Quirk *et al.*, 1989) and claustrophobia (Granet and Gelber, 1990). It has also been shown to interfere with perceptual judgements (e.g., Boyle *et al.*, 2006). Of particular concern for fMRI studies of sound processing is the masking produced by the background noise of stimuli presented to the listener through headphones and its interference with the measures of

[a]Author to whom correspondence should be addressed. Electronic mail: d.hall@ihr.mrc.ac.uk
[b]Electronic mail: john.chambers@intelligent-energy.com
[c]Electronic mail: maa@ihr.gla.ac.uk
[d]Electronic mail: john@ihr.mrc.ac.uk
[e]Electronic mail: ron.coxon@nottingham.ac.uk
[f]Electronic mail: alan@ihr.mrc.ac.uk

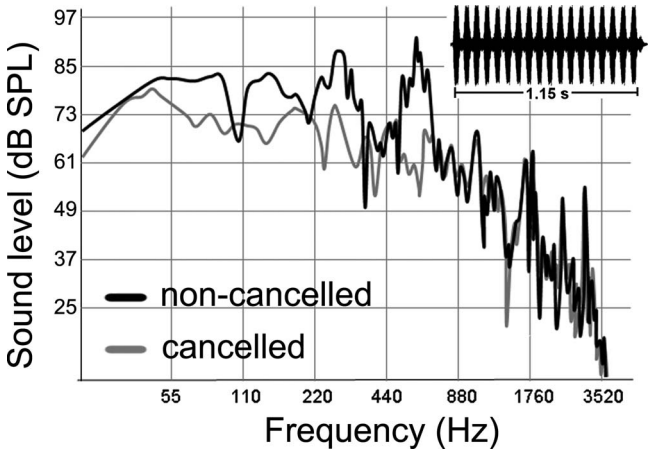


FIG. 1. The average power spectrum of the scanner noise generated during functional brain imaging (EPI) using a Philips Intera 3 T MR scanner. The black line (noncancelled) indicates the spectrum of the noise recorded under normal operation and shows a peak of power at 600 Hz and lesser peaks at 150 and 300 Hz. The gray line (cancelled) indicates the spectrum when the ANC is operative. The power around the dominant peak is reduced by 35 dB. The inset (upper right) shows the time-domain waveform of a single burst of EPI noise, a sequence of 18 discrete pings corresponding to a series of 18 consecutive slices through the brain. Each ping lasts about 0.06 s, and the total duration of the burst is about 1.15 s.

stimulus-related brain activation (Elliott *et al.*, 1999; Shah *et al.*, 1999). Moreover, EPI noise itself produces measurable brain responses not only within the primary auditory cortex (Bandettini *et al.*, 1998), but also within surrounding nonprimary auditory regions (Hall *et al.*, 2000; Ulmer *et al.*, 1998). It is therefore preferable to reduce the noise level by other means.

Three different approaches to noise reduction have been reported. One is to reduce the noise at source, but since the source of the sound—the gradient coil switching—is an essential part of the scanning process, the technical challenges in reducing the induced mechanical vibrations of the scanner are considerable. A second approach has been to develop scanning protocols that minimize the instantaneous level or the total dosage of the background noise, but these either require specialized pulse programming (Hennel *et al.*, 1999; Schwarzbauer *et al.*, 2006) or dramatically increase the scanning time required to obtain the same number of brain images (Hall *et al.*, 1999); for a review, see Amaro *et al.* (2002). A third approach is to acoustically reduce the level of the sound using the technique of active noise control (ANC). This technique is examined here.

The spectral characteristics of the EPI noise are mainly a function of the "readout" gradient pulses of the chosen MR pulse sequence and comprise a series of harmonically related peaks together with a background of broadband noise. The intensity of the peaks is primarily attributable to the features of the gradient-coil assembly and so can differ across manufacturers and models (More *et al.*, 2006). The peak level of the EPI noise produced by the MR scanner that was used in the current series of experiments is 99 dB SPL measured at the center of the bore. The dark line in Fig. 1 illustrates the spectrum of this noise. It has a peak at 600 Hz with two other prominent peaks, 5–10 dB lower in level, at 150 and 300 Hz. The harmonic structure of the noise, and its spectral invariance, make it an ideal candidate for active cancellation. Moreover, the noise repeats at an exact period determined by the EPI pulse sequence. In the example shown in Fig. 1 (see inset), there is a rapid succession of 18 "pings," corresponding to a series of 18 consecutive slices acquired through the brain. The repeating period is 64 ms and the whole noise period lasts approximately 1.15 s. This temporal predictability is also critical to the success of ANC.

One ANC approach is to acoustically control the gradient coils themselves, a method which has been shown to greatly reduce the noise generated at the source within the coil structure (Chapman *et al.*, 2003). However, so far this method amounts only to a technical demonstration and it has not been incorporated into commercial MR systems. The second approach that we have taken has been to design an active acoustically controlled headset. This method uses a cancellation signal to add destructively with the noise at the listener's ear. It has been shown to greatly reduce the noise at the eardrum of the listener and poses no restrictions on the MR scanning hardware (Chambers *et al.*, 2001, 2007). Previously, we have reported the technical specification of a two-channel sound system for the active control of EPI noise that worked in conjunction with passive attenuation (using circumaural ear defenders) (Chambers *et al.*, 2001) and subsequent modifications to optimize the performance of this system (Chambers *et al.*, 2007). This prototype system was originally evaluated in the laboratory using a plastic-and-wood replica of a scanner in which recordings of EPI noise were presented over loudspeakers. For both recorded examples of noise and for simulated scanner sounds with dominant frequencies from 500 Hz to 3 kHz, the ANC system reduced the levels of the peak components by about 40 dB, when measured with a microphone placed underneath the circumaural ear defenders. The perceptual benefits of noise reduction were ascertained by making loudness matches between cancelled and noncancelled targets that were created from recordings of EPI noise that had been filtered to remove energy at the frequencies remote from the fundamental (here the dominant components were 600 and 1900 Hz). In contrast to the acoustic cancellation, perceptual cancellation reached, at most, only 13 dB. We proposed that the difference between acoustic and perceptual measures of noise reduction was likely to be due to the rise of the bone conduction route as the primary path of sound to the cochlea when attenuation of the airborne sound had reached its limit. Accordingly, Chambers *et al.* (2001) predicted that ANC should be most effective when the most intense component of the EPI noise occurs at low frequencies ($<$1000 Hz), where the difference between bone and air conduction is greatest (Berger, 1983). While circumaural headphones are known to produce an occlusion effect that can amplify the bone-conducted physiological noise, such effects appear only to reach significant levels at 250 Hz and below (of the order of 5 dB; Berger and Kiervan, 1983). For the present purposes, where the EPI noise is well above this frequency, the occlusion effect is therefore unlikely to significantly limit cancellation performance.

The main goal of the three experiments reported here was to examine the effectiveness of this ANC system under

true fMRI conditions. We used a Philips 3 T MR scanner for acquiring EPI scans during all of our tests. The noise generated by this MR scanner contains three prominent peaks below 1 kHz (Fig. 1) and is representative of the EPI noise generated by other commercial MR scanners that are currently available. However, body vibrations are much greater in a real MR scanner than in a simulation, through the coupling between the bed and the gradient coil system. The unknown characteristics of the bone-transmitted energy mean that it is uncertain how much perceptual benefit might be provided by ANC. Both experiments 1 and 3 investigated the perceptual benefits of the reduction in the EPI noise by measuring detection thresholds of targets presented at different frequencies, with and without the ANC system operating. Experiment 2 evaluated the effects of ANC on the detectability of stimulus-related activation measured by fMRI using sounds whose frequency was close to the dominant peak at 600 Hz. In all three experiments the EPI noise itself served as the masker.

An important component of the ANC system is the adaptive filter which optimizes the cancellation signal. The coefficients of this filter are determined during an initial "training" period, but they can then either be fixed for the remainder of the test or allowed to continually adjust. Previously, we used a fixed-coefficient ANC system (Chambers *et al.*, 2001, 2007), but if either the spectral content of the scanner noise or the position of the listener's head changes sufficiently over time, then the effectiveness of this is diminished. Given that it takes several minutes to arrive at the detection threshold by adaptive procedures, fixing the cancellation signal may pose a real concern for the practical implementation of the system. Accordingly, experiments 1 and 2 were conducted using an ANC system in which the filter coefficients were continually updated. For comparison, experiment 3 was conducted using an ANC system in which the filter coefficients were fixed.

II. EXPERIMENT 1: EFFECTS OF CANCELLATION AT THE PROMINENT FREQUENCY OF THE EPI NOISE

A. Methods

1. Description of the ANC system

The present ANC system enables the presentation of high-fidelity sound to participants. It uses Sennheiser HE60 electrostatic headphones incorporated into Bilsom 2452 circumaural ear defenders, together with an external high-voltage amplifier (Sennheiser HEV70).

Only a brief summary of the ANC system is given here; detailed technical specification can be found in Chambers *et al.* (2001, 2007). The ANC system operates independently in the left and right channels. The system is similar in concept to a standard feed-forward noise control system (Nelson and Elliott, 1992), which has two microphones, termed the "reference" and "error," and which perform two different functions. The reference microphone receives the sound that is propagating toward the point of cancellation. This sound is termed the reference signal. The error microphone monitors the residual sound after cancellation. Ideally, in a MR scanner, the reference microphone would be near the noise source (i.e., the gradient coils) and the error microphone would be near to the listener's eardrum. However, the two microphones must be sufficiently separated in space to enable a long enough propagation delay between them for generating the cancellation signal. In the MR scanner, this cannot be achieved because the gradient coils are too close to the listener's head. Accordingly, our system has only one microphone that serves both functions, but at different times. This microphone uses optical technology so that the radio-frequency energy that is used in the scanning process does not affect its operation, and its presence does not produce any deterioration of the brain images acquired. It is positioned under the ear defender, close to the entrance of the ear canal. The electrostatic transducer of the headset delivers both the experimental stimuli and the cancellation signal.

The standard function of the reference microphone is replaced in our system by a prerecording of the EPI noise that serves as the reference signal. This recording is taken from the microphone during one of the initial bursts of EPI noise immediately before cancellation is started. It cannot be used directly as the cancellation signal because it is affected by the transfer function of the ANC system, including the electrostatic transducer and the error microphone in the headset. Therefore the transfer function is compensated for by a digital filter (termed the "plant model"), which is calculated before the start of cancellation by presenting a maximum length sequence (MLS) through the system and receiving it on the microphone. This is done during the initialization phase of the MR scanner, when the coolant pump (which recycles liquid gases surrounding the superconducting magnet) automatically switches off and before the gradient switching begins[2].

During ANC operation, the cancellation signal is presented through the transducer, so that it adds destructively with the EPI noise at the listener's ear. The level of residual noise at the microphone (now operating as the error microphone) is continually monitored and used to fine tune the cancellation signal. The timing of the cancellation signal is determined by an electrical trigger pulse, produced by the MR scanner 25 ms in advance of the EPI noise. The cancellation signal itself is derived from the reference signal via an adaptive digital filter, whose coefficients are set so that cancellation is as effective as possible (i.e., the resultant signal at the error microphone is as small as possible). The initial setting is found during four training scans and uses a fast rate of change of adaptation. These coefficients are then subsequently adjusted at a slower rate of adaptation, to track slow changes in the characteristics of the EPI noise. This two-step process improves the immunity of the adaptive filter from the stimuli that are being presented during the fMRI experiment, and hence, helps to maximize the acoustic noise reduction.

In practice, the ANC system is easy to operate, and it is integrated into the same PC software used for presenting the sound stimuli during the fMRI experiment. The three stages involved in setting-up and using the system (determining the plant model, acquiring the reference signal, and optimizing

the coefficients of the adaptive filter) are all automatically controlled within this software and are specified by only a few parameters.

2. Stimuli

The targets were diotic narrowband noises, centered at 600 Hz, with a bandwidth of 50 Hz, a duration of 500 ms and sampled at a rate of 44 100 Hz. Stimuli were first created in the spectral domain (using a fast Fourier transform with 2-Hz resolution) with rectangular edges to the passband, before transforming to the time domain and windowing using 50-ms raised-cosine ramps applied to the onset and offset. One hundred examples of the target were created in advance, and the total root-mean-square (rms) power of each was equated. On each trial, a sound file was chosen randomly from the set as the target. The presentation level of the target was varied according to am adaptive procedure by applying a digital attenuation before digital-to-analog conversion; 0-dB attenuation corresponded to a level of 94 dB SPL. This, and subsequent, levels were calibrated by mounting the headset on a KEMAR manikin equipped with a free-field response microphone (Brüel and Kjær, type 4134) and Zwislocki Coupler (Brüel and Kjær, type DB-100), and connected to a measuring amplifier (Brüel and Kjær, type 2636).

The maskers were the noises generated by the MR scanner (a Philips 3 T Intera equipped with an eight-channel receiver head coil) as it acquired sets of EPI brain images. Each noise burst comprised a rapid succession of 18 pings occurring over 1.15 s (see inset, Fig. 1). Each "ping" corresponded to a brain slice and the set of slices was angled in an oblique horizontal plane sufficient to cover the superior temporal gyrus (auditory cortex) in both hemispheres. Note that in experiments 1 and 3, these brain images were not analyzed because only the psychophysical data were relevant to the question.

Conventional fMRI has equal intervals between each set of image acquisitions, and so the noises bursts are equally spaced. However, this sequence would have caused difficulties for creating the timing of the standard psychophysical method of two closely spaced intervals followed by an intertrial gap. Accordingly, we used a bespoke EPI pulse sequence to generate a noise sequence that would permit us to conduct a two-interval psychophysical paradigm. In this paradigm, a "trial" comprised a pair of masker noise bursts separated by a gap of 0.85 s. An intertrial interval of 2.85 s provided a suitable response window. The response window between each pair of masker noise bursts was created by applying a modified gradient waveform comprising only slice-selective excitation when the radio-frequency excitation pulse was applied (see Schwarzbauer *et al.*, 2006). Because this event did not generate an intense noise nor create a brain image, it is termed a "dummy" scan. Note that this bespoke sequence was only required for the psychophysical evaluation and would not be necessary for the more routine implementation of active noise control in auditory fMRI.

3. Procedures

The detection threshold for the target was measured using two-interval, two-alternative forced-choice (2I-2AFC) task with a two-down, one-up adaptive procedure that converged upon the 70.7% correct point on the psychometric function (Levitt, 1971). The target was presented randomly within one of the noise maskers and its onset always occurred 0.40 s after the onset of the masker. The listener was required to judge whether the target appeared in the first or second interval, responding by pressing one of two buttons with the index finger of their left or right hand. At its initial level, the target was clearly audible (79 dB SPL in the cancelled condition and 89 dB SPL in the noncancelled condition). This level was used for the first two trials of each run. Thereafter, the level changed in steps of 5 dB until two reversals occurred, and by 2 dB until four more reversals occurred. The mean of the levels at the last four reversal points was taken as the detection threshold. During each run, the residual sound at the error microphone inside the earcups was recorded onto a digital audio tape (DAT) recorder (Sony DTC-690) for off-line evaluation of the acoustic cancellation over time. For each listener, ten adaptive runs were completed for both the noncancelled and the cancelled conditions. Each run lasted about $3\frac{1}{2}$ min. To avoid fatigue, these measurements were spread across two $1\frac{1}{2}$ h testing sessions. In each session, five adaptive runs were completed for each condition, the order of conditions being counterbalanced across sessions and listeners in an ABBA manner. Listeners were allowed a short break after each adaptive run.

4. Listeners

Four right-handed listeners participated in experiment 1. Participants were two males aged 42 and 39 years, and two females aged 18 and 36 years (henceforth referred to as subjects 1–4, respectively). Listeners 2 and 4 were authors of this paper. All listeners had normal hearing as assessed using standard pure-tone audiometry [$<$20 dB hearing level (HL) over octave frequencies between 250 and 8000 Hz].

B. Analysis

The residual sound that was recorded from the error microphone was analyzed to determine the sound level at the error microphone over the entire duration of the runs with cancellation, including the training period. Since experiment 1 is concerned only with the amount of cancellation at the frequency where the EPI noise is most intense, the digitized recordings were first filtered with a 100-Hz-wide filter centered on 600 Hz (Adobe Audition fast Fourier transform (FFT) filter) to remove sound energy that was remote from the frequency of interest. We chose a bandwidth of 100 Hz as a close approximation to the equivalent rectangular bandwidth of the psychophysical auditory filter at 600 Hz (Glasberg and Moore, 1990). These filtered recordings were then processed using a customized program to calculate the rms value for each burst of EPI noise. The program detected the portions of the recording in which the residual sound exceeded a certain threshold and it then calculated a rms value over a central 1-s time window that contained either the

TABLE I. Individual sound levels measured in dB SPL at the listeners' ears in experiment 1. Values in parentheses represent the standard error of the measurement. The three different masker levels were computed from the DAT recording of the error microphone using a 100-Hz wide filter centered on 600 Hz. The masker levels reported for the cancelled condition were measured for trials at the reversal points in the adaptive run. The individual threshold levels of the 600-Hz signal are provided for comparison with the masker levels in the better ear.

	Noncancelled	Cancelled			
Listener	Masker (better and worse ear mean)	Masker in better ear	Masker in worse ear	Signal level	Acoustic noise reduction (better and worse ear mean)
1	94 (0.3)	57 (0.3)	62 (0.7)	59 (0.7)	37 (0.5)
2	94 (0.7)	59 (0.3)	62 (0.6)	55 (0.7)	37 (0.4)
3	92 (0.6)	53 (0.3)	59 (0.4)	53 (0.7)	38 (0.4)
4	90 (0.2)	53 (0.3)	57 (0.5)	50 (0.6)	37 (0.3)
Mean	93	56	60	54	37

masker alone or a combination of the masker and the 600-Hz target. The recorded levels were then calibrated against the known level of the target presented on each trial so that we obtained an accurate rms level estimate over time at the error microphone. Note that, due to the 2I-2AFC procedure, for each pair of noise bursts, one value represented the combined sound level of the target and masker, while the other represented only the masker. The rms levels of the masker noise in the cancelled condition are highly informative because the difference between the levels of the masker noise at the start and end of the run defines the amount of acoustic cancellation achieved by the ANC system. The absolute values also provide information about differences across listeners and across ears (within a listener). We also compared the values of the masker to the presentation level of the target on the corresponding trial, for the last four threshold reversal points, to investigate whether the level of the masker was the limiting factor for perception. Masker levels were measured using only those intervals that did not contain the 600-Hz target and are reported separately for the better and worse ears since the amount of acoustic noise reduction was not always equal at the two ears. It is reasonable to assume that listeners performed the detection task using their better ear, irrespective of whether this was the left or the right.

C. Results

1. Acoustic noise reduction during the cancelled condition

Table I shows the noncancelled and cancelled masker levels for each listener. On average these were 93 and 58 dB SPL, respectively. Thus, the ANC system provided 35 dB of acoustic noise reduction around 600 Hz, ranging from 33 and 36 dB across individuals. The corresponding reduction in the broadband rms noise level was, on average, 14 dB.

A typical adaptive run recorded from listener 4 is illustrated in Fig. 2(a). The events in the track include the initial recording scan and the four training scans, followed by 38 stimulus trials. After the initial rapid training phase, the residual sound at the error microphone had reduced from 91 down to 54 dB SPL—a drop of 37 dB. For every adaptive run, the training phase provided at least 30 dB of acoustic cancellation. A substantial benefit was reached over the first 1 min or so of the adaptive run. Allowing the filter coefficients of the ANC system to slowly adapt throughout the run enabled further slow-rate optimization which provided an additional 5 dB or so of acoustic benefit. However, the acoustic reduction was not absolutely stable over time and whether the better ear, acoustically speaking, was the left or the right varied from trial to trial.

Figure 3 shows the three examples of the masker envelopes taken from recordings at the error microphone. The top panel shows one of the noncancelled maskers, illustrating the periodicity of peaks and dips caused by the sequential acquisition of the slices in the scan. The remaining panels show two of the cancelled maskers. The ANC system reduced the level by 30–40 dB, but there was a noteworthy moment-to-moment variation in its performance. The middle panel shows a general improvement across the masker, but the system did not preserve the periodicity evident in the noncancelled masker, whereas the bottom panel shows slightly better performance as well as an overall periodicity.

The amount of cancellation on average differed by 4 dB across ears. There was a significant influence of gender on the final level of the cancelled masker ($F[319,1]=111.1$, $p<0.001$). The level of the residual noise both before and after cancellation was higher for the two male listeners (1 and 2, mean) than for the two female listeners (3 and 4). We speculate that the effect of head size may be important here. When the male listeners were wearing the headset, it was a tight fit to position them inside the headcoil and there was no available space between them and the coil itself. Although the sides of the headset were padded with two thin layers of foam, it is likely that this contact provided an additional route for the conduction of vibrational energy from the MR scanner. In contrast, the two female listeners had smaller heads and so there was no direct coupling between the headset and the headcoil.

2. Perceptual benefits of noise reduction

For each listener, the perceptual benefit of noise reduction was computed by subtracting the mean threshold during the ten adaptive runs in the noncancelled condition from

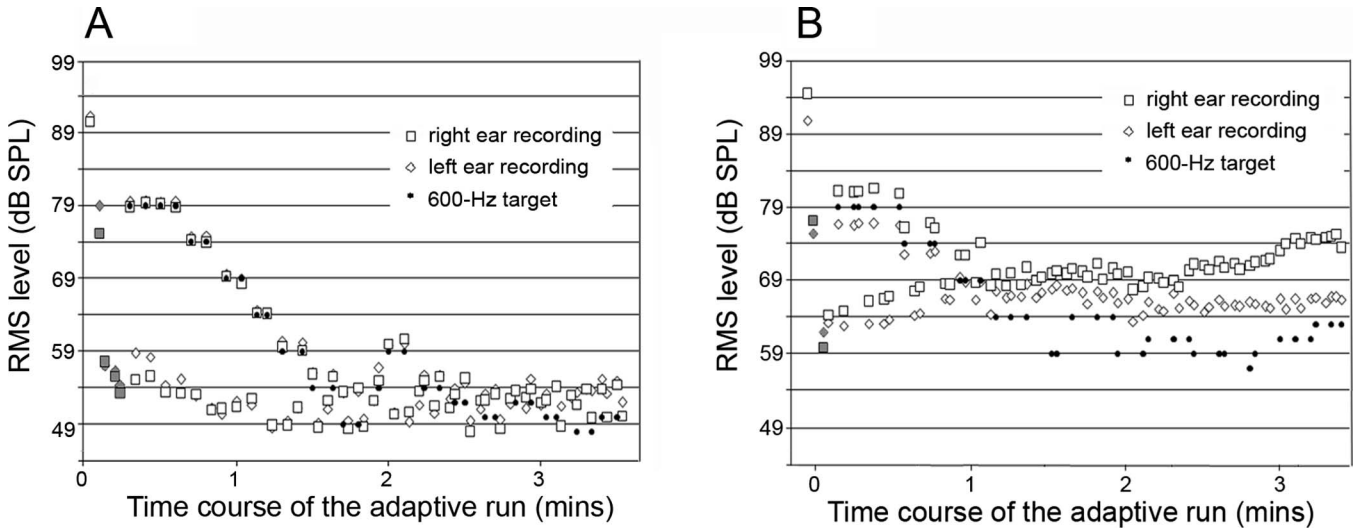


FIG. 2. Examples of the time course of acoustic noise reduction and its relationship to detection thresholds for a 600-Hz target. Diamonds and squares denote the rms level at the left and right ears, respectively. The first scan corresponds to the recording of the noise reference and indicates the noncancelled level of the EPI noise. (A) Example taken from listener 4 in experiment 1. The level of the EPI noise rapidly reduces over the four training scans (gray infills) as the ANC system optimises the adaptive filter. In each pair of subsequent trials, the level in one interval is higher than in the other because it contains both the masker and the target signal. The black dots denote the actual presentation level of the 600-Hz target, the first two practice trials being at 79 dB SPL. The target level tracks the interval that represents the combined level of the noise and target reasonably well. (B) Example taken from listener 6 in experiment 3. The level of the EPI noise rapidly reduces over the two training scans (gray infills), but the amount of noise reduction is generally less than for experiment 1 and the masker attenuation in the right ear can be seen to deteriorate by over 15 dB over the course of the adaptive run.

each of their thresholds measured in the cancelled condition. Across the four listeners, the average detection threshold for the 600-Hz target masked by the noncancelled masker was 74 dB SPL. With cancellation, the average was 54 dB SPL, an improvement of 20 dB. Table II reports the size of the perceptual benefit of noise reduction for individual listeners. There was a significant difference in benefit between listeners ($F[39,1]=5.4$, $p<0.01$), with post-hoc testing revealing that listener 4 gained significantly greater benefit than listeners 1 and 2 ($p<0.05$). We note that this is consistent with the gender difference in the levels of residual acoustic noise reported above.

3. Interim summary

These results clearly demonstrate that the ANC system can provide useful amounts of both acoustic and perceptual noise reduction under true fMRI conditions. Our listeners gained about 35 dB of acoustic reduction and 20 dB of perceptual benefit. The perceptual benefit was smaller than the acoustic reduction. Some of the difference may be due to differences in the "peakiness" of the envelope of the masker across the cancelled and noncancelled conditions (Fig. 3). It is well known that signal-detection thresholds can depend upon the peakiness of the masker envelope (Kohlrausch and Sander, 1995). Also, Gockel *et al.* (2002), who compared detection thresholds for noise signals masked by a complex tone with a 62.5-Hz fundamental frequency with either a repeated, modulated envelope or with a random envelope, found that the detection threshold in the repeated-envelope condition was between 5 and 20 dB lower than in the random-envelope condition (depending on the masker level). One possibility here is that detecting a signal in EPI noise is facilitated by the regular dips between the pings of the noncancelled masker (Fig. 3, top panel) relative to the irregular envelopes of the cancelled masker (Fig. 3, bottom two panels).

In our earlier tests of the ANC system (Chambers *et al.*, 2001), we had proposed that the difference between the acoustic and perceptual measures was due to the emergence of bone conduction as the primary transmission path of the EPI noise to the cochlea. The present results remain consistent with that notion, although they do not decide the issue. The sound may be transmitted indirectly through the head and body via the coupling inherent to the person lying on the MR scanner bed (cf., above-noted greater amount of residual acoustic noise and smaller perceptual benefit for the male than female listeners). As has been suggested previously by Ravicz *et al.* (2001), it is likely that additional modifications to reduce these sources of bone conduction by isolating the head and body from the surrounding noise field would improve the perceptual performance of the ANC system.

III. EXPERIMENT 2: BENEFITS FOR DETECTABILITY OF SOUND-RELATED BRAIN ACTIVITY MEASURED USING fMRI

Given that the ANC system significantly improves the detectability of a 600-Hz target that coincides with the main peak in the spectrum of the EPI noise, cancellation should also improve the statistical detection of frequency-dependent activation produced by a 600-Hz signal in an fMRI experiment. We expected to obtain an improved activation pattern with cancellation not only because the target is more easily perceived, but also because the effect of the background noise in the baseline condition is reduced. The rationale for the latter effect is as follows. The most straightforward fMRI paradigm detects sound-related activity in a relative manner

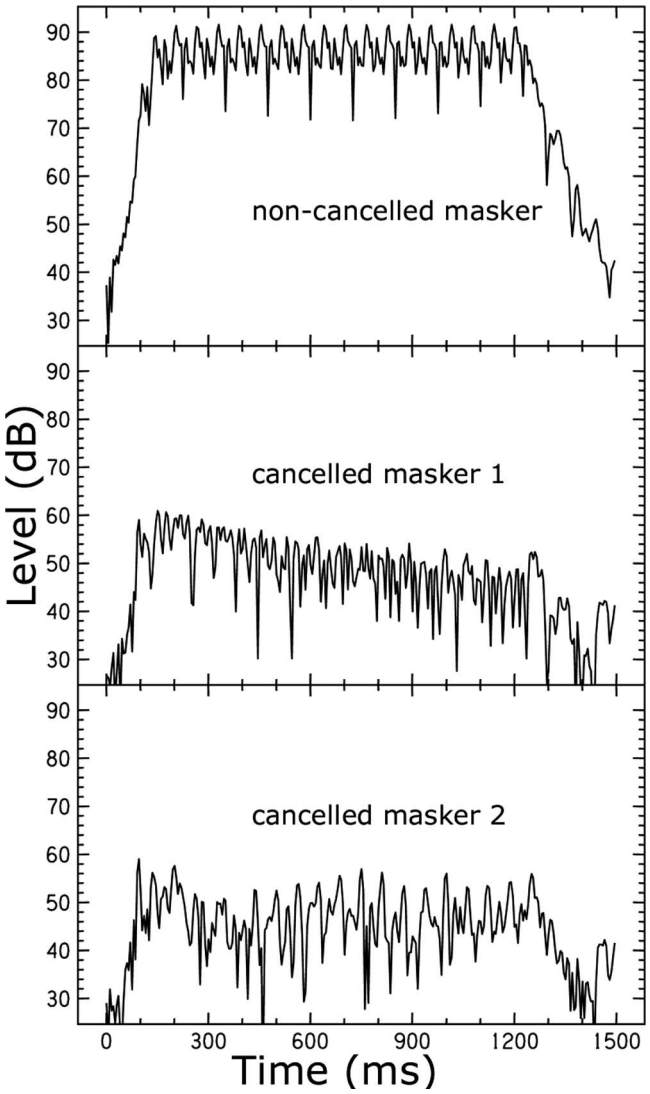


FIG. 3. Envelopes of the recordings of three noise maskers taken from one adaptive track of listener 4. The top panel shows one of the non-cancelled maskers at the beginning of the run, and the other two panels show two separate cancelled maskers from later in the run. The envelopes were calculated using the Hilbert transform, using edited waveforms of 1500-ms duration that were approximately centered on the scan. The waveforms were initially filtered to a 100-Hz wide band, centered on 600 Hz (see Sec. II B).

by subtracting brain images acquired during a baseline condition in which the sound is absent from those for which sound stimuli are presented. Given that the EPI noise itself is known to evoke significant auditory cortical activity (e.g., Bandettini *et al.*, 1998; Hall *et al.*, 2000; Ulmer *et al.*, 1998), a high intensity of the background noise elevates the baseline of the auditory cortical response, thus reducing the dynamic range for detecting the additional activity evoked by the stimulus of interest.

Accordingly, experiment 2 sought to demonstrate the neuroimaging benefit of the ANC system by measuring frequency-dependent activation for three listeners using signals presented at two intensities, with and without cancellation, in a same-different frequency-discrimination task. This paradigm addresses the most fundamental question in auditory fMRI, namely the sensitivity for detecting sound-related activity. The attenuated condition presents an interesting case

TABLE II. Masked perceptual thresholds in the cancelled and noncancelled conditions for experiments 1 and 3. Mean levels are reported in dB SPL, with standard errors in parentheses.

Center frequency	Listener	Noncancelled threshold	Cancelled threshold	Perceptual benefit
Exp. 1				
600 Hz	1	76 (0.6)	59 (1.2)	17 (1.2)
	2	73 (1.0)	55 (1.3)	18 (1.3)
	3	73 (1.0)	53 (1.3)	20 (1.3)
	4	73 (1.7)	50 (1.1)	23 (1.1)
Exp. 3				
600 Hz	5	81 (1.1)	64 (1.0)	17 (1.0)
	6	76 (0.7)	65 (1.2)	11 (1.2)
	7	76 (0.8)	59 (0.6)	17 (0.6)
	8	81 (0.8)	66 (0.8)	15 (0.8)
Exp. 3				
150 Hz	Mean	69 (0.5)	65 (0.4)	4 (0.6)
300 Hz	Mean	69 (0.8)	59 (0.5)	10 (1.2)
450 Hz	Mean	64 (1.1)	54 (0.6)	10 (1.4)
600 Hz	Mean	78 (1.1)	63 (0.6)	15 (1.1)

in which cancellation exerts the greatest effect on stimulus audibility. The target was intended to be easy to hear when the ANC system was cancelling the masker noise, but inaudible when the ANC system was not operational.

A. Methods

1. Stimuli

The stimuli were constructed in the same manner as for experiment 1, except that they had center frequencies of 570, 600, or 630 Hz, and a duration of 450 ms. Stimuli were concatenated to form pairs, separated by a 50-ms silent gap. In total, there were 60 same-frequency pairs (20 of each combination of 570/570, 600/600, and 630/630 Hz) and 60 different-frequency pairs (30 of each combination of 570/600 and 600/630 Hz). The order of the two stimuli in the different-frequency pairs was counterbalanced. The total rms power was equated across stimulus pairs. Each condition contained a sequence of eight stimulus pairs, four same-frequency and four different-frequency trials, presented in a randomized order.

Six conditions were formed by crossing three signal levels ("94 dB," "attenuated," and "no signal") with two ANC states ("noncancelled" and "cancelled"). The 94-dB condition represents the maximum possible level of the 600-Hz signal and the stimuli at this level were clearly audible in both cancelled and noncancelled conditions. The signal level in the attenuated condition was 3 dB below that individual's mean 600-Hz detection threshold in the noncancelled ANC condition (as measured in experiment 1). The audibility of the attenuated stimuli was therefore dependent on the ANC state. The physical levels of the stimuli and the resulting sensation levels for each listener are reported in Table III. In the no-signal condition, the signal was not present. This condition served as the baseline comparison in a typical subtraction analysis of the fMRI data.

TABLE III. The physical levels of the stimuli and the resulting sensation levels, for each listener who participated in experiment 2. The sensation levels for the 94-dB conditions are determined from the individual cancelled and noncancelled thresholds reported in experiment 1 (Table II).

		Sound level, dB SPL			Sensation level, dB		
		Listener			Listener		
Stimulus	ANC state	1	2	4	1	2	4
94 dB	Noncancelled	94	94	94	18	21	21
94 dB	Cancelled	94	94	94	35	39	44
Attenuated	Noncancelled	73	70	70	−3	−3	−3
Attenuated	Cancelled	73	70	70	14	15	20

A trial comprized a stimulus pair that was presented at the same time as the MR scanner acquired an 18-slice set of brain images. These scans, and the characteristics of their EPI noises, were the same as those used in experiment 1 (see Fig. 1). However, in the present experiment, a conventional MR pulse sequence was used, i.e., sets of brain images were acquired repeatedly (at 2.7 s intervals) and there were no dummy scans. The onset of the stimulus pair always occurred 0.05 s after the onset of the masker, and again stimulus pairs were presented at a rate of one every 2.7 s.

2. Procedures

Listeners were required to decide whether the two members of each pair were the same or different and to respond by pressing one of two buttons with the index finger of their right or left hand. The quiet interval of 1.55 s between trials was sufficient to allow this response. Listeners were instructed to guess when their response was uncertain and also to guess when the stimulus pair was inaudible.

The presentation order was chosen to ensure equal distribution of the six conditions across four, 9-min experimental runs. The conditions cycled in a pseudorandom order, such that there were never more than 64 stimulus pairs between repetitions of the same condition. In addition, throughout the experiment, the ANC system cycled between cancelled and noncancelled conditions after every 24 stimulus pairs (i.e., after every three conditions). Each of the six conditions occurred 16 times, giving a total dataset of 6×128 ($N=768$) trials and the same number of brain scans.

The fMRI procedure positioned the brain slices in the same orientation as in experiment 1. To optimize coverage of the brain within a slice, the in-plane field of view was chosen to be 192 mm with a 64×64 matrix, corresponding to an image resolution (i.e., voxel size) of $3\times3\times3$ mm. At the end of the four experimental runs, we also acquired a $1\times1\times1$ mm resolution anatomical scan of the whole brain to enable precise anatomical localization of auditory cortical activity. The anatomical scan took $4\frac{1}{2}$ min.

3. Listeners

Three right-handed listeners participated in experiment 2. They were listeners 1, 2, and 4 from experiment 1.

4. fMRI analysis

A standard fMRI analysis was conducted separately for each listener using Statistical Parametric Mapping v2 (SPM2) software running on MATLAB v6.5.0. To minimize the effects of minor head movements that typically occur over the scanning period, all of the scans in the four runs were realigned using the last scan of the first run as the reference. Within runs, the adjustment for movement was never more than 3 mm and 1°, indicating that the listeners had complied with the instructions to remain as still as possible. To interpret the anatomical location of sound-evoked activation using a standard atlas of the brain and to compare across different brains, the individual brain images were next transformed into the brain space defined by the template of the Montreal Neurological Institute. The aim of this procedure was to match the cortical volume of each individual brain to that of the "standard" brain. The final spatial processing step involved smoothing these scans by a Gaussian kernel of 4 mm full width at half maximum. The spatially transformed images of each participant provided the input to the analysis of the sound-evoked brain activation. The analysis used the General Linear Model that is implemented in SPM2. The model partitions the observed response according to a sum of weighted variables (defining the experimental conditions). We used a model that specified the chronological sequence of events using 12 experimental variables for each run; six comprised the main sustained response associated with each experimental condition and six others were included to capture any differences in the latency of the peak response from the canonical form of the haemodynamic response function by using its temporal derivative. A final set of four variables accounted for any large between-run differences in the mean image intensity. Low-frequency artifacts in the time series were removed by applying a high-pass filter with a cutoff of 0.003 Hz. After model estimation, statistical contrasts between conditions were specified by linear combinations of these variables and the significance of each contrast was determined based on the scan-to-scan residual variability. Four major contrasts were specified. Each contrast was a pair-wise comparison between a particular sound condition [namely: the (i) cancelled 94 dB, (ii) cancelled attenuated, (iii) noncancelled 94 dB, or (iv) noncancelled attenuated] and its corresponding no signal baseline condition, with a probability criterion for statistical significance of 0.001. For listeners 1 and 2, the variables for the main sus-

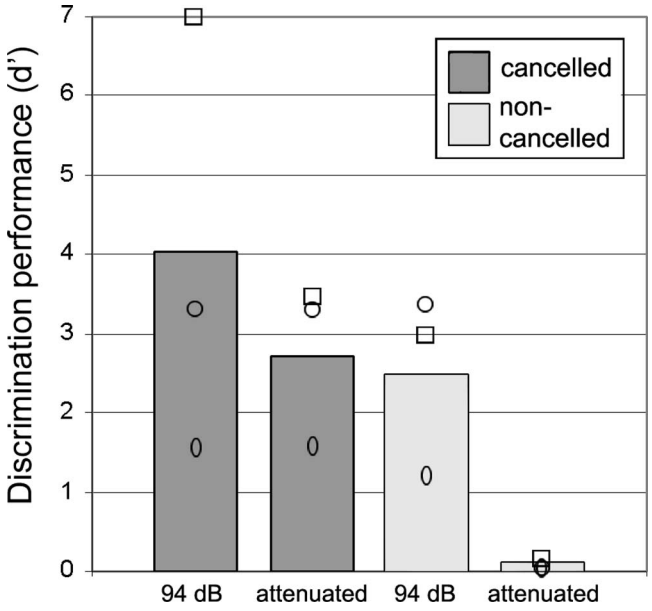


FIG. 4. Same/different frequency discrimination performance for experiment 2 measured using d'. Mean performance is shown by the height of the bar, whilst individual scores are represented by the ellipse (listener 1), the square (listener 2), and the circle (listener 4).

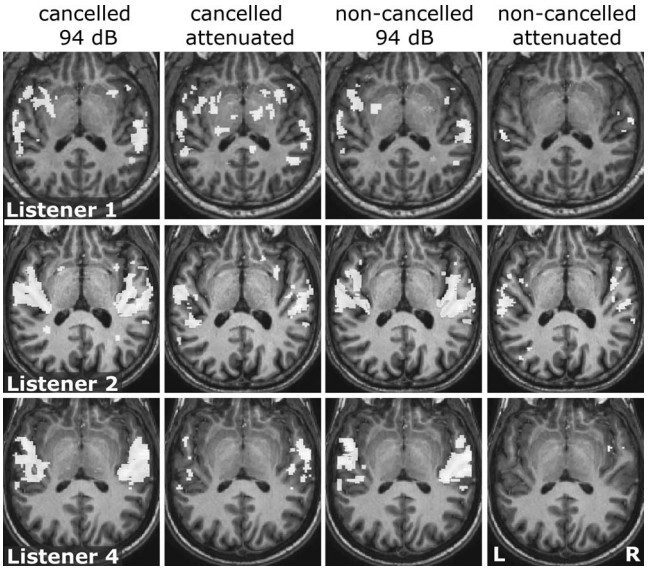


FIG. 5. Sound-evoked brain activity for experiment 2. Each row displays the results for an individual listener overlaid onto a slice that is oriented through auditory cortical regions on the upper surface of the superior temporal gyrus. Each column displays the pattern of significant brain activity, in white, for one of the four major contrasts (voxel-level threshold, $p<0.001$).

tained response were sufficient to capture sound-evoked brain activity, whereas for listener 4 it was necessary to add in the variables for additional shifts in the onset and offset of the evoked response.

The extent of sound-evoked activity was quantified using two auditory regions of interest that were defined using the outer-most borders from two standardized templates. The primary auditory cortex on Heschl's gyrus was defined using criteria from Morosan *et al.* (2001), while the posterior nonprimary auditory cortex on the planum temporale was defined using criteria from Westbury *et al.* (1999). The latter region excluded any voxels that overlapped with the primary region to ensure that voxels in the two regions were mutually exclusive.

B. Results

1. Behavioral performance

All three listeners reliably judged ($d'>1.2$) whether the stimulus pair were same or different in all the three conditions in which the stimuli were clearly detectable (cf., Reed and Bilger, 1973), including the attenuated signals when the masker level was reduced by ANC (Fig. 4). In contrast, performance was poor ($d'<0.2$) for the attenuated signals presented without ANC.

2. Sound-evoked brain activity

Brain activity was examined by overlaying voxels that showed a significant pair-wise effect onto the listener's anatomical scan. Suprathreshold activation occurred in auditory cortical regions, but not in subcortical structures, such as the medial geniculate body. For all three listeners, activity encompassed parts of the primary auditory cortex on Heschl's gyrus and the posterior nonprimary auditory cortex on planum temporale, although for listener 1 activation was more restricted to lateral regions. This is illustrated in Fig. 5, which shows a benefit of cancellation in the attenuated condition for all three listeners. The extent of auditory cortical activity is quantified in Fig. 6. Although the extent of sound-evoked activity was reduced for stimuli in the attenuated condition, a significant advantage for using ANC as found for the nonprimary auditory region ($t[5]=4.5$, $p<0.01$). There was also a trend towards an advantage in primary auditory cortex, but this did not reach significance ($p=0.21$).

In summary, experiment 2 demonstrated that for individual listeners, ANC offers significant advantages for the fMRI detection of sound-related activity at a low level. The extent of activity did not, however, simply increase monotonically with the sensation level of the stimulus. This is due to the fact that sound-evoked activation is determined by

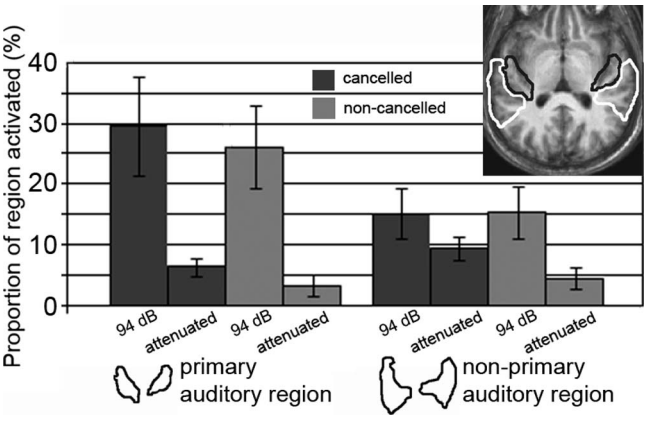


FIG. 6. The extent of sound-evoked auditory activity (voxel-level threshold, $p<0.001$) within two auditory cortical regions. The primary auditory region contains a total of 2172 voxels across both hemispheres. The (posterior) nonprimary region contains a total of 7982 voxels. Extent of activation is plotted separately for the different sound levels and for the different ANC conditions, with standard error bars.

subtracting images for the baseline from the stimulus conditions and so it is influenced by the level of the EPI noise, as well as the level of the stimulus (see also Hall *et al.*, 2001; Hart *et al.*, 2003; Scott *et al.*, 2004). The particular benefit provided by ANC for detecting sound-related activation in nonprimary regions may reflect the decreased listening effort required to segregate stimuli from the background noise (Griffiths and Warren, 2002).

IV. EXPERIMENT 3: EFFECTS OF CANCELLATION AT OTHER FREQUENCIES OF THE EPI NOISE

Experiment 3 investigated the acoustical and perceptual benefits of cancellation for target stimuli presented at frequencies away from the prominent peak of energy in the EPI noise. In this experiment, we measured detection thresholds for narrow bands of noise whose center frequencies corresponded to other prominent low-frequency peaks in the spectrum of the EPI noise—namely 150 and 300 Hz—and which were attenuated by at least 10 dB by ANC. We also examined the effects for an additional frequency (450 Hz) that corresponds to a "dip" in the spectrum of the EPI noise.

A secondary goal was to explore the effect of fixing the coefficients of the adaptive filter after the initial training period and to document the stability of the noise reduction over time. Fixing these coefficients effectively "freezes" the cancellation signal. This was deemed a reasonable way to implement ANC because both the gradient coils that generate the EPI noise and the trigger pulse that synchronises the DSP contain a clock that is controlled by a crystal oscillator, and are therefore highly stable over successive repetitions (Chambers *et al.*, 2007). For example, Chambers *et al.* (2007) reported that time variations ($<\pm 0.1$ μs) between a trigger being received and the DSP system starting a conversion were insufficient to affect cancellation performance. Documenting the levels of the masker noise and the target over time enabled us to quantify the performance of the "fixed" ANC system.

A. Methods

1. Stimuli

The targets were diotic narrowband noises centered on 150, 300, 450, and 600 Hz. The stimuli were created in the same way as for experiment 1. They had 50-Hz bandwidth and 500-ms duration (including 50-ms raised-cosine ramps applied to the onset and offset) and were sampled at a rate of 44 100 Hz. The presentation level was varied according to the adaptive procedure by applying a digital attenuation before digital-to-analog conversion. For the 150-, 300-, 450-, and 600-Hz targets, 0-dB attenuation corresponded to levels of 94, 95, 94, and 94 dB SPL, respectively. The acoustic structure of the masking noise is described in Sec. II A 2. The target was presented randomly within one of the two maskers and its onset always occurred 0.40 s after the onset of the masker.

2. Procedures

We used the same bespoke EPI pulse sequence to generate a noise sequence that allowed us to conduct a two-interval psychophysical paradigm. Detection thresholds for the targets were measured using a 2I-2AFC three-down, one-up adaptive procedure to track 79% correct. The listener was required to judge whether the target appeared in the first or second interval, responding appropriately by pressing one of two buttons with the index finger of either their left or right hand. As in experiment 1, listeners completed a practice-and-familiarization session detecting the 600-Hz target in the plastic-and-wood replica of an MR scanner.

The procedure for initializing the ANC system was as described before (Sec. II A 1), except that the coefficients of the adaptive filter (and hence, the characteristics of the cancellation signal) were fixed after the training scans. On the first two practice trials, the initial attenuation levels of the target were 15 dB in the cancelled condition and 5 dB in the noncancelled condition. Hence, the target always started off being clearly audible. Thereafter, the level changed in 5-dB steps until two reversals occurred, and 2-dB steps thereafter. For the first two listeners (5 and 6), each adaptive run ended after six reversals. Three runs were obtained for each condition and the mean detection threshold was taken as the average of the levels at the last four reversal points across runs. Each run contained 30–40 trials and lasted about $3\frac{1}{2}$ min. Although a total time of 21 min was required to obtain the cancelled and noncancelled thresholds for one target frequency, the significant amount of setting-up time in-between runs made these sessions very long and tiring. Consequently, the final two listeners (7 and 8) completed just two runs for each condition, each terminating after ten reversals. The mean detection threshold was taken as the average of the levels at the last eight reversal points across runs. Using this protocol, each adaptive run contained 50–60 trials and lasted about $5\frac{1}{2}$ min. Although the total acquisition time remained about the same as before, the sessions were considerably shortened by the reduction in setting-up time. During each adaptive run, the residual sound was monitored by the error microphone under the left and right ear defenders and recorded onto a DAT recorder (Sony DTC-690) for off-line evaluation. Individual measurements were completed in two $1\frac{1}{2}$ h testing sessions, the order of conditions being counterbalanced across sessions and listeners.

3. Listeners

Four listeners took part in experiment 3, one male (5) and three females (6–8). Listeners were recruited from the undergraduate population at Nottingham University and had a median age of 21.5 years (range 20–22). All listeners had normal hearing as assessed using standard pure-tone audiometry (<20 dB HL over octave frequencies between 250 and 8000 Hz).

B. Analysis

The DAT recordings were analyzed to determine the sound level at the error microphone over the course of each adaptive run using the procedures reported in experiment I (Sec. II B). Again, we were interested only in the amount of cancellation at the frequency corresponding to the target and so the digitized recordings were first filtered to remove sound

TABLE IV. Mean sound levels measured in dB SPL at the listeners' ears across different frequencies of interest in experiment 3. The masker levels reported for the cancelled condition correspond to those trials at the reversal points in the adaptive run. The mean threshold level of each target signal is provided for comparison with the masker levels in the better ear. Values in parentheses represent the standard error of the measurement.

Center frequency (Hz)	Recording scan (noncancelled)	Cancelled			Acoustic noise reduction (better and worse ear mean)
	Masker (better and worse ear mean)	Masker in better ear	Masker in worse ear	Signal level	
150	—	—	—	51 (0.4)	—
300	91 (0.8)	66 (0.5)	72 (0.6)	59 (0.5)	22 (1.0)
450	75 (0.9)	62 (0.6)	68 (0.2)	54 (0.6)	10 (2.3)
600	94 (0.4)	70 (0.7)	75 (0.9)	63 (0.6)	21 (2.0)

energy that was remote from the frequency of interest. We chose bandwidths of 60, 80, and 100 Hz as convenient approximations to the equivalent rectangular bandwidths for 300, 450, and 600 Hz, respectively (Glasberg and Moore, 1990). Filtering the digitized recordings for the 150-Hz condition, using a filter bandwidth of 40 Hz, did not yield meaningful estimates because the level of the cancelled masker approached that of the ambient noise. This was continually present and primarily reflects the air conditioning in the room.

The perceptual benefits of noise reduction for each individual were computed by subtracting the mean detection threshold in the noncancelled condition from the mean detection threshold in the cancelled condition.

C. Results

1. Acoustic noise reduction during the cancelled condition

The results for target frequencies of 300, 450, and 600 Hz demonstrated that the ANC system operated effectively across a range of low frequencies and attenuated both peaks and dips in the acoustic spectrum. Table IV shows the noncancelled and cancelled masker levels for each center frequency, computed from the recording made at the microphone under the earcups during the tests. On average, the ANC system provided similar amounts of acoustic noise reduction at 300 and 600 Hz (22 and 21 dB, respectively), but somewhat less noise reduction at 450 Hz (10 dB)[3]. Once again, there were differences of about 4 dB in the degree of cancellation across the ears of a given subject.

We quantified the effect of fixing the adaptive-filter coefficients by comparing the performance of the ANC system at 600 Hz in experiment 3 with its corresponding performance in experiment 1, in which the coefficients were allowed to adapt slowly over the run. The noise reduction was smaller in experiment 3 than in experiment 1 (21 dB versus 35 dB), as the residual noise level was higher in experiment 3 (72 versus 58 dB SPL). An example of the performance of the ANC system over the course of an adaptive run is shown in Fig. 2(b) (recorded from listener 6). Despite an initial noise reduction of over 30 dB, the noise reduction began to diminish immediately after the second training scan (the moment at which the adaptive filter coefficients were fixed). Thereafter, there was a steady deterioration over time, especially for the right ear.

2. Perceptual benefits of noise reduction

The mean perceptual benefits are reported in Table II. As expected, the perceptual benefit was greatest at 600 Hz, the frequency at which the energy of the EPI noise was maximal, although reliable improvements were obtained at all other frequencies.

We quantified the effect of fixing the adaptive-filter coefficients by comparing target detection thresholds at 600 Hz across Experiments 1 and 3 (see Table II). Although the noncancelled thresholds were broadly similar, the cancelled thresholds were always higher in experiment 3. This is consistent with the view that listening performance in the cancelled condition was compromised by the poorer acoustic performance of the ANC system, due to the fixing of the adaptive-filter coefficients.

V. GENERAL DISCUSSION

These experiments demonstrate that ANC can be achieved in the electromagnetically hostile and physically compact neuroimaging environment. The ANC system does not involve a significant increase in scanning time and can be used with minimal training. Although the present paper reports the levels of acoustic noise reduction over time, these detailed measurements are not necessary under normal operating conditions. More routine on-line monitoring of the residual sound at the listener's ears is provided by two level meters on the front of the DSP system. Consequently, routine use of the ANC system does not prevent scanning large numbers of participants (N=10–20). We, therefore, recommend that ANC should be implemented whenever possible during auditory fMRI research.

As well as improving the audibility of sounds presented to participants, ANC increased the sound-evoked fMRI activity for stimuli presented at low levels. The noise reduction was obtained across a range of low frequencies corresponding to both peaks and dips in the spectrum of the EPI noise and reduction was greatest when the ANC system continually adjusted the cancellation signal to maintain optimal performance over time.

One of the main advantages of ANC is that it enables experimenters to present sound stimuli at much lower levels than has been possible hitherto during fMRI. Typically in such studies, signals are delivered at moderate to high sound levels (see Hall, 2005, for a review). Although fMRI research has made important advances in characterizing the functional anatomy of human tonotopy, especially in the primary auditory region on the medial two-thirds of Heschl's gyrus, it is important to remember that the response pattern of a given neuron can change as a function of level. Thus, it may not be valid to directly compare these human fMRI results with data from microelectrode recordings in animals that localize the best frequency response near threshold because the cortical area responsive to a suprathreshold stimulus cannot be predicted from the threshold map (Phillips *et al.*, 1994). ANC offers the possibility to reexamine issues such as human tonotopy using sounds with levels as much as 20 dB below those of the noncancelled equivalent.

ANC also provides an opportunity to address questions that have previously been confounded by the excessive EPI noise levels. One example concerns measures of the neural correlates of speech intelligibility. Davis and Johnsrude (2003) have shown a clear association between the speech-to-noise ratio and speech-related activation in many brain areas including the superior and middle temporal gyri, left inferior frontal gyrus, and left hippocampus. Activity within these brain areas increases as a function of speech intelligibility. Clearly, a reduction in background noise level by 20 dB should markedly improve speech intelligibility, and hence, influence the resulting pattern of activity. A second example concerns measures of the neural basis of temporal coding for complex tones (Hall *et al.*, 2006). Coding of periodicity pitch can be examined using harmonic complex tones in which the first few harmonic components are missing from the spectrum. However, nonlinear interactions on the basilar membrane can reconstitute the lower part of the harmonic series as a series of distortion products. Psychophysical evidence demonstrates that, although the components of the distortion spectrum are reduced in level by 10 dB and more compared to the primaries, many exceed the hearing threshold (Pressnitzer and Patterson, 2001). Furthermore, the higher the presentation level of the complex tone, the greater the intensity and number of distortion components. The use of ANC allows a lower presentation level than otherwise in an fMRI experiment, and so any distortion products will be less of a concern[4].

While it is possible that future improvements to the ANC system may increase the acoustic noise reduction, if the limiting step is bone conduction, then such gains will not be apparent to the listener. Ravicz *et al.* (2001) evaluated different combinations of earplugs, ear defenders and/or a sound-absorbing helmet and concluded that (at least for frequencies <500 Hz) air was the dominant conduction route for every device configuration tested. Using an ANC system in a laboratory setting to measure detection thresholds for a tone masked by white noise, Chambers *et al.* (2001) found that the detection thresholds were improved by a substantial amount, but not by as much as the acoustic noise reduction for frequencies below 1 kHz. For example, at 500 Hz, the average perceptual and acoustic benefits were 17 and 23 dB, respectively. The maximum perceptual benefit that can be gained using methods that treat only sound conduction through the head (e.g., ear defenders combined with ANC) is determined by whether the dominant transmission path of sound to the cochlea is airborne or bone conducted. It is likely that accurate estimates of the upper bound of noise reduction for vibrations that are mechanically coupled to the head (as in this case when listeners are lying in the head coil of the MR scanner) would be difficult to ascertain, and indeed have not been attempted to the best of our knowledge. Standard laboratory measurements of free-field bone conduction are often made using one or more loudspeakers at some distance from the listener (e.g., Berger, 1983; Berger and Kerivan, 1983) and so may not be directly applicable to the MR environment, where the listener is lying inside the source of the noise and, moreover, the source is physically large and cannot be assumed to be a point source. Thus, while bone conduction could be limiting the perceptual benefits from cancellation, we do not have the measurements to quantify its effect.

In summary, the present experiments have demonstrated the effectiveness of an ANC system for fMRI experimentation, although we found that the perceptual benefit was somewhat less than the acoustic benefit. Further improvements to the ANC system could reduce this difference. For example, placing the error microphone nearer to the listener's ear drum rather than just under the ear defender could provide additional improvements in the efficacy of the cancellation signal. However, if the perceptual benefit is indeed limited by bone conduction, then a combined approach will be required to treat both transmission routes. A final point to bear in mind is that not only does ANC improve auditory perception, but it also provides important subjective benefits in terms of the well-being of the participant who is tolerating the rather hostile environment.

ACKNOWLEDGMENTS

Dr. Yuvi Kahana at Optoacoustics, Ltd., Or Yehuda, Israel, kindly provided the optical microphones. Kimberley Quinn helped to test participants in experiment 3 and also contributed to a pilot version of experiment 2 as part of her final-year, BSc Physics project at Nottingham Trent University. We are also grateful to Brian Moore (Associate Editor) and two anonymous reviewers for their insightful comments on an earlier version of the manuscript.

[1]The term "noise" is commonly used in the imaging literature to characterize the distraction and annoyance of the sound produced by the scanner, and also to emphasize the desirability of attenuating, or removing, it. Perceptually, however, the acoustic signal generated by the scanner during EPI sounds more like a sharp, intense ping, with a clear tonal character, rising above a broadband spectrum of noise that is at a much lower intensity.

[2]This is the only period in which the scanner room is quiet. Although the scanner room itself was not sound treated, at 57 dB(A), the background level was well below the level of the EPI noise.

[3]Although Fig. 1 indicates that there is no acoustic reduction at 450 Hz, our report of 10 dB of acoustic reduction is based on analysis using a narrowband filtered signal, with cutoffs at 410 and 490 Hz.

[4]The reader is reminded that the EPI noise is a rapid sequence of sharp

pings that does not form a continuous noise and it so cannot be assumed to mask the distortion products whatever their level and whenever they occur.

Amaro, E., Jr., Williams, S. C. R., Shergill, S. S., Fu, C. H. Y., MacSweeney, M., Picchioni, M. M., Brammer, M. J., and McGuire, P. K. (**2002**). "Acoustic noise and functional magnetic resonance imaging: Current strategies and future prospects," J. Magn. Reson Imaging **16**, 497–510.

Bandettini, P. A., Jesmanowicz, A., Kylen, J. Van, Birn, R. M., and Hyde, J. S. (**1998**). "Functional MRI of brain activation induced by scanner acoustic noise," Magn. Reson. Med. **39**, 410–416.

Berger, E. H. (**1983**). "Laboratory attenuation of earmuffs and earplugs both singly and in combination," Am. Ind. Hyg. Assoc. J. **44**, 321–329.

Berger, E. H., and Kerivan, J. E. (**1983**). "Influence of physiological noise and the occlusion effect on the measurement of real-ear attenuation at threshold," J. Acoust. Soc. Am. **74**, 81–94.

Boyle, Y., Bentley, D. E., Watson, A., and Jones, A. K. P. (**2006**). "Acoustic noise in functional magnetic resonance imaging reduces pain unpleasantness ratings," Neuroimage **31**, 1278–1283.

Chambers, J., Akeroyd, M. A., Summerfield, A. Q., and Palmer, A. R. (**2001**). "Active control of the volume acquisition noise in functional magnetic resonance imaging: Method and psychoacoustical evaluation," J. Acoust. Soc. Am. **110**, 3041–3054.

Chambers, J., Bullock, D., Kahana, Y., Kots, A., and Palmer, A. R. (**2007**). "Developments in active noise control sound systems for magnetic resonance imaging," Appl. Acoust. **68**, 281–295.

Chapman, B. L. W., Haywood, B., and Mansfield, P. (**2003**). "Optimized gradient pulse for use with EPI employing active acoustic control," Magn. Reson. Med. **50**, 931–935.

Davis, M. H., and Johnsrude, I. S. (**2003**). "Hierarchical processing in spoken language comprehension," J. Neurosci. **23**, 3423–3431.

Elliott, M. R., Bowtell, R. W., and Morris, P. G. (**1999**). "The effect of scanner sound in visual, motor, and auditory functional MRI," Magn. Reson. Med. **41**, 1230–1235.

Foster, J. R., Hall, D. A., Summerfield, A. Q., Palmer, A. R., and Bowtell, R. W. (**2000**). "Sound-level measurements and calculations of safe noise dosage during EPI at 3 T," J. Magn. Reson Imaging **12**, 157–163.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gockel, H., Moore, B. C. J., and Patterson, R. D. (**2002**). "Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression," J. Acoust. Soc. Am. **111**, 2759–2769.

Granet, R., and Gelber, L. (**1990**). "Claustrophobia during MR imaging," N. J. Med. **87**, 479–482.

Griffiths, T. D., and Warren, J. D. (**2002**). "The planum temporale as a computational hub," Trends Neurosci. **25**, 348–353.

Hall, D. A. (**2005**). "Representations of spectral coding in the human brain," in *Auditory Spectral Processing. International Review of Neurobiology, Volume 70*, edited by M. S. Malmierca and D. R. F. Irvine (Elsevier, San Diego), pp 331–369.

Hall, D. A. (**2006**). "Haemodynamic-based imaging of central auditory function," in *Auditory Evoked Potentials: Basic Principles and Clinical Application*, edited by R. Burkard, M. Don, and J. Eggermont (Lippincott, Williams and Wilkins, Baltimore), Chap. 2, pp. 22–40.

Hall, D. A., Edmondson-Jones, A. M., and Fridriksson, J. (**2006**). "Frequency and periodicity coding in human auditory cortex," Eur. J. Neurosci. **24**, 3601–3610.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E., and Bowtell, R. W. (**1999**). "Sparse temporal sampling in auditory fMRI," Hum. Brain Mapp **7**, 213–223.

Hall, D. A., Haggard, M. P., Summerfield, A. Q., Akeroyd, M. A., Palmer, A. R., and Bowtell, R. W. (**2001**). "FMRI measurements of sound-level encoding in the absence of background scanner noise," J. Acoust. Soc. Am. **109**, 1559–1570.

Hall, D. A., Hart, H. C., and Johnsrude, I. S. (**2003**). "Relationships between human auditory cortical structure and function," Audiol. Neuro-Otol. **8**, 1–18.

Hall, D. A., Summerfield, A. Q., Gonçalves, M. S., Foster, J. R., Palmer, A. R., and Bowtell, R. W. (**2000**). "Time-course of the auditory BOLD response to scanner noise," Magn. Reson. Med. **43**, 601–606.

Hart, H. C., Hall, D. A., and Palmer, A. R. (**2003**). "The sound-level-dependent growth in the extent of fMRI activation in Heschl's gyrus is different for low- and high-frequency tones," Hear. Res. **179**, 104–112.

Hennel, F., Girard, F., and Loenneker, T. (**1999**). " 'Silent' MRI with soft gradient pulses," Magn. Reson. Med. **42**, 6–10.

Johnsrude, I. S., Giraud, A. L., and Frackowiak, R. S. J. (**2002**). "Functional imaging of the auditory system: The use of positron emission tomography," Audiol. Neuro-Otol. **7**, 251–276.

Kohlrausch, A., and Sander, A. (**1995**). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets," J. Acoust. Soc. Am. **97**, 1817–1829.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

More, S. R., Lim, T. C., Li, M. F., Holland, C. K., Boyce, S. E., and Lee, J. H. (**2006**). "Acoustic noise characteristics of a 4 Tesla MRI scanner," J. Magn. Reson Imaging **23**, 388–397.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., and Zilles, K. (**2001**). "Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system," Neuroimage **13**, 684–701.

Nelson, P. A., and Elliott, S. J. (**1992**). *Active Control of Sound* (Academic, London).

Phillips, D. P., Semple, M. N., Calford, M. B., and Kitzes, L. M. (**1994**). "Level-dependent representation of stimulus frequency in cat primary auditory cortex," Exp. Brain Res. **102**, 210–226.

Pressnitzer, D., and Patterson, R. D. (**2001**). "Distortion products and the perceived pitch of harmonic complex tones," in *Physiological and Psychophysical Bases of Auditory Function*, edited by A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker Publishing, Maastricht), pp. 97–104.

Quirk, M., Letendre, A., Ciottone, R., and Lingley, J. (**1989**). "Anxiety in patients undergoing MR imaging," Radiology **170**, 463–466.

Ravicz, M. E., and Melcher, J. R. (**2001**). "Isolating the auditory system from acoustic noise during functional magnetic resonance imaging: Examination of noise conduction through the ear canal, head, and body," J. Acoust. Soc. Am. **109**, 216–231.

Ravicz, M. E., Melcher, J. R., and Kiang, N. Y.-S. (**2000**). "Acoustic noise during functional magnetic resonance imaging," J. Acoust. Soc. Am. **108**, 1683–1696.

Reed, C. M., and Bilger, R. C. (**1973**). "A comparative study of S/N and E/N_0," J. Acoust. Soc. Am. **53**, 1039–1045.

Schwarzbauer, C., Davis, M. H., Rodd, J. M., and Johnsrude, I. S. (**2006**). "Interleaved silent steady state (ISSS) imaging: A new sparse imaging method applied to auditory fMRI," Neuroimage **29**, 774–782.

Scott, S. K., Rosen, S., Wickham, L., and Wise, R. S. J. (**2004**). "A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception," J. Acoust. Soc. Am. **115**, 813–821.

Shah, N. J., Jancke, L., Grosse-Ruyken, M. L., and Muller-Gartner, H. W. (**1999**). "Influence of acoustic masking noise in fMRI of the auditory cortex during phonetic discrimination," J. Magn. Reson Imaging **9**, 19–25.

Ulmer, J. L., Biswal, B. B., Yetkin, F. Z., Mark, L. P., Mathews, V. P., Prost, R. W., Estkowski, L. D., McAuliffe, T. L., Haughton, V. M., and Daniels, D. L. (**1998**). "Cortical activation response to acoustic echo planar scanner noise," J. Comput. Assist. Tomogr. **22**, 111–119.

Westbury, C. F., Zatorre, R. J., and Evans, A. C. (**1999**). "Quantifying variability in the planum temporale: A probability map," Cereb. Cortex **9**, 392–405.

www.ihr.mcr.ac.uk/research/technical.php (Last viewed 11/10/2008).

www.fil.ion.ucl.ac.uk/spm (Last viewed 11/10/2008).

Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids

Tim Van den Bogaert[a)]
ExpORL, K.U. Leuven, O&N2, Herestraat 49, Bus 721, B-3000 Leuven, Belgium

Simon Doclo
ESAT-SCD, K.U. Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

Jan Wouters
ExpORL, K.U. Leuven, O&N2, Herestraat 49, Bus 721, B-3000 Leuven, Belgium

Marc Moonen
ESAT-SCD, K.U. Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

(Received 28 March 2008; revised 1 August 2008; accepted 21 October 2008)

This paper evaluates speech enhancement in binaural multimicrophone hearing aids by noise reduction algorithms based on the multichannel Wiener filter (MWF) and the MWF with partial noise estimate (MWF-N). Both algorithms are specifically developed to combine noise reduction with the preservation of binaural cues. Objective and perceptual evaluations were performed with different speech-in-multitalker-babble configurations in two different acoustic environments. The main conclusions are as follows: (a) A bilateral MWF with perfect voice activity detection equals or outperforms a bilateral adaptive directional microphone in terms of speech enhancement while preserving the binaural cues of the speech component. (b) A significant gain in speech enhancement is found when transmitting one contralateral microphone signal to the MWF active at the ipsilateral hearing aid. Adding a second contralateral microphone showed a significant improvement during the objective evaluations but not in the subset of scenarios tested during the perceptual evaluations. (c) Adding the partial noise estimate to the MWF, done to improve the spatial awareness of the hearing aid user, reduces the amount of speech enhancement in a limited way. In some conditions the MWF-N even outperformed the MWF possibly due to an improved spatial release from masking.
 [DOI: 10.1121/1.3023069]

PACS number(s): 43.66.Pn, 43.66.Ts, 43.60.Fg, 43.66.Qp [RYL] Pages: 360–371

I. INTRODUCTION

Hearing aid users often have great difficulty understanding speech in a noisy background. They typically require a signal-to-noise ratio (SNR) of about 5–10 dB higher than normal hearing listeners to achieve the same level of speech understanding. Therefore, several single- and multimicrophone noise reduction strategies have been developed for modern hearing aids. Multimicrophone noise reduction systems are able to exploit spatial in addition to spectral information and are hence preferred to single-microphone systems (Welker *et al.*, 1997; Lotter, 2004). However, the noise reduction systems currently implemented in modern hearing aids, typically adaptive directional noise reduction systems, are designed to optimize speech in noise monaurally (Wouters and Vanden Berghe, 2001; Luo *et al.*, 2002; Maj *et al.*, 2004). In a bilateral hearing aid configuration, these systems do not take the contralateral ear into account and hence may incorrectly represent the binaural cues (Keidser *et al.*, 2006; Van den Bogaert *et al.*, 2006, 2008). The main binaural cues are interaural time differences (ITDs) and interaural level differences (ILDs). These cues play a major role in directional hearing in the horizontal plane and in spatial awareness and also contribute to an improved speech understanding in a noisy environment due to spatial release from masking also known as "the cocktail party effect" (Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1988; Bronkhorst, 2000). By combining the microphones of the left and right hearing aids into one binaural hearing aid configuration, adaptive algorithms may be controlled more easily to preserve binaural cues, thereby enhancing directional hearing and speech perception in noisy environments. Moreover, an additional improvement in speech perception may be obtained by an increased noise reduction performance due to the advanced signal processing on an increased number of available microphone signals. Spectral cues, more related to resolving front-back confusions and elevation, are not discussed in this manuscript.

Several algorithms have been studied in the past decennium to combine noise reduction with the preservation of binaural localization cues. First, Wittkop and Hohmann (2003) proposed a method based on computational auditory scene analysis in which the input signal is split into different frequency bands. By comparing the estimated binaural properties, such as the coherence, of each frequency band with the expected properties of the signal component (typically it is assumed that the signal component arrives from the frontal area with ITD and ILD values close to 0 μs and 0 dB), these frequencies are either enhanced or attenuated. By applying identical gains to the left and the right hearing aid, binaural cues should be preserved. However, spectral enhancement

artifacts such as "musical noise" will typically occur. Moreover, localization performance when using this technique was never evaluated.

A second class of systems is based on fixed or adaptive beamforming. Desloge *et al.* (1997) introduced two methods to combine fixed beamforming strategies with the preservation of localizing abilities in a binaural hearing aid. In the first method, the amount of ITD distortion introduced by the fixed beamformer, averaged over all directions, was constrained to 40 μs. In the second method, the fixed directional beamformer was limited to frequencies higher than 800 Hz. The monaural output of the beamformer was then combined with the unprocessed low ($f<800$ Hz) frequencies of the omnidirectional microphone at each hearing aid. These frequencies could then be used to localize sound sources. Both of these methods are inspired by observations that the ITD information, present at the lower frequencies, is a dominant localization cue compared to the ILD information, present at the higher frequencies (Wightman and Kistler, 1992). With both systems, a reasonable localization performance was obtained with a root mean square error smaller than 20°. The high pass–low pass method was expanded by Welker *et al.* (1997). An adaptive beamformer was used to process the high-frequency part of the signal. When using this approach with hearing impaired subjects, Zurek and Greenberg (2000) obtained a noise reduction performance of 2.0 dB. However, these systems usually rely on the assumption that the speech signal is arriving from the frontal hemisphere and that the noise signal is arriving from the back hemisphere. Therefore, a good noise reduction performance is only obtained in these specific scenarios. Moreover, localization cues are typically preserved for the targeted speech component but not for the noise component, and this only when speech is arriving from the forward field of view (Van den Bogaert *et al.*, 2008).

A third class of systems is based on the multichannel Wiener filter (MWF). In general, the goal of the Wiener filter is to filter out noise corrupting a desired signal. By using the second-order statistical properties of the desired speech signal and the noise, the optimal filter or Wiener filter can be calculated. It generates an output signal that estimates the desired signal in a minimum mean square error sense. In contrast with a standard beamformer, it can do so without any prior assumption on the angle of arrival of the signal. In Doclo and Moonen (2002), it was shown that a MWF can be used for monaural hearing aid applications. Later on, this approach was extended to a binaural hearing aid configuration in which one or more contralateral microphone signals can be added. One of the main benefits of a MWF is that it inherently preserves the interaural cues of the estimated speech component. This was mathematically proven in the work of Doclo *et al.* (2006). However, it was also proven that the interaural cues of the noise component are distorted into those of the speech component. To preserve binaural information for both the speech and the noise component, an extension, the MWF with partial noise estimation (MWF-N), was proposed by Klasen *et al.* (2007). The rationale of the MWF-N is to remove only part of the noise component. The remaining unprocessed part of the noise signal then restores the spatial cues of the noise component in the signal at the output of the algorithm. Obviously this may come at the cost of a reduced noise reduction. In a way, this is similar to the work of Noble *et al.* (1998) and Byrne *et al.* (1998), in which improvements in localization performance were found when using open instead of closed earmolds. The open earmolds enables the use of the direct unprocessed sound at frequencies with low hearing loss to improve localization performance.

In Van den Bogaert *et al.* (2008) it was shown perceptually that in a binaural hearing aid configuration, the MWF and the MWF-N, have advantages in terms of spatial awareness for the hearing aid user in comparison with an adaptive directional microphone (ADM), which is the most frequently implemented adaptive multimicrophone noise reduction system in modern digital hearing aids. This was done by using a localization experiment in the frontal horizontal hemisphere with a realistic environment ($T_{60}=0.61$ s). In contrast with the ADM, the MWF preserves the location of the target speech sound, independently of its angle of arrival. However, in some conditions subjects located the noise source at the place of the speech source as mathematically predicted by the work of Doclo *et al.* (2006). When using the MWF-N, however, subjects correctly localized both the speech and the noise source.

Until now, noise reduction and speech enhancement performance of the MWF and MWF-N have not been evaluated thoroughly. The study of Klasen *et al.* (2007) focused on the concept of partial noise estimation and how it can decrease ITD and ILD errors. Only a limited set of objective measurements of monaural SNR improvements, done in anechoic conditions with a single noise source fixed at 90°, was reported. The study of Van den Bogaert *et al.* (2008) mainly focused on localization performance. A limited set of speech perception data of three-microphone MWF and MWF-N was also presented. This was done for two single noise source scenarios in a realistic environment. In both scenarios—S_0N_{60} and $S_{90}N_{270}$ with S_xN_y defining the spatial scenario with speech arriving from angle x and a noise signal arriving from angle y—the MWF and MWF-N outperformed a two-microphone ADM. The MWF and MWF-N can increase noise reduction performance by using microphone signals from both the ipsilateral and the contralateral hearing aid. However, transmitting microphone signals between hearing aids comes at the large cost of power consumption and bandwidth, especially since commercial manufacturers prefer a wireless connection between both devices. Therefore, a thorough evaluation in realistic listening conditions is needed on the obtained gain in speech understanding when transmitting no (a bilateral configuration), one, or all contralateral microphone signals. A commonly used ADM is used as a reference noise reduction system. The algorithms discussed in this manuscript are evaluated using monaural and binaural presentations.

This paper presents objective and perceptual evaluations of the noise reduction and speech enhancement performance of the MWF and MWF-N approaches using different microphone combinations under several spatial sound scenarios in different acoustical environments. The main research questions answered in this manuscript are the following: (a) What

is the speech enhancement performance of a MWF in comparison with a standard bilateral ADM in a monaural and a binaural hearing aid configuration? (b) What is the gain in speech enhancement when evolving from a monaural hearing aid design to a binaural hearing aid design, i.e., adding a third and/or a fourth microphone, positioned at the contralateral hearing aid, to a MWF already using two microphones of the ipsilateral hearing aid? (c) What is the cost in speech enhancement performance when adding a partial noise estimate into the MWF-scheme, i.e., the MWF-N, which enables a correct sound localization of both the speech and the noise component? (Van den Bogaert *et al.*, 2008). All three questions will be evaluated using both objective performance measures, using a semianechoic and a realistic reverberant environment, and perceptual performance measures, only for the realistic environment, in different single and multiple noise source scenarios. The correlation between both performance measures is also discussed.

II. HEARING AID CONFIGURATION

The hearing aid configuration used in this study is identical to the one used in Van den Bogaert *et al.* (2008). The microphone array of the left and right behind-the-ear hearing aids consists of two omnidirectional microphones with an intermicrophone distance of approximately 1 cm. In a general binaural configuration, microphone signals from the ipsilateral (M_I) and contralateral (M_C) hearing aids can be used to generate an output signal for each ear. Three different noise reduction algorithms were evaluated with these hearing aids: the MWF, the MWF-N, and the ADM. For all algorithms a sampling frequency of f_s=20 480 Hz was used.

A. MWF and MWF-N

Different microphone combinations were evaluated to measure the benefit of adding one or two contralateral microphone signals to the MWF or MWF-N algorithm active at the ipsilateral hearing aid. A monaural system with each hearing aid using only its own two microphone signals was first evaluated. The MWF-based systems were then extended by transmitting one or two contralateral microphone signals to the ipsilateral hearing aid. The three different implementations of the MWF algorithm used in this study are denoted as MWF_{2+M_C}, with $0 \leqslant M_C \leqslant 2$. The three different implementations of the MWF-N algorithm are denoted similarly as MWF_{2+M_C}-N. A list of algorithms evaluated during this study is given in the left column of Table I. A description of the algorithmic aspects of the MWF and MWF-N algorithms is already presented in Van den Bogaert *et al.* (2008). A brief summary is given here. The algorithms are described in the frequency domain.

Transmitting M_C contralateral microphone signals to the ipsilateral hearing aid results in an M-dimensional ($M=M_I+M_C$) input vector $\mathbf{Y}_L(\omega)$ and $\mathbf{Y}_R(\omega)$ for the left and right hearing aid, respectively. Each signal vector $\mathbf{Y}(\omega)$ can be written as a sum of a speech component $\mathbf{X}(\omega)$ and a noise component $\mathbf{V}(\omega)$, which are equal to the speech and noise source signals convolved with the impulse responses of the room. The output signal of the noise reduction algorithm at the left and the right hearing aid can be described by the filtered input vectors, i.e.,

$$Z_L(\omega)=\mathbf{W}_L^H(\omega)\mathbf{Y}_L(\omega), \quad Z_R(\omega)=\mathbf{W}_R^H(\omega)\mathbf{Y}_R(\omega), \tag{1}$$

where $\mathbf{W}_L(\omega)$ and $\mathbf{W}_R(\omega)$ are M-dimensional complex vectors representing the calculated Wiener filters for each hearing aid. The MWF uses the M available microphone signals at each hearing aid to produce the filters $\mathbf{W}_L(\omega)$ and $\mathbf{W}_R(\omega)$. These filters create a minimum mean square error estimate of the speech component at the reference microphone, usually the front omnidirectional microphone for the left [for $\mathbf{W}_L(\omega)$] and for the right [for $\mathbf{W}_R(\omega)$] hearing aid, respectively. By doing so, an MWF inherently preserves the binaural cues of the speech component. Through the remainder of the paper, the frequency domain variable ω is omitted for conciseness.

The filter $\mathbf{W}=[\mathbf{W}_L^T\mathbf{W}_R^T]^T$ with T the transpose operator, is calculated by minimizing the cost function

$$J_{\mathrm{MWF}}(\mathbf{W})=\mathcal{E}\left\{\left\|\begin{bmatrix} X_{L,1}-\mathbf{W}_L^H\mathbf{X}_L \\ X_{R,1}-\mathbf{W}_R^H\mathbf{X}_R \end{bmatrix}\right\|^2+\mu\left\|\begin{bmatrix} \mathbf{W}_L^H\mathbf{V}_L \\ \mathbf{W}_R^H\mathbf{V}_R \end{bmatrix}\right\|^2\right\}, \tag{2}$$

with H the Hermitian transpose operator and $\mathcal{E}$ the expected value operator. μ is a parameter which trade offs noise reduction performance and speech distortion (Spriet *et al.*, 2004). The rationale of the MWF-N is to remove not the full noise component from the reference microphone signal but to remove only a part $(1-\eta)$ of it. The other part (η) remains unprocessed. This changes the original cost function to

$$J_{\mathrm{MWF\text{-}N}}(\mathbf{W})=\mathcal{E}\left\{\left\|\begin{bmatrix} X_{L,1}-\mathbf{W}_L^H\mathbf{X}_L \\ X_{R,1}-\mathbf{W}_R^H\mathbf{X}_R \end{bmatrix}\right\|^2+\mu\left\|\begin{bmatrix} \eta V_{L,1}-\mathbf{W}_L^H\mathbf{V}_L \\ \eta V_{R,1}-\mathbf{W}_R^H\mathbf{V}_R \end{bmatrix}\right\|^2\right\}. \tag{3}$$

TABLE I. The list of algorithms, microphone combinations, and spatial scenarios evaluated in this paper. All algorithms use two microphone signals of the left/right hearing aid and M_C (0,1 or 2) microphone signals of the contralateral hearing aid to generate a signal for the left/right ear. This is depicted as $2+M_C$. The second column represents whether a bilateral/binaural (b) and/or a monaural presentation (m) was used during the perceptual evaluation of the corresponding algorithm. The third and fourth columns represent the list of spatial scenarios, S_xN_y, evaluated during the objective evaluations. x represents the location of the speech source; y represents the location of the noise source(s). The conditions S_0N_{60}, $S_{90}N_{270}$, and S_0N_3 were also evaluated perceptually.

Evaluated algorithms		Spatial scenarios	
MWF_{2+0}	$b+m$	S_0N_x	x between 0° and 330°
MWF_{2+1}	b	$S_{90}N_{180}$	Single noise source N at 180°
MWF_{2+2}	b	$S_{90}N_{270}$	Single noise source N at 270° (=−90°)
MWF_{2+0}-$\mathrm{N}_{0.2}$	b	$S_{45}N_{315}$	Single noise source N at 315° (=−45°)
MWF_{2+1}-$\mathrm{N}_{0.2}$	b	S_0N_{2a}	Noise sources at −60° and +60°
MWF_{2+2}-$\mathrm{N}_{0.2}$	b	S_0N_{2b}	Noise sources at −120° and +120°
ADM	$b+m$	S_0N_{2c}	Noise sources at 120° and 210°
Unproc	$b+m$	S_0N_3	Noise sources at 90°, 180° and 270°
		S_0N_{4a}	Noise sources at 60°, 120°, 180° and 210°
		S_0N_{4b}	Noise sources at 60°, 120°, 180° and 270°

Note that Eq. (2) is a special case of Eq. (3) with $\eta=0$. Both cost functions are minimized by using estimates of the speech and noise correlation matrices (Klasen *et al.*, 2007; Van den Bogaert *et al.*, 2008). The Wiener solution minimizing $J_{\text{MWF-N}}(\mathbf{W})$ equals

$$\mathbf{W}=\begin{bmatrix}\mathbf{R}_{x,L}+\mu\mathbf{R}_{v,L} & 0_M\\ 0_M & \mathbf{R}_{x,R}+\mu\mathbf{R}_{v,R}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{R}_{x,L}\mathbf{e}_L\\ \mathbf{R}_{x,R}\mathbf{e}_R\end{bmatrix}, \quad (4)$$

where $\mathbf{e}_L$ and $\mathbf{e}_R$ are all zero vectors, except for a "1" in the position corresponding to the selected reference microphone, i.e., $\mathbf{e}_L(1)=1$ and $\mathbf{e}_R(1)=1$. $\mathbf{R}_x$, and $\mathbf{R}_v$, are defined as the $(M\times M)$-dimensional speech and noise correlation matrices, e.g., for the left hearing aid $\mathbf{R}_{x,L}=\mathcal{E}\{\mathbf{X}_L\mathbf{X}_L^H\}$ and $\mathbf{R}_{v,L}=\mathcal{E}\{\mathbf{V}_L\mathbf{V}_L^H\}$. A voice activity detector (VAD) is used to discriminate between "speech and noise periods" and "noise only periods." The noise correlation matrix $\mathbf{R}_v$ was calculated during the noise only periods. The speech correlation matrix $\mathbf{R}_x$ was estimated by subtracting $\mathbf{R}_v$ from the correlation matrix $\mathbf{R}_y$ of the "speech and noise" signal vector $\mathbf{Y}$. For both the MWF and MWF-N algorithms, speech and noise and noise only correlation matrices were calculated using a perfect VAD. A filter length of 96 taps was used per microphone channel. When using block processing, an overlap of 48 samples was used, leading to a total delay of approximately 4.7 ms for the MWF and MWF-N algorithms. Pilot experiments showed that $\mu=5$ provides a good trade-off between noise reduction and speech distortion. In the work of Van den Bogaert *et al.* (2008), it was shown that $\eta=0.2$ resulted in a good localization performance. Therefore these parameter settings were used throughout this study.

B. Adaptive directional microphone

An ADM was used as a reference multimicrophone noise reduction algorithm. This algorithm is commonly used in modern digital hearing aids (Luo *et al.*, 2002; Maj *et al.*, 2004). Unlike the MWF-based algorithms, the ADM relies on the assumption that the target signal arrives from the frontal field of view and that jammer signals arrive from the back hemisphere. The ADM exploits the time of arrival differences between the microphones on a hearing aid to improve the SNR by steering a null in the direction of the jammer signals. The ADM used the two omnidirectional microphones of the ipsilateral hearing aid. A first stage generated two software directional microphone signals corresponding to, respectively, a front and a back oriented cardioid pattern. These signals were then combined by an adaptive scalar β to minimize the energy arriving from the back hemisphere at the output of the algorithm (Maj *et al.*, 2006). The parameter β was constrained between 0 and 0.5 to avoid noise reduction in the frontal hemisphere.

III. METHODS

A. General

First, different sets of impulse responses were measured between a loudspeaker and the microphones in two behind-the-ear hearing aids worn by a CORTEX MK2 manikin. Loudspeakers were placed at 1 m distance of the center of the head, and impulse responses were measured in the horizontal plane in steps of 30°. Measurements were done in a room with dimensions $5.50\times4.50\times3.10$ m^3 (length $\times$ width $\times$ height), and acoustical curtains were used to change its acoustical properties. Two different acoustical environments were studied with a reverberation time, linearly averaged over all one-third octave bands between 100 and 8000 Hz, of, respectively, $T_{60}=0.21$ s and $T_{60}=0.61$ s, with the latter value corresponding to a realistic living room condition.

The measured impulse responses were convolved with the appropriate speech and noise material to generate the four microphone signals for the different spatial scenarios used in the perceptual and the objective evaluations. A spatial scenario, with a target signal (S) arriving from angle x and one or multiple noise sources (N) arriving from angle(s) y, is denoted as S_xN_y. The angles were defined clockwise with 0° being in front of the subject. The generated microphone signals were used as input for the different algorithms. Besides the different algorithms, an unprocessed condition, using the front microphones of each hearing aid, was used as a reference condition. In each spatial scenario, the input SNR was calibrated to 0 dBA, measured in absence of the head. A full list of tested conditions is given in Table I. Evaluations were done after convergence of the filters for all algorithms.

B. Objective evaluation

The improvement in speech intelligibility weighted SNR ($\Delta\text{SNR}_{\text{SI}}$), defined by Greenberg *et al.* (1993), was used to evaluate the noise reduction performance of the algorithms. This is defined as the difference between the output SNR_{SI} and the input SNR_{SI}. The input SNR_{SI} was calculated between the front omnidirectional microphone of the left and the right hearing aid. For the left hearing aid, this gives

$$\Delta\text{SNR}_{\text{SI},L}=\sum_i I(\omega_i)\text{SNR}_{\text{out},L}(\omega_i)-I(\omega_i)\text{SNR}_{\text{in},L}(\omega_i), \quad (5)$$

with $\text{SNR}(\omega_i)$ as the SNR measured in the ith third-octave band and $I(\omega_i)$ as the importance of the ith frequency band for speech intelligibility, as defined by ANSI-SII (1997).

Noise reduction performance was evaluated using an average speech spectrum of a Dutch male speaker from the VU test material (Versfeld *et al.*, 2000) as target sound (S) and multitalker babble (Auditec of St. Louis) as jammer sound (N). The long term average spectrum of both the speech and the noise material is given in Van den Bogaert *et al.* (2008). For multiple noise source scenarios, time-shifted versions of the same noise source signal were generated to obtain "uncorrelated" noise sources. Since simulations are a time-efficient way to assess the performance of noise reduction algorithms, a large number of spatial conditions were examined using one target signal and one to four noise sources. A full list of studied spatial scenarios is given in the right column of Table I. Simulations were done for both $T_{60}=0.21$ s and $T_{60}=0.61$ s to evaluate the influence of reverberation on the algorithms.

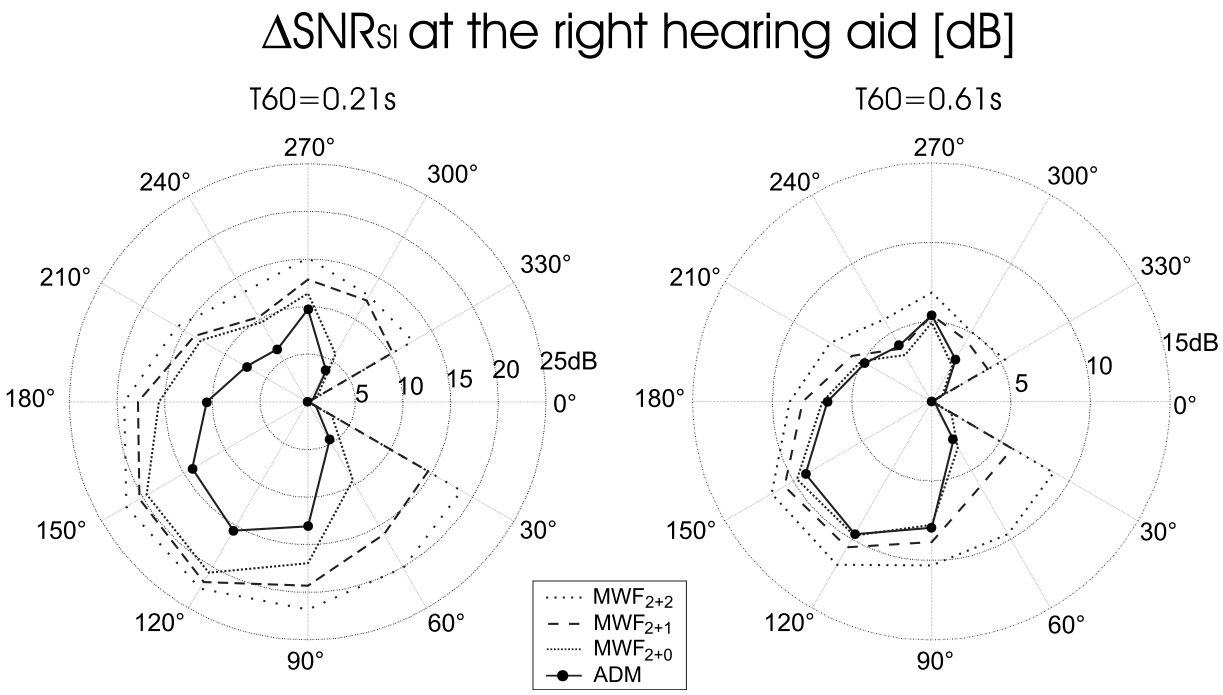


FIG. 1. ΔSNR_{SI} of the ADM and the MWF with different microphone combinations (denoted as MWF_{2+M_C}) for single noise source scenarios with speech arriving from 0° and noise arriving from $x°$ (S_0N_x). The data of the right hearing aid are presented in two reverberant environments, with x being varied per 30°.

C. Perceptual evaluation

Speech reception thresholds (SRTs) were measured with ten normal hearing subjects using an adaptive test procedure (Plomp and Mimpen, 1979). The procedure adjusts the level of the speech signal in steps of 2 dB to extract the 50% SRT. The level of the noise signal was calibrated with the sound pressure level, averaged over the left and the right ear, equal to 65 dBA. The male sentences of the VU test material (Versfeld *et al.*, 2000) were used as speech material, and a multitalker babble (Auditec of St. Louis) was used as a noise source.

The algorithms were perceptually evaluated using a binaural presentation with signals presented to both the left and the right ear. The MWF_{2+0} and the ADM were also tested for one ear only with a monaural presentation of the stimuli (for the full list of conditions, see Table I). In the monaural evaluation, signals were presented to the right ear of the subjects. In both the binaural and the monaural presentation, an unprocessed condition was used as a reference, bringing the total of tested conditions to 11. The speech enhancement achieved by each algorithm was calculated by subtracting the SRT score (in dB SNR) of the algorithm from the unprocessed SRT score, i.e.,

$$\Delta SRT_{algo} = SRT_{unproc} - SRT_{algo}. \quad (6)$$

Tests were performed in a double walled sound booth under headphones (TDH-39) using an RME Hamerfall Multiface II soundcard and a Tucker Davis HB7 headphone driver. The perceptual evaluations were carried out using the impulse responses of the acoustical environment with T_{60} =0.61 s, i.e., a realistic living room condition. Because of practical considerations, three spatial scenarios, selected from the list of scenarios tested in the objective evaluation, were perceptually evaluated, i.e., S_0N_{60}, $S_{90}N_{270}$ and a triple noise source condition $S_0N_{90/180/270}$.

IV. RESULTS AND ANALYSIS

A. Objective evaluation

First, the noise reduction performance of the MWF is discussed and compared with the ADM. Second, the MWF-N is evaluated.

1. MWF

Figure 1 shows the measured speech intelligibility weighted gain in SNR, ΔSNR_{SI}, for a target speech source arriving from 0° and a single noise source arriving from $x°$ (S_0N_x) for the ADM and the three different MWF algorithms. This is done for a room with a low (T_{60}=0.21 s) and a living room (T_{60}=0.61 s) reverberation time, respectively. The data are given only for the right hearing aid as, for a single noise source scenario, the directivities of the left and the right hearing aid are almost identical (if one changes positive angles into negative angles). The noise reduction data of more challenging scenarios, with multiple noise sources or a nonzero speech source angle, are shown in Fig. 2.

For both the single noise source data and the more complex spatial scenarios, it was observed that the acoustical parameters have a very large effect on the noise reduction performance of the algorithms. Due to the presence of reflections, the performance of all algorithms decreased significantly, which is a well known effect from literature. In case of a low reverberant condition, gains of up to 23 dB were obtained. In a more realistic environment, this performance

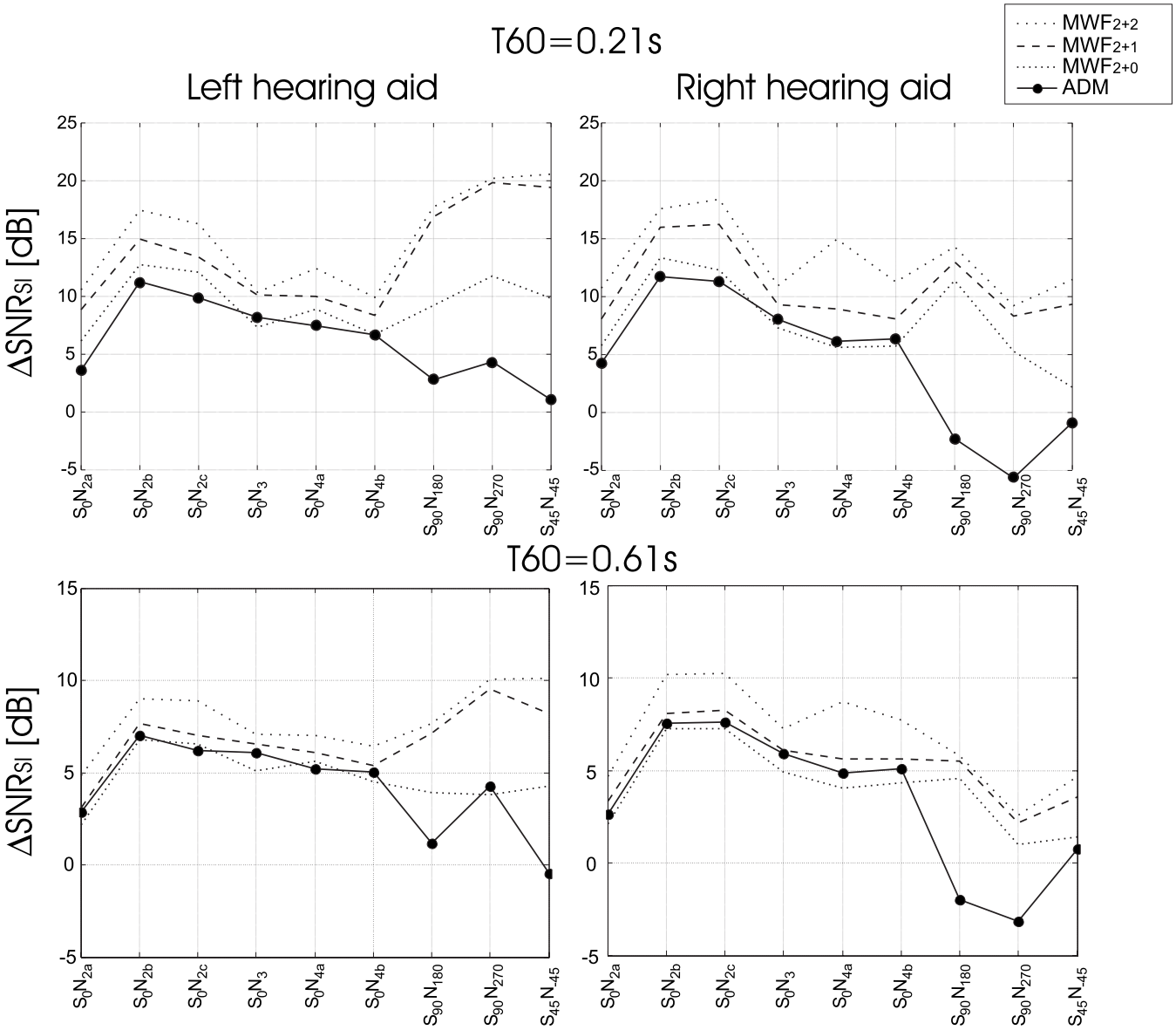


FIG. 2. ΔSNR_{SI} of the MWF with different microphone combinations and the ADM for multiple noise sources. The abbreviations of the spatial scenarios are explained in Table I. Two different acoustical environments are evaluated.

dropped to 12 dB for the same spatial scenario and the same hearing aid, i.e., the scenario S_0N_{120} at the right hearing aid.

In single noise source scenarios (Fig. 1), extending the MWF_{2+0} with contralateral microphone signals substantially increased noise reduction performance, especially if the speech and the noise source were positioned within 60° of each other. In these spatial scenarios, an additional gain of 7.5–14 dB in T_{60}=0.21 s and of 3.1–7.6 dB in T_{60}=0.61 s was obtained for the right hearing aid when going from the MWF_{2+0} to the MWF_{2+2}. In the other single noise source scenarios, the benefit was much more modest. An average difference (and standard deviation) between the MWF_{2+0} and, respectively, the MWF_{2+1} and the MWF_{2+2} of 1.4 ± 0.7 and 3.3 ± 1.0 dB for T_{60}=0.21 s and of 0.8 ± 0.3 and 2.2 ± 0.3 dB for T_{60}=0.61 s was measured over these spatial scenarios. Interestingly the MWF_{2+0} outperformed the ADM in low reverberant conditions. However, in a realistic environment both bilateral algorithms had a similar performance.

For the multiple noise source scenarios, as shown in Fig. 2, the same trends were observed, with the MWF_{2+2} outperforming the MWF_{2+1}, which in turn performed better than the MWF_{2+0} and ADM. For both acoustic environments, both two-microphone algorithms, i.e., the ADM and the MWF_{2+0}, tend to have a similar performance. However, for the spatial scenarios with the target signal not arriving from 0°, all MWF-based algorithms easily outperformed the ADM. In these scenarios, the ADM only showed very small improvements or even a decrease in ΔSNR_{SI} (up to −5 and −2.5 dB for T_{60}=0.21 s and T_{60}=0.61 s, respectively). For the more complex spatial scenarios shown in Fig. 2, it is observed that the gain in noise reduction achieved by extending the MWF_{2+0} with contralateral microphone signals was highly dependent on the spatial scenario and the ear of interest. For instance, a large gain in ΔSNR_{SI} for the left hearing aid is observed in $S_{90}N_{270}$, while a more modest gain is present at the right hearing aid. For the right hearing aid, a large gain is observed for, e.g., condition S_0N_{4a}, while a more modest gain is observed in, e.g., condition S_0N_3.

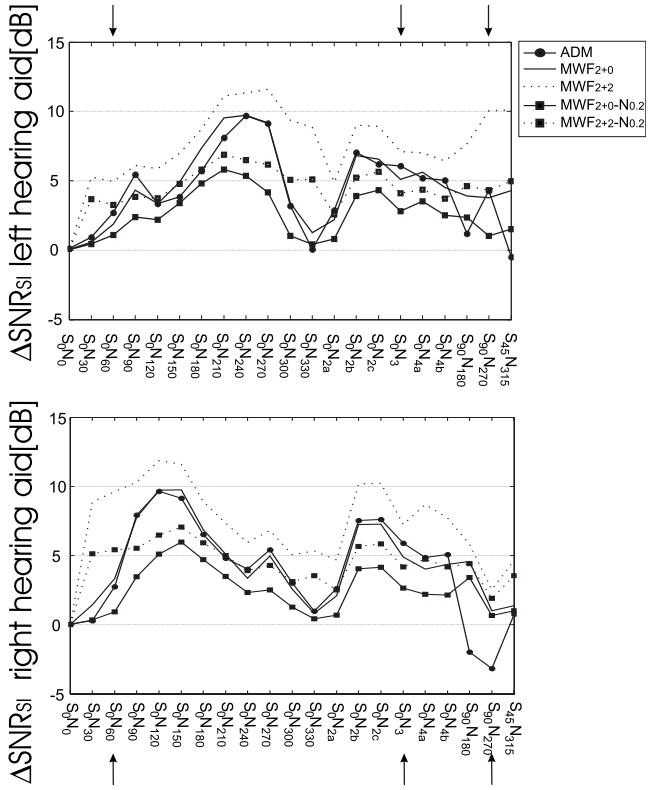


FIG. 3. The influence of η=0.2 on ΔSNR_{SI} of the MWF, the MWF-$N_{0.2}$, and the ADM for T_{60}=0.61 s. A four- and two-microphone MWF-based system have been tested. The abbreviations of the spatial scenarios are explained in Table I. The arrows highlight the spatial scenarios that have been evaluated perceptually.

2. MWF-N

As discussed in the Introduction, the MWF-N enables the user to correctly localize the speech and the noise component when used in a binaural hearing aid configuration. This is in contrast with other signal processing schemes for hearing aids, e.g., the ADM and partly (only for the noise component) the MWF (Van den Bogaert *et al.*, 2008). The parameter η controls the amount of noise that remains unprocessed by the algorithm.

Figure 3 illustrates the influence of the parameter η =0.2 on the estimated noise reduction performance of the MWF_{2+2} and the MWF_{2+0}. The performance of the ADM is also shown as a reference noise reduction system. This figure illustrates that when adding a partial noise estimate to the MWF algorithm (MWF-Nη), the loss in noise reduction is not only dependent on the parameter η, but also on the amount of noise reduction originally obtained by the MWF. Larger losses are observed if a high noise reduction performance was already obtained by the MWF algorithm. As a consequence, the influence of the parameter η is more pronounced on the MWF_{2+2} than on the MWF_{2+0} algorithm. The figure shows that when using η=0.2, the estimated noise reduction performance of the MWF_{2+2}-$N_{0.2}$ drops, in most conditions, below the performance of the ADM and the MWF_{2+0}. Other simulations have shown that when using η =0.1, the MWF-N still outperforms the ADM. If the speech source is located outside the forward field of view, all MWF- and MWF-N-based algorithms outperform the ADM.

B. Perceptual evaluation

To further validate the performance of the MWF and MWF-N, a number of perceptual evaluations were performed. Three spatial scenarios were selected (see Table I or the arrows in Fig. 3). Table II shows the improvement in SRT relative to an unprocessed condition averaged over ten normal hearing subjects obtained when using the different algorithms. The bottom two rows show the SRT levels of the unprocessed reference condition. The gains in ΔSNR_{SI} measured during the objective evaluation were added for both the left and the right hearing aid. All statistical analyses were done using SPSS 15.0. For conciseness, the term "factorial repeated measures analysis of variance (ANOVA)" is abbre-

TABLE II. The gain in SRT, ΔSRT_{algo}, averaged over ten normal hearing subjects. The bottom rows show the SNRs at which the unprocessed reference SRTs have been measured for the monaural and the binaural presentations. A "*" depicts a significant noise reduction performance ($p<0.05$) compared to the unprocessed condition. ΔSNR_{SI}, calculated for the left and right hearing aids in the objective evaluation, is also added to the table.

	S_0N_{60}			$S_{90}N_{270}$			$S_0N_{90/180/270}$		
Bilat/bin ΔSRT (dB)	Perceptual	Left	Right	Perceptual	Left	Right	Perceptual	Left	Right
ADM	2.1 ± 1.9	2.7	2.8	−4.3 ± 1.3*	4.3	−3.2	1.3 ± 1.4	6.0	5.9
MWF_{2+2}	4.3 ± 1.5*	4.9	9.6	0.7 ± 1.4	10.0	2.5	4.6 ± 0.8*	7.1	7.2
MWF_{2+1}	3.8 ± 1.6*	4.0	6.2	0.3 ± 2.0	9.6	2.1	4.0 ± 1.5*	6.6	6.0
MWF_{2+0}	1.0 ± 0.7*	1.9	3.3	−1.2 ± 1.6	3.8	1.0	2.8 ± 1.3*	5.1	4.9
MWF_{2+2}-$N_{0.2}$	3.6 ± 1.4*	3.3	5.4	2.0 ± 1.4*	4.3	1.9	3.2 ± 0.8*	4.1	4.2
MWF_{2+1}-$N_{0.2}$	2.7 ± 1.3*	2.6	3.0	1.5 ± 1.6	3.9	1.6	3.4 ± 0.8*	3.7	3.3
MWF_{2+0}-$N_{0.2}$	1.0 ± 2.1	1.1	0.9	0.0 ± 1.5	1.0	0.7	2.3 ± 1.4*	2.8	2.6
Monaural ΔSRT (dB)									
ADM	5.4 ± 2.0*		2.8	−5.4 ± 1.2*		−3.2	3.4 ± 2.3*		5.9
MWF_{2+0}	3.4 ± 1.3*		3.3	−0.7 ± 1.4		1.0	5.0 ± 1.6*		4.9
SNR-unproc (dB)									
Binaural	−6.2 ± 1.8			−9.1 ± 1.7			−7.2 ± 1.6		
Monaural	2.8 ± 2.0			−8.0 ± 1.7			−3.0 ± 2.1		

viated as ANOVA, and pairwise comparisons discussed throughout the document were always Bonferroni corrected for multiple comparisons. The reported p-values of the pairwise comparisons are lower bound values. A p-value of $p=0.05$ was used as a threshold for significance.

1. Bilateral/binaural presentation

To compare the different algorithms, an ANOVA is carried out on the SRT data. These data were also used to calculate the average gains shown in Table II [see Eq. (6)]. The ANOVA was carried out using the factor algorithm (seven algorithms and an unprocessed condition) and spatial scenario (three spatial scenarios). An interaction was found between both factors ($p=0.005$). This was expected since the performance of the algorithms was clearly dependent on the location of the speech and the noise source(s). Therefore an ANOVA and pairwise comparisons were carried out for each spatial scenario. For all three spatial scenarios, a main effect for the factor algorithm was found ($p=0.002$, $p<0.001$, and $p<0.001$ for respectively, S_0N_{60}, $S_{90}N_{270}$, and $S_0N_{90/180/270}$).

First, an overview is given of the comparisons made between the algorithms and the unprocessed condition. An " *" was added in Table II if the algorithm generated a significant gain in SRT compared to the unprocessed condition. For the scenario S_0N_{60}, a significant gain in noise reduction was achieved by all algorithms except for the ADM ($p=0.155$) and the MWF_{2+0}-$\mathrm{N}_{0.2}$ ($p=1.000$). The highest significant gain was obtained by the MWF_{2+2} algorithm (4.3 dB, $p<0.001$). The lowest significant gain was obtained when using the MWF_{2+0} (1.0 dB, $p=0.036$). For the scenario $S_{90}N_{270}$, a significant gain was achieved only by the MWF_{2+2}-$\mathrm{N}_{0.2}$ algorithm (2.0 dB, $p=0.047$). When using the ADM, a significant decrease in speech understanding was observed (−4.3 dB, $p<0.001$). For the triple noise source scenario, all MWF algorithms showed a significant gain in speech understanding ranging from 2.3 dB for the MWF_{2+0}-$\mathrm{N}_{0.2}$ ($p=0.019$) to 4.6 dB for the MWF_{2+2} ($p<0.001$). The ADM showed no significant improvement compared to the unprocessed condition ($p=0.435$).

Second, an overview is given of the pairwise comparisons between the ADM and all MWF and MWF-N approaches. For the spatial scenario S_0N_{60}, only the MWF_{2+2} showed a significant gain in speech enhancement compared to the ADM (2.2 dB, $p=0.013$); the MWF_{2+1} showed a nonsignificant gain of 1.6 dB ($p=0.061$). The performance of the MWF_{2+0} showed no significant difference with the ADM (which is also a two microphone algorithm). For the scenario $S_{90}N_{270}$ all MWF and MWF-N algorithms showed a clear significant benefit (all p-values $p\leqslant 0.001$) compared to the ADM. This benefit is in the range of 3.1 dB for the MWF_{2+0} to 6.3 dB for the MWF_{2+2}-$\mathrm{N}_{0.2}$. For the triple noise source scenario, a significant benefit is found for the MWF_{2+2} (3.3 dB, $p<0.001$), the MWF_{2+1} (2.7 dB, $p<0.001$), and the MWF_{2+1}-$\mathrm{N}_{0.2}$ (2.1 dB, $p=0.002$). Since the MWF_{2+1}-$\mathrm{N}_{0.2}$ showed a significant gain compared to the ADM, it was expected that also the MWF_{2+2}-$\mathrm{N}_{0.2}$, which has an extra microphone input, would show this benefit. However, no statistically significant difference is found between this algorithm and the ADM (1.9 dB, $p=0.164$).

Third, the influence of adding contralateral microphones to the original two-microphone MWF-scheme (MWF_{2+0}) can be observed. For S_0N_{60} both the MWF_{2+1} and MWF_{2+2} showed a significant increase in performance of, respectively, 2.8 dB ($p=0.022$) and 3.3 dB ($p=0.001$) compared to the MWF_{2+0}. The MWF_{2+2} and MWF_{2+1} were statistically not significantly different. For the MWF-$\mathrm{N}_{0.2}$ algorithms the same trends were observed, but these differences were not statistically significant [MWF_{2+1}-$\mathrm{N}_{0.2}$ and MWF_{2+2}-$\mathrm{N}_{0.2}$ show an average improvement of, respectively, 1.7 dB ($p=0.341$) and 2.5 dB ($p=0.125$) compared to the MWF_{2+0}-$\mathrm{N}_{0.2}$]. For the spatial scenario $S_{90}N_{270}$, the same observations are made, with MWF_{2+1} and MWF_{2+2} performing statistically better than MWF_{2+0} (respectively, 1.5 dB, $p=0.033$ and 1.8 dB, $p=0.001$) and with no significant difference between MWF_{2+2} and MWF_{2+1}. Again both the MWF_{2+1}-$\mathrm{N}_{0.2}$ and MWF_{2+2}-$\mathrm{N}_{0.2}$ show the same nonsignificant trend compared to the MWF_{2+0}-$\mathrm{N}_{0.2}$ (with, respectively, a gain of 1.5 dB, $p=0.454$ and 1.9 dB, $p=0.215$). For the triple noise source scenario, only the MWF_{2+2} performed significantly better than the MWF_{2+0} (1.7 dB, $p=0.004$). Again both the MWF_{2+1}-$\mathrm{N}_{0.2}$ and MWF_{2+2}-$\mathrm{N}_{0.2}$ show a nonsignificant improvement compared to the MWF_{2+0}-$\mathrm{N}_{0.2}$.

Finally the last comparisons examine the impact of introducing the partial noise estimate using $\eta=0.2$ to the original MWF algorithm (MWF versus MWF-$\mathrm{N}_{0.2}$). In the three different ANOVAs, one for each spatial scenario, only one significant difference was found when comparing the performance of the MWF_{2+M_C} with the MWF_{2+M_C}-$\mathrm{N}_{0.2}$, with M_C ranging from 0 to 2. A significant decrease in performance of −1.4 dB is observed ($p=0.016$) when comparing the MWF_{2+2}-$\mathrm{N}_{0.2}$ with the MWF_{2+2} in the triple noise source scenario. Some other nonsignificant trends were also observed. In the triple noise source scenario and in scenario S_0N_{60}, the MWF_{2+M_C}-$\mathrm{N}_{0.2}$ tends to have a decreased performance compared to the MWF_{2+M_C} condition, which was expected since the parameter $\eta=0.2$ introduces an unprocessed noise component at the output of the noise reduction algorithm. Interestingly this trend is not observed in the scenario $S_{90}N_{270}$. In this scenario the MWF-N algorithms typically outperformed the MWF algorithms.

These trends were verified by a different ANOVA. In this refined analysis, the factor algorithm (three different MWF algorithms: MWF_{2+0}, MWF_{2+1}, and MWF_{2+2}) and eta ($\eta=0$ and $\eta=0.2$) were used per spatial condition. For all three ANOVAs, no interactions were found between both factors. For the scenario S_0N_{60}, no significant effect is observed. For the condition $S_{90}N_{270}$, a significant increase in performance of 1.3 dB ($p=0.002$) is observed when comparing $\mathrm{MWF}_{2+M_C}-\mathrm{N}_{0.2}$ with MWF_{2+M_C}. For the triple noise source scenario, a significant decrease in performance of 0.8 dB ($p=0.001$) is observed when introducing $\eta=0.2$. In all three of these ANOVAs, a significant increase in performance is found when introducing one or two contralateral microphones, but no significant difference is observed between the three and four-microphone algorithms, confirming the observations made in the paragraph on contralateral mirophones.

2. Monaural presentation

The monaural SRT data, used to calculate the gains shown in Table II, were used in an ANOVA. Again the factor algorithm (two algorithms and an unprocessed condition) and spatial scenario were used. Similar to the analysis of the bilateral/binaural presentation, an interaction is found between both factors ($p<0.001$). This leads to a separate ANOVA and separate pairwise comparisons for each spatial scenario.

In the scenario S_0N_{60}, both algorithms perform significantly better than the unprocessed condition with an average gain of 3.4 dB by the MWF_{2+0}, $p<0.001$ and an average gain of 5.4 dB by the ADM, $p<0.001$. Both algorithms are significantly different from each other, with the performance of the ADM being 2.0 dB better than the MWF_{2+0} ($p=0.007$). For the scenario $S_{90}N_{270}$, the MWF_{2+0} is not significantly different from the unprocessed condition. The ADM shows a significant decrease in performance compared to both the MWF_{2+0} and the unprocessed condition (respectively, 5.4 and 4.7 dB, both $p<0.001$). In the triple noise source scenario, both the MWF_{2+0} and the ADM show a significant improvement compared to the unprocessed condition (respectively, 5.0 dB, $p<0.001$ and 3.4 dB, $p=0.004$).

3. Comparison with the objective data

In Table II, the noise reduction gains (ΔSNR_{SI}) calculated during the objective evaluations are shown together with the speech enhancement data of the perceptual evaluations. Large correlations are present between the data of both evaluations. In the bilateral/binaural configuration, perceptual results correlated best with ΔSNR_{SI} of the hearing aid that had the best input SNR (e.g., the left ear for S_0N_{60}, the right ear for $S_{90}N_{270}$, and both ears for $S_0N_{90/180/270}$). It was observed that this hearing aid is typically the device with the lowest gain in noise reduction, ΔSNR_{SI}. Although large correlations between both performance measures were observed, Table II illustrates that the performance of the ADM and the MWF seems to be overestimated by approximately 2 dB in $S_{90}N_{270}$ and the triple noise source scenario.

V. DISCUSSION

This paper evaluates two recently introduced MWF-based noise reduction algorithms for multimicrophone hearing aids, which offer the ability to preserve the spatial awareness of hearing aid users. A verification of the speech enhancement and the noise reduction performance of the algorithms is presented in this study. A bilateral ADM was used as a reference noise reduction algorithm as this is commonly implemented in current bilateral hearing aids. Three research questions on combining noise reduction with preserving sound source localization in multimicrophone noise reduction algorithms were raised in the Introduction. The results and analysis from the previous sections will be used to answer these questions.

A. Noise reduction performance of the MWF

In Sec. IV A the performance of the MWF was evaluated objectively in two different acoustical environments, i.e., $T_{60}=0.21$ s and $T_{60}=0.61$ s. In the low reverberant condition, the two-microphone MWF, i.e., the MWF_{2+0}, outperformed the ADM, especially in single noise source scenarios (Fig. 1) and in conditions in which the target signal was not arriving from the forward field of view (the three rightmost data-points of Fig. 2). The performance of all the adaptive algorithms dropped significantly in a more realistic acoustic environment. This phenomenon is well known and commonly found in literature [e.g., see Kompis and Dillier (2001) and Greenberg and Zurek (1992)]. In this more realistic acoustic environment, the MWF_{2+0} outperformed the ADM only if the speech source is not arriving from the forward field of view. In all other spatial scenarios, both two-microphone algorithms had approximately the same performance. The perceptual evaluation, also carried out with $T_{60}=0.61$ s, supported these conclusions. When using a bilateral configuration that consists of two independent monaural systems, no significant differences were apparent between the ADM and the MWF_{2+0} if the speech source arrives from 0° (Sec. IV B). Still, unlike the ADM, the MWF preserves the binaural cues of the speech component independent of the angle of arrival of the signal (Doclo *et al.*, 2006; Van den Bogaert *et al.*, 2008).

Why the ADM caught up with the performance of the MWF in more reverberant conditions can be explained by the MWF, unlike an ADM, not performing any dereverberation. The MWF is designed to estimate the speech component, X, present at a reference microphone, which is the convolution of the target signal S with the room impulse response. Hence, no dereverberation is performed. The ADM, on the other hand, is designed to preserve signals arriving from the frontal hemisphere. In other words, reflections arriving from the back hemisphere are reduced in amplitude. However, this also implies that the ADM will reduce speech perception if the target signal arrives from the side or the back of the head. Therefore the ADM was significantly outperformed by the MWF in these spatial scenarios. This was also validated by the perceptual evaluation in which all MWF-based algorithms outperformed the ADM in the condition $S_{90}N_{270}$.

The two bottom rows of Table II show the SNRs at which the unprocessed reference SRTs were measured. It is observed that if a bilateral/binaural configuration was used, subjects always benefited from the best ear advantage. This means that if both ears are available, one of the ears has a better SNR than the other ear due to the headshadow effect and the positioning of the sound sources. This enables the human auditory system to focus on the ear with the best SNR. In condition S_0N_{60}, the noise source was close to the right ear, i.e., the ear used in the monaural evaluation. Therefore, the SRT level was much higher in the monaural presentation compared to the binaural presentation. Overall, it is observed that a binaural presentation, i.e., accessing the signals from both ears, always resulted in lower SRT values

compared to the monaural presentations. This has motivated the standard use of bilateral hearing aids in case of a bilateral hearing deficit (Libby, 2007).

During this study a perfect VAD was used to demonstrate the potential of the noise reduction performance of the MWF-based algorithms. It is clear that VAD performance will have an impact on the noise reduction performance of the algorithms. Simulations of Doclo *et al.* (2007) with a monaural spatially preprocessed MWF show that no large degradations (<1 dB) in performance should be expected when using an energy-based VAD at input SNRs higher than −2 dB. In the work of Wouters *et al.* (2008), hearing impaired subjects were evaluated with an adaptive version of this algorithm using a monaural energy-based VAD. Also in their experiments, a clear and robust gain in speech perception of several decibels was observed in multisource setups. Binaural algorithms also offer the possibility of integrating contralateral information into the VAD, which could lead to an improved VAD performance.

B. Adding contralateral microphones

Adding contralateral microphone signals to the ipsilateral hearing aid clearly comes at the cost of transmitting and processing those signals. To evaluate this trade-off, different microphone combinations were evaluated.

The objective evaluations showed that in single noise source scenarios with speech arriving from 0°, adding contralateral microphones introduced a large gain in noise reduction performance if the speech and noise sources were relatively close to each other (Fig. 1), i.e., within 60°. In other words the directional pattern generated by the MWF became more narrow when more microphones were used. This effect is well known in sensor array processing. Typically a large impact is obtained if additional sensors, in our case the contralateral front microphone, are placed sufficiently far away from the original sensors, thereby enhancing the size of the array. Extreme examples of this phenomenon are, in the specific case of hearing aids, often referred to as tunnel-hearing (Stadler and Rabinowitz, 1993). Soede *et al.* (1993) proved that very narrow beams in the horizontal hemisphere could be created when using several (4–17) microphones positioned on eyeglasses. If the speech and the single noise source were more spatially separated, adding more microphones did not result in large improvements in noise reduction performance (Fig. 1). This is due to the fact that in single noise source scenarios, the MWF only has to create a single null pointed toward the location of the noise source. As a consequence, adding more degrees of freedom, i.e., more microphones, to a two-microphone system does not significantly improve noise reduction performance.

Significant gains in noise reduction performance were also obtained during the objective evaluations for some asymmetric single noise source scenarios. In these scenarios, i.e., $S_{90}N_{270}$, $S_{45}N_{315}$, and $S_{90}N_{180}$, a significant improvement in performance were observed at the ear with the worst input SNR, i.e., the left ear. This is due to the asymmetrical setting of the speech source. Since the microphone inputs of the left hearing aid had a low input SNR, due to the headshadow effect, the noise reduction algorithm on this hearing aid produced a nonoptimal estimate of the speech component. However, if a contralateral microphone signal, which has a higher SNR, was added to the system, a better estimate of the speech component could be generated and noise reduction performance increased. One may interpret this as introducing the best ear advantage, used by our own auditory system, into the noise reduction algorithm.

One should be aware that this increased performance at the hearing aid with the worst SNR may be limited in daily life. The human auditory system focuses on the ear with the best SNR to listen to speech. The hearing aid with the large gain in SNR, obtained at the ear with the worst input SNR, will typically produce a similar output SNR as the hearing aid on the other side of the head. Therefore, perceptual SRT measurements with a bilateral/binaural hearing aid configuration will not show the large predicted gain in SNR. This was confirmed when comparing the objective and the perceptual data. It was shown that the actual gain in SRT correlates best with the predicted $\Delta\mathrm{SNR}_{\mathrm{SI}}$ performance obtained at the ear with the best input SNR. The more spectacular improvements found during the simulations, obtained at the ear with the worst input SNR, were not realistic predictions of the SRT gains. This illustrates that input as well as output SNRs or the best ear advantage should be taken into account when interpreting measurements of noise reduction gains for binaural or bilateral hearing aid configurations.

For the multiple noise source scenarios (Fig. 2), objective evaluations demonstrated that adding more microphones or more degrees of freedom does result in a significant gain in noise reduction. For the very asymmetrical condition, i.e., $S_0N_{60/120/180/210}$ (S_0N_{4a}), it was again observed that a larger benefit was obtained at the ear with the worst input SNR, i.e., the right ear.

In the perceptual evaluations, it was observed that in the scenarios S_0N_{60} and $S_{90}N_{270}$ the MWF_{2+1} and MWF_{2+2} outperformed the MWF_{2+0}. These observations confirm the objective evaluation, discussed earlier. In the triple noise source scenario only the MWF_{2+2} significantly outperformed the MWF_{2+0}, which can be explained by taking into account the degrees of freedom needed to reduce three noise sources. The grouped analysis of the perceptual data of the MWF and MWF-N showed that, in general, a three-microphone system, consisting of two ipsilateral and one contralateral microphone outperformed the two-microphone system. Adding a fourth microphone did not, in general, add a significant improvement over the three-microphone system. Intuitively this can be explained by the fact that adding a third microphone placed at the other side of the head will introduce a significant amount of "new information" to the noise reduction system. The fourth microphone will increase the degrees of freedom of the system, but its impact will be much smaller since it is located very close to the third microphone.

C. Noise reduction performance of the MWF-N

Van den Bogaert *et al.* (2008) showed that adding a partial noise estimate with $\eta=0.2$ to the MWF algorithm not only preserves the capability to localize the targeted speech component but also restores the capability to localize the

noise component. This is important for hearing aid users in terms of spatial awareness and release from masking. However, this clearly comes at the cost of some noise reduction. Figure 3 demonstrates the influence of the parameter $\eta = 0.2$ on ΔSNR_{SI} of the MWF_{2+0} and MWF_{2+2} in an environment with a realistic reverberation. It showed that the loss in noise reduction due to the partial noise estimate was dependent on its original noise reduction performance. This can be explained by using the relation between the output of the MWF and MWF-N. The output of the MWF-N [Eq. (1)] can be written as the sum of a scaled proportion of the input signal added to the output of the MWF (Van den Bogaert *et al.*, 2008), i.e.,

$$Z_{\text{MWF-N}_\eta L}(\eta) = \eta Y_{L,1} + (1-\eta) Z_{\text{MWF},L}, \tag{7}$$

$$Z_{\text{MWF-N}_\eta R}(\eta) = \eta Y_{R,1} + (1-\eta) Z_{\text{MWF},R}. \tag{8}$$

It was also observed that when adding a partial noise estimate with $\eta = 0.2$, the predicted performance, ΔSNR_{SI}, could drop below the performance of the ADM for some spatial scenarios (Fig. 3). This may be interpreted as a cost to sufficiently preserve the binaural cues of the speech and the noise component. However, during the perceptual evaluations, no significant difference was found between the ADM and the MWF_{2+0}-$N_{0.2}$ in scenarios S_0N_{60} and $S_0N_{90/180/270}$. Moreover, the ADM showed a significant loss in performance compared to all MWF-$N_{0.2}$ algorithms in the scenario $S_{90}N_{270}$ for reasons already discussed in the previous section. The lack of a significant SRT difference (ΔSRT_{algo}) between the ADM and the MWF_{2+0}-$N_{0.2}$, which was in contrast with the objective evaluation, may be explained by spatial release from masking. Since the MWF-$N_{0.2}$ preserved the localization of both the speech and noise component, a slightly better speech perception in noise compared to the performance predicted by ΔSNR_{SI} could be expected. The same spatial release from masking may also explain why the MWF_{2+M_C}-$N_{0.2}$ outperformed the MWF_{2+M_C} in the condition with the largest spatial separation between speech and noise sources, i.e., $S_{90}N_{270}$. In this condition, the MWF_{2+M_C}-$N_{0.2}$ produced a worse ΔSNR_{SI}, but since it preserves the user's ability to localize both the speech and the noise component correctly, a significantly better SRT could be obtained. This also illustrates that although ΔSNR_{SI} is a useful tool for predicting noise reduction performance, other factors such as binaural cues should be taken into account when evaluating speech enhancement by noise reduction algorithms in hearing aids.

VI. CONCLUSION

In Van den Bogaert *et al.* (2008), it was shown that MWF-based noise reduction approaches have interesting features in terms of preserving binaural cues and hence spatial awareness for hearing aid users. Unlike other noise reduction approaches, the MWF and MWF-N approaches are capable of using multimicrophone information; they can easily integrate contralateral microphone signals, and they inherently preserve the binaural cues of the speech component, independent of the angle of arrival of the signal. By preserving part of the noise component (MWF-N), the ability to localize both the speech and the noise component can be preserved. This paper presented a thorough evaluation of the noise reduction performance of the MWF and MWF-N algorithms in comparison with an unprocessed condition and an ADM, which is a commonly used noise reduction system in commercial digital hearing aids. This was done by evaluating noise reduction and speech perception performance in different speech-in-multitalker-babble scenarios. Three different research questions have been addressed.

First, it was shown that a two-microphone MWF (MWF_{2+0}) has approximately the same performance as an ADM. It does so while preserving the binaural cues of the speech component. Since the MWF operates independently of the angle of arrival of the signal, it easily outperformed the ADM if the speech signal was not arriving from the forward field of view. Moreover, in these scenarios the ADM may even reduce the speech perception of the hearing aid user compared to the unprocessed condition. This was observed during the perceptual evaluation of both the monaural (−5.4 dB) and the bilateral (−4.3 dB) ADM configuration in the spatial scenario $S_{90}N_{270}$. Large differences were observed when comparing the monaural with the bilateral data. It was observed that a bilateral presentation leads to an improved speech perception in noisy environments due to the best ear benefit. This confirms, although tests were performed with normal hearing subjects, the common practice of using bilateral/binaural hearing aids for a bilateral hearing impaired subject.

Second, different microphone combinations were evaluated. A significant gain in performance was found if one contralateral microphone signal was added to the ipsilateral hearing aid. This shows that transmitting microphone signals can result in a significant gain in noise reduction, especially in multiple noise source scenarios or if the speech and the noise source(s) are placed asymmetrically around the head. Adding a second contralateral microphone signal to the ipsilateral hearing aid did not, in general, show a significant SRT improvement in the perceptual evaluations.

Finally, it was demonstrated that adding a partial noise estimate to the MWF, large enough to sufficiently preserve the binaural cues to restore the directional hearing and spatial awareness (MWF-$N_{0.2}$), only slightly affects noise reduction performance. Moreover, perceptual evaluations showed that in some conditions ($S_{90}N_{270}$) the MWF-$N_{0.2}$ could even outperform the MWF, which may be due to improved spatial release from masking. The parameter η controls the amount of noise reduction. Therefore it may also be used as a control mechanism to maximize or to limit the amount of noise reduction if necessary. This can be done adaptively using sound classification algorithms, which are often available in present-day high-end digital hearing aids.

The study also demonstrated that carefully selected objective performance measures can be very useful in predicting the performance of noise reduction algorithms. However, one has to take into account psychophysical properties of the auditory system for a correct interpretation of these objective measures, e.g., the best ear benefit and spatial release from masking effects.

The chosen experimental setup, used to investigate and demonstrate the previously mentioned effects, does not represent all of the many conditions and noise sources encountered by hearing impaired subjects. The effect of head movements, which may interfere with the adaptation of the filters, other noise source scenarios, and a real-time implementation with a realistic, perhaps binaural, multimicrophone VAD were not discussed in this manuscript. Therefore more validation is preferred before implementing these algorithms into hearing aids. However, recent research with hearing impaired subjects indicates that a robust gain in speech perception is found when using a monaural real time MWF algorithm together with an energy-based VAD (Wouters *et al.*, 2008).

In conclusion, it seems that the binaural MWF-based algorithms offer a valid alternative for standard adaptive directional algorithms. Unlike these algorithms, the MWF does not rely on the direction of arrival of the speech signal nor on assumptions of the microphone characteristics of the hearing aids. In this paper, it was shown that the bilateral and the binaural MWF are capable of offering a good noise reduction performance in an environment with realistic acoustical parameters. Since it is often assumed that localization performance is mainly dominated by low-frequency ITD cues, future research may also include the investigation of a frequency dependent parameter η.

ACKNOWLEDGMENTS

T.V.d.B is funded by a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This research was partly carried out in the frame of the K.U. Leuven Concerted Research Action GOA-AMBioRICS. The scientific responsibility is assumed by its authors.

ANSI-SII (**1997**). "American National Standard Methods for Calculation of the Speech Intelligibility Index ANSI S3.5-1997," Acoustical Society of America, Melville, NY.

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acust. Acta Acust. **86**, 117–128.

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Byrne, D., Sinclair, S., and Noble, W. (**1998**). "Open earmold fittings for improving aided auditory localization for sensorineural hearing losses with good high-frequency hearing," Ear Hear. **19**, 62–71.

Desloge, J. G., Rabinowitz, W. M., and Zurek, P. M. (**1997**). "Microphone-array hearing aids with binaural output—part I: Fixed processing systems," IEEE Trans. Speech Audio Process. **5**, 529–542.

Doclo, S., Klasen, T. J., Van den Bogaert, T., Wouters, J., and Moonen, M. (**2006**). "Theoretical analysis of binaural cue preservation using multichannel Wiener filtering and interaural transfer functions," in Proceedings International Workshop on Acoustic Echo and Noise Control (IWAENC), Paris, France, pp. 1–4.

Doclo, S., and Moonen, M. (**2002**). "GSVD-based optimal filtering for single and multi-microphone speech enhancement," IEEE Trans. Signal Process. **50**, 2230–2244.

Doclo, S., Spriet, A., Moonen, M., and Wouters, J. (**2007**). "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," Speech Commun. **49**, 636–656.

Greenberg, J. E., Peterson, P. M., and Zurek, P. M. (**1993**). "Intelligibility-weighted measures of speech to interference ratio and speech system performance," J. Acoust. Soc. Am. **94**, 3009–3010.

Greenberg, J. E., and Zurek, P. M. (**1992**). "Evaluation of an adaptive beamforming method for hearing aids," J. Acoust. Soc. Am. **91**, 1662–1676.

Keidser, G., Rohrseitz, K., Dillon, H., Hamacher, V., Carter, L., Rass, U., and Convery, E. (**2006**). "The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers," Int. J. Audiol. **45**, 563–579.

Klasen, T. J., Van den Bogaert, T., Moonen, M., and Wouters, J. (**2007**). "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," IEEE Trans. Signal Process. **55**, 1579–1585.

Kompis, M., and Dillier, N. (**2001**). "Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. II. Experimental verification of the predictions," J. Acoust. Soc. Am. **109**, 1134–1143.

Libby, E. R. (**2007**). "The search for the binaural advantage revisited," Hear. Rev. **14**, 22–26.

Lotter, T. (**2004**). "Single and multimicrophone speech enhancement for hearing aids," Ph.D. thesis, RWTH Aachen.

Luo, F.-L., Yang, J., Pavlovic, C., and Nehorai, A. (**2002**). "Adaptive null-forming scheme in digital hearing aids," IEEE Trans. Signal Process. **50**, 1583–1590.

Maj, J. B., Royackers, L., Moonen, M., and Wouters, J. (**2006**). "Comparison of adaptive noise reduction algorithms in dual microphone hearing aids," Speech Commun. **48**, 957–970.

Maj, J. B., Wouters, J., and Moonen, M. (**2004**). "Noise reduction results of an adaptive filtering technique for dual microphone behind the ear hearing aids," Ear Hear. **25**, 215–229.

Noble, W., Sinclair, S., and Byrne, D. (**1998**). "Improvements in aided sound localization with open earmolds: Observations in people with high-frequency hearing loss," J. Am. Acad. Audiol **9**, 25–34.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.

Plomp, R., and Mimpen, A. M. (**1981**). "Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech reception threshold for sentences," Acustica **48**, 325–328.

Soede, W., Berkhout, A. J., and Bilsen, F. A. (**1993**). "Development of a directional hearing instrument based on array technology," J. Acoust. Soc. Am. **94**, 785–798.

Spriet, A., Moonen, M., and Wouters, J. (**2004**). "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," Signal Process. **84**, 2367–2387.

Stadler, R. W., and Rabinowitz, W. M. (**1993**). "On the potential of fixed arrays for hearing aids," J. Acoust. Soc. Am. **94**, 1332–1342.

Van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. (**2008**). "The effect of multimicrophone noise reduction systems on sound source localization in binaural hearing aids," J. Acoust. Soc. Am. **124**, 485–497.

Van den Bogaert, T., Klasen, T. J., Van Deun, L., Wouters, J., and Moonen, M. (**2006**). "Localization with bilateral hearing aids: Without is better than with," J. Acoust. Soc. Am. **119**, 515–526.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.

Welker, D. P., Greenberg, J. E., Desloge, J. G., and M., Z. P. (**1997**). "Microphone-array hearing aids with binaural output-part II: A two-microphone adaptive system," IEEE Trans. Speech Audio Process. **5**, 543–551.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.

Wittkop, T., and Hohmann, V. (**2003**). "Strategy selective noise reduction for binaural digital hearing aids," Speech Commun. **39**, 111–138.

Wouters, J., Luts, H., Eneman, K., Spriet, A., Moonen, M., Büchler, M., Dillier, N., Dreschler, W. A., Froehlich, M., Grimm, G., Volker, H., Houben, R., Leijon, A., Lombard, A., Mauler, D., Puder, H., Schulte, M., and Vorman, M. (**2008**). "Signal processing in hearing aids: Results of the HEARCOM project," J. Acoust. Soc. Am. **123**, 3166.

Wouters, J., and Vanden Berghe, J. (**2001**). "Speech recognition in noise for cochlear implantees with a two- microphone monaural adaptive noise reduction system," Ear Hear. **22**, 420–430.

Zurek, P. M., and Greenberg, J. E. (**2000**). "Two-microphone adaptive array hearing aids with monaural and binaural outputs," in Proceedings of the Ninth IEEE DSP Workshop, Hunt, TX.

Speech recognition by bilateral cochlear implant users in a cocktail-party setting

Philipos C. Loizou[a] and Yi Hu
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688

Ruth Litovsky and Gongqiang Yu
Binaural Hearing and Speech Laboratory, Waisman Center, 1500 Highland Avenue, Madison, Wisconsin 53705

Robert Peters and Jennifer Lake
Dallas Otolaryngology Associates, 7777 Forest Lane A-103, Dallas, Texas 75230

Peter Roland
Department of Otolaryngology/Head and Neck Surgery, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390

(Received 21 February 2008; revised 30 October 2008; accepted 5 November 2008)

Unlike prior studies with bilateral cochlear implant users which considered only one interferer, the present study considered realistic listening situations wherein multiple interferers were present and in some cases originating from both hemifields. Speech reception thresholds were measured in bilateral users unilaterally and bilaterally in four different spatial configurations, with one and three interferers consisting of modulated noise or competing talkers. The data were analyzed in terms of binaural benefits including monaural advantage (better-ear listening) and binaural interaction. The total advantage (overall spatial release) received was 2–5 dB and was maintained with multiple interferers present. This advantage was dominated by the monaural advantage, which ranged from 1 to 6 dB and was largest when the interferers were mostly energetic. No binaural-interaction benefit was found in the present study with either type of interferer (speech or noise). While the total and monaural advantage obtained for noise interferers was comparable to that attained by normal-hearing listeners, it was considerably lower for speech interferers. This suggests that bilateral users are less capable of taking advantage of binaural cues, in particular, under conditions of informational masking. Furthermore, the use of noise interferers does not adequately reflect the difficulties experienced by bilateral users in real-life situations.
 [DOI: 10.1121/1.3036175]

PACS number(s): 43.66.Sr, 43.66.Pn [JCM] Pages: 372–383

I. INTRODUCTION

It is well established that normal-hearing (NH) listeners have a remarkable ability to perceptually segregate a target voice amid a background of competing voices, a formidable task that has been termed the "cocktail-party" problem (e.g., Cherry, 1953). When the target voice and interfering voices (or noise) are spatially separated, listeners are able to take advantage of the favorable signal -to-noise ratio (SNR) at the "better" ear owing to the head-shadow effect. In addition, listeners are able to receive binaural advantage resulting from binaural unmasking in the low frequencies, facilitated by interaural time difference (ITD) differences between competing sources (Bronkorst and Plomp, 1988; Zurek, 1993). Aside from the use of interaural (time and level) differences, NH listeners exploit a number of other cues that help them cope with the cocktail-party problem. Much research (see review by Bronkorst, 2000) has been done to understand the perceptual processes used by NH listeners to segregate a target voice from competing, interfering voices, but relatively little is known about the processes used by bilateral cochlear implant (CI) users.

Bilateral cochlear implantation seeks to restore the advantages of listening with two ears. A number of studies have assessed speech recognition performance of adult (Tyler *et al.*, 2002; Gantz *et al.*, 2002; Muller *et al.*, 2002; van Hoesel and Tyler, 2003; Schleich *et al.*, 2004; Buss *et al.*, 2008) and pediatric (e.g., Litovsky *et al.*, 2006a) bilateral CI users in situations where the target and masker are either spatially coincident or separated. In the study by Tyler *et al.* (2002) data from nine adult subjects were collected three months after bilateral implantation. Speech intelligibility was tested both in quiet and in broadband noise presented from the left (−90°) or right (+90°). The level of the noise was adjusted for each subject to minimize ceiling or floor effects. When the noise was spatially separated from the speech signals, the subjects showed a significant head-shadow advantage but only a few subjects received benefit known as the binaural-interaction benefit, arising from the use of both ears over the ear with better SNR. Muller *et al.* (2002) reported data from nine bilateral implant users. Speech was presented from the

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

front and steady speech-shaped noise was presented at either +90° or −90° azimuth at a fixed SNR (10 dB). Results indicated significant head-shadow benefits as well as an additional, albeit small binaural-interaction benefit. Performance with bilateral implants for monosyllabic word recognition in quiet also showed improvement compared to that obtained with the better ear alone. Buss *et al.* (2008) reported large head-shadow benefits (37–38 percentage points) for 26 bilateral implant users participating in a multicenter clinical trial. Small binaural-interaction benefits (3–10 percentage points) were observed, but only after 1 yr postimplantation. In the bilateral CI studies described above, fixed SNRs were used for testing, which can produce data dominated by floor or ceiling effects. The study by van Hoesel and Tyler (2003) used an adaptive procedure to assess speech recognition performance. Broadband noise (nonmodulated) was used as a masker and presented to the subjects at 0°, 90°, or −90° (target was presented from the front). Overall, subject's performance improved with two implants, and the overall benefit (4–5 dB) was dominated by better-ear listening (head-shadow effect). A considerably smaller improvement of 1–2 dB was attributed to binaural interaction. Schleich *et al.* (2004) measured speech reception thresholds (SRTs) for 21 Med-El Combi 40/40+ bilateral users with continuous noise presented from the left or right. Results indicated a 6.8 dB head-shadow effect, a 0.9 dB binaural-interaction effect, and a 2.1 dB binaural summation effect. Litovsky *et al.* (2006b) measured SRTs in 34 simultaneously implanted adult Nucleus 24 users, after 3 months of bilateral hearing experience. With both target and competing speech in front, 15/34 subjects (44%) demonstrated a "binaural redundancy" effect, whereby the bilateral listening mode produced an advantage over one of the two unilateral conditions. With target in front and competing speech to the side, head-shadow effects averaged 5–6 dB and were found in 32/34 (94%) of subjects for at least one of the head-shadow comparisons (right or left). Binaural interaction, in contrast, averaged 1.95 dB and was found in 16/34 (47%) of subjects.

The above studies provided undoubtedly valuable information as to the benefit introduced via bilateral implants, but were limited in scope in several respects. First, with the exception of the study by Litovsky *et al.* (2006b), most studies used a single noise source making it difficult to predict the bilateral implant user's true performance in more realistic listening scenarios wherein multiple noise sources might be present. It is known from the NH literature that the number of masking sources as well as the spatial configuration of those noise sources can significantly affect performance (Bronkorst and Plomp, 1992; Yost *et al.*, 1996; Peissig and Kollmeier, 1997; Hawley *et al.*, 1999; 2004). Bronkorst and Plomp (1992), for instance, showed that speech intelligibility is reduced when noise sources are placed symmetrically around the target (i.e., across the two hemifields) than when they are placed asymmetrically, in part because the benefit from head shadow is obliterated.

Second, the temporal properties and spectral content of the masker can also affect performance. Hawley *et al.* (2004) observed a larger spatial release from masking by NH listeners when the maskers are comprised of speech or reversed speech, that is, when they contain linguistic content or context, compared to when noise (modulated or nonmodulated) maskers were used. Studies with bilateral implant users to date have been restricted to a single masker type; thus the extent to which the content and/or context of the masker are important remains to be understood. One important issue to recognize from the work of Hawley *et al.* (2004) is that in competing talker listening situations the interfering speech is likely to contain linguistic information which could be distracting or confused with the content of the target speech. This confusion is often classified as a form of "informational masking" (Brungart, 2001). Using a nonspeech pattern identification task, Kidd *et al.* (1998) showed that NH listeners benefited more from the spatial separation of the target and masker signals when the masker was informational in nature (no spectral overlap between target and masker) than when it was energetic (masking caused by the mere spectral overlap between the masker and target signals). When speech intelligibility is assessed using similar paradigms, the advantage of spatial separation is larger when there are substantial similarities in the information transmitted by the target and interferers, thus forcing listeners to rely more heavily on spatial cues to segregate competing sources from the target.

The effect of masker types on spatial separation benefits underscores the need to evaluate performance of bilateral CI users with both speech and nonspeech maskers; the use of nonspeech maskers might underestimate the advantage of bilateral implants for spatially segregated conditions in real-world situations. None of the aforementioned bilateral CI studies focused on this issue. Finally, it is of great interest to know how bilateral implant users perform compared to NH listeners. Such a comparison, however, is difficult to make given the differences in spatial configurations, testing environment (e.g., reverberation), and test material used in the various CI and NH studies. Several studies examined the effect of spectral/temporal characteristics of various maskers on performance in unilateral implant users (Nelson *et al.*, 2003; Stickney *et al.*, 2004) or with CI simulations (Qin and Oxenham, 2003). These studies showed that contrary to the benefit received by NH listeners when the masker is modulated, unilateral implant users did not benefit from such masker modulations (Stickney *et al.*, 2004).

In all, there are multiple factors that may influence the performance of bilateral implant users in real-world listening situations where multiple interfering sources might be present. The influence of these factors on bilateral CI performance is not well understood. The present study aims to assess the performance of bilateral users in more complex listening situations (cocktail party) with multiple competing sources emanating from various directions in space. The purpose of the study is to explore the interaction between the number of interfering noise sources, the magnitude of benefit incurred by better-ear listening for different target-masker spatial configurations, and the effect of informational/energetic masking on speech recognition. In this study we used the same simulated anechoic environment and the same stimuli presented to NH listeners in the study of Hawley *et al.* (2004). We will thus be in a unique position to compare

TABLE I. Biographical data for the bilateral CI subjects tested.

Subject	Duration of deafness (yrs)	Age (yrs)	CI use (yrs) left/right	Speech coding strategy	Probable cause of hearing loss
S1	19	61	5/5	ACE	Noise
S2	38	58	4/4	ACE	Measles
S3	17	36	3/4	ACE	Unknown
S4	11	65	4/3	ACE	Congenital
S5	22	68	5/5	ACE	Unknown
S6	>10	38	5/5	Speak	Unknown
S7	15	36	4/3	ACE	Unknown
S8	22	67	6/6	ACE	Hereditary

the bilateral users' performance against NH listeners' performance in the same listening conditions and ascertain the true benefit of bilateral implantation in more realistic noisy situations.

II. METHODS

A. Subjects

Eight postlingually deafened adults were recruited for testing. The participants were all bilateral CI patients fitted with the Nucleus 24 multichannel implant device manufactured by Cochlear Corporation. They were all native speakers of American English and were paid for their participation. All subjects had a minimum of three years experience with their implant devices. Biographical data for the subjects tested are given in Table I.

B. Experimental research processor

All subjects wore the Cochlear Esprit BTE processor on a daily basis. During their visit, subjects were temporarily fitted with the SPEAR3 wearable research processor. The SPEAR3 processor was developed by the Cooperative Research Center (CRC) for Cochlear Implant and Hearing Aid Innovation, Melbourne, Australia, in collaboration with HearWorks. The SPEAR3 has been used in a number of investigations to date as a way of controlling inputs to the CI system (e.g., van Hoesel and Tyler, 2003). Prior to the subjects' scheduled visit, the Seed-Speak Graphical User Interface (GUI) application was used to program the SPEAR3 processor with the individual users' threshold (T) and comfortable loudness levels (C). In addition, all participants (except subject S6) used the device programmed with the advanced combination encoder (ACE) speech coding strategy (e.g., see Vandali *et al.*, 2000) with all parameters (e.g., stimulation rate, number of maxima, frequency allocation table, etc.) matched to their clinical settings.

C. Speech and interferer stimuli

The speech stimuli were taken from the IEEE corpus (IEEE, 1969). The recordings were produced by two male speakers, each contributing half of the sentences (same stimuli that were used in Hawley *et al.*, 2004). Four of the longest sentences were reserved for use as interferers to ensure that all targets were shorter than the interferers. The remaining sentences were made into 64 lists of ten sentences each maintaining a single talker for each list. The mixture stimuli were constructed by having the interferers precede the target sentence (for about a second), and following the target sentence for another second. The interferer was either a female talker or speech-modulated noise that was computed using one of the four interferer sentences. For the speech-modulated noise, the envelope was extracted from the speech interferer and was used to modulate noise (originally filtered to match the long-term spectrum of the male talker), giving the same coarse temporal structure as speech. The envelope of running speech was extracted using a method similar to that described by Festen and Plomp (1990) by low-pass filtering a rectified version of the waveform. A first-order Butterworth low-pass filter was used with the 3-dB cutoff set at 40 Hz.

D. Simulated anechoic space

A set of free-field-to-eardrum (or anechoic) head-related transfer functions (HRTFs) previously measured in an acoustic manikin (Head Acoustics, HMS II.3) as described in the AUDIS catalog (see Blauert *et al.*, 1998) was used to simulate different spatial locations of the speech target and the interferer signals. HRTFs provide a measure of the acoustic transfer function between a point in space and the eardrum of the listener, and also include the high-frequency shadowing component due to the presence of the head and the torso. It should be noted that the use of HRTFs may not simulate accurately the intended source locations for CI users wearing the behind-the-ear microphones (as they lack pinna directionality), but rather for CI users (e.g., Advanced Bionics Corporation) wearing the in-the-canal microphones. On this regard, the data obtained with HRTFs might slightly overestimate the performance of CI users wearing behind-the-ear microphones. The duration of the impulse response was 256 sample points (at 16 kHz sampling frequency), amounting to a relatively short impulse response duration of 16 ms and therefore negligible reverberation. To generate the multisensor composite signals observed at the pair of microphones, the target and interferer stimulus for each position were convolved with the set of HRTFs for the left and right ears, respectively, thus generating a set of mixture signals for each of the two ears. In all experiments, HRTFs were used for stimuli simulating sources at a conversational distance of 1 m, with the vertical position (or elevation) adjusted at ear level.

All stimuli were presented to the listener through the auxiliary input jack of the SPEAR3 processor in a double-walled sound-attenuated booth (Acoustic Systems, Inc.). During the practice session, subjects were allowed to adjust the volume to reach comfortable level in both ears. For the unilateral conditions, either the left or right implant was activated. In the majority of simulated configurations the interfering virtual sound sources were situated to the listeners' right and were therefore less intense at the left than the right ear.

TABLE II. List of spatial configurations tested.

No. of interferers	Interferer type	Front	Left or distributed on both sides	Right or distributed on right	Right
One interferer	Modulated noise	0	−30°	60°	90°
Three interferers	Modulated noise	0°, 0°, 0°	−30°, 60°, 90°	30°, 60°, 90°	90°, 90°, 90°
One interferer	Female talker	0	−30°	60°	90°
Three interferers	Female talker	0°, 0°, 0°	−30°, 60°, 90°	30°, 60°, 90°	90°, 90°, 90°

E. Conditions

The simulated target location was always at the front (0° azimuth). Subjects were tested in conditions with either one or three interferers. Up to three interferers were placed either in the front (0°,0°,0°), distributed on both sides (−30°,60°,90°), distributed on the right side (30°,60°,90°), or from the same location on the right side (90°,90°,90°). Note that −90° means that the interferer was located to the left of the listener, and 90° means that it was located to the right. Table II summarizes these conditions. The level of each interferer was fixed and the overall level of the interferers was thus naturally increased as more interferers were added.

Each listener completed testing in six to ten sessions of 1–1.5 h each, spanning 2 days. During these sessions, two SRT measurements for each of the 16 conditions (2 numbers of interferers × 4 spatial configurations × 2 interferer types) were obtained. To minimize any order effects, all conditions were randomized among subjects. Different sets of sentences were used in each condition. Subjects S5 and S6 were not available for testing in a subset of the conditions (single interferer with speech-modulated noise).

F. SRT measurement

SRTs were measured using a method similar to that developed by Plomp (1986) and used in the NH study (Hawley *et al.*, 2004) with which the data will be compared. Listeners were seated in the sound-attenuated booth in front of a terminal screen. At the start of each session practice SRTs were given with three interferers for each interferer type to familiarize the subject with the interferer types and the task. At the start of each SRT measurement, the level of the target was initially very low. The subject heard the same target sentence and interferer combination repeatedly. After each presentation, the subject's task was to repeat as many words as possible. After each response, the experimenter pressed the return key and the same target sentence and interferer combination was replayed, but with the signal-to-interferer ratio (computed based on the ratio of signal-to-interferer energies) increased by 4 dB. The subject repeated the words/sentence he/she heard orally, and when the experimenter determined (based on a written sentence transcript) that the subject reproduced more than half of the sentence correctly, the first recording was made of the number of keywords correct. From that point on, a SRT was measured using a one-down/one-up adaptive SRT technique targeting 50% correct speech reception (Levitt, 1971) using an approach that was successfully used in studies with NH listeners (Hawley *et al.*, 2004). After listening to each sentence, the subjects were asked to repeat what they heard. Each IEEE sentence had five designated keywords and these words were in capital letters in the transcript, e.g., "The BIRCH CANOE SLID on the SMOOTH PLANKS." The experimenter compared what was repeated by the subject with the displayed text and typed in the number of keywords found correct. The SNR of the next trial was raised by 2 dB if two or fewer keywords were correct and the SNR was lowered by 2 dB if three or more keywords were correct. The entire transaction was logged in a data file and displayed on the experimenter's computer monitor for verification of scoring reliability. The SRT was determined by averaging the level presented in the last eight trials.[1] The content, number, and locations of the interferers were fixed throughout the run in each condition.

G. Data analysis

The data were analyzed in a similar way to the approach taken by Hawley *et al.*, (2004) in order to draw comparisons from the present results with those of NH listeners. More precisely, we used the raw SRTs for unilateral and bilateral stimulations to derive the following three advantages that are potentially introduced by the availability of binaural listening: total advantage, monaural advantage, and binaural advantage (also known as binaural interaction).

The total advantage is the improvement in performance (decrease in SRT) observed when the masker-target spatial separation is introduced compared with when the interferer and target are both presented from front (0° azimuth). It is determined by subtracting the bilateral SRT of a given spatially-separated condition from the SRT of the corresponding unseparated condition. This overall benefit is also known as spatial release from masking and is assumed to contain the advantages from both head shadow and binaural advantage (binaural interaction).

The monaural advantage is defined as the improvement in performance (decrease in SRT) observed when listening with the better ear, i.e., the ear with the more favorable SNR. It is determined by subtracting the SRT of a given unilateral spatially separated condition corresponding to the ear contralateral to the interferer location, from the SRT of the cor-

responding unilateral unseparated condition (0° azimuth). So, for instance, if the interferer is presented from the left hemifield (e.g., −30°), the monaural advantage is computed by subtracting the SRT obtained with the right implant from the SRT obtained in the unseparated condition (0°) with the right implant. For the (−30°,60°,90°) condition, the monaural advantage was computed using the SRTs obtained with the left implant since the majority of the interferers came from the right hemifield. Note that the head-shadow advantage was computed differently in other studies in which similar effects in bilateral CI users were measured (van Hoesel and Tyler, 2003; Schleich *et al.*, 2004; Buss *et al.*, 2008). In those studies, head shadow was computed by subtracting the unilateral SRT obtained when the interferer was on the contralateral side of the implant from the SRT obtained when the interferer was on the ipsilateral side of the implant. A different method for measuring the head-shadow advantage is used in the present study for two reasons. First, the intent was to be consistent with the method used in Hawley *et al.* (2004) for assessing better-ear listening. The adoption of the same definition of monaural advantage will enable appropriate comparisons between the two studies. Second, methods used in other studies are better suited for making comparison when the interferer(s) is (are) placed symmetrically across the two hemifields. In this study the interferer(s) was (were) placed mostly on the right and in asymmetrical configurations; thus we are unable to compute the head-shadow advantage in the manner done by others (e.g., van Hoesel and Tyler, 2003; Schleich *et al.*, 2004).

The binaural advantage (or binaural interaction) is thought to assess the contribution of binaural processing to advantages introduced in spatial separation. This advantage reflects the benefit from listening binaurally over listening with just the better unilateral ear (i.e., implant contralateral to the interferer), and is determined by subtracting the monaural advantage with the better-ear condition from the total advantage of separation. That is, binaural advantage is equal to total advantage minus monaural advantage. Binaural advantage data are reported for all conditions, including the condition in which the interferer originated from −30°. Note that the study by Hawley *et al.* (2004) did not report binaural advantage data for interferers at −30°, as they only tested their subjects monaurally with the left ear.

III. RESULTS AND DISCUSSION

The three advantages are discussed next, along with the raw SRT values obtained in the various conditions (see Figs. 1 and 2). For each of the two types of interferers, Analysis of Variances (ANOVAs) were conducted to assess the interaction between the number of interferers and other factors on performance (SRT values). For the speech interferer, a three-way ANOVA (2 numbers of interferers × 3 listening modes × 4 interferer locations) revealed a significant effect [$F(1,5)=16.6$, $p=0.01$] of the number of interferers (1 versus 3), with SRTs being significantly higher in the presence of 3 versus 1 interferer(s). In addition, there was a significant effect [$F(2,10)=15.9$, $p=0.001$] of listening mode (bilateral versus unilateral left or unilateral right), a nonsignificant effect of interferer location, a significant interaction [$F(6,30)=5.9$, $p<0.005$] between listening mode and interferer location, and a nonsignificant interaction between number of interferers and other factors. Similar effects were noted with the noise interferer. Similar ANOVAs were conducted for data analyzed in terms of advantage, and noted no significant interactions between number of interferers and most of other factors. A significant interaction [e.g., for female talker, $F(4,20)=3.7$, $p=0.02$] was only found with listening mode by interferer location.

Given the absence of significant interaction between number of interferers and other factors (with the exception of one interaction found when the data were analyzed in terms of advantage), the data were reanalyzed for main effects and interactions separately for one and three interferers. Results from the one-interferer conditions will be discussed first, followed by results from the three-interferer conditions.

A. One interferer

The results for a single interferer are shown in Fig. 1 (upper left panels).

1. Raw SRTs

Figure 1 [panels (A) and (B)] shows the raw SRTs obtained with a single noise and speech interferers. The discussion on SRTs that follows focuses on differences in performance relative to conditions in which both target and interferers were at 0°. As shown in Fig. 1 [panels (A) and (B)], mean SRT values decreased in the bilateral condition as the interferer moved away from the target (located at 0°) regardless of the interferer type. For the unilateral condition with the left implant alone, mean SRTs increased when the interferer was at −30° and then dropped when the interferer was at 60° and 90° [Fig. 1, panels (A) and (B)]. The increase in SRT was expected since the left ear was on the same side of the interferer. For the unilateral condition with right implant alone, SRTs increased as expected when the interferer was on the right (i.e., at 60° and 90°) and decreased when the interferer was on the left (−30°) [see Fig. 1, panels (A) and (B)]. Overall, the left and right unilateral SRTs mirrored each other, as expected [Fig. 1, panels (A) and (B)]. A three-way ANOVA (2 types of interferers × 3 listening modes × 4 interferer locations) revealed a significant effect [$F(2,8)=5.4$, $p=0.03$] of listening mode, a significant effect [$F(3,12)=4.1$, $p=0.032$] of interferer location, a nonsignificant effect of interferer type, a significant interaction [$F(6,24)=33.3$, $p<0.005$] between listening mode and interference location, and a marginally significant interaction [$F(2,8)=4.4$, $p=0.049$] between interferer type and listening mode. *Posthoc* analyses on effect of listening mode indicated a significant difference in SRTs [$F(1,6)=8.5$, $p=0.026$] between the bilateral and right-implant conditions, a nonsignificant difference [$F(1,6)=2.03$, $p=0.21$] between the bilateral and left-implant conditions and a nonsignificant difference [$F(1,6)=1.4$, $p=0.27$] between left- and right-implant conditions (a significant interaction was noted, however, in the left- versus right-implant analysis). *Post hoc* analyses on effect of interferer location suggested that SRTs were lower [$F(3,4)=7.8$,

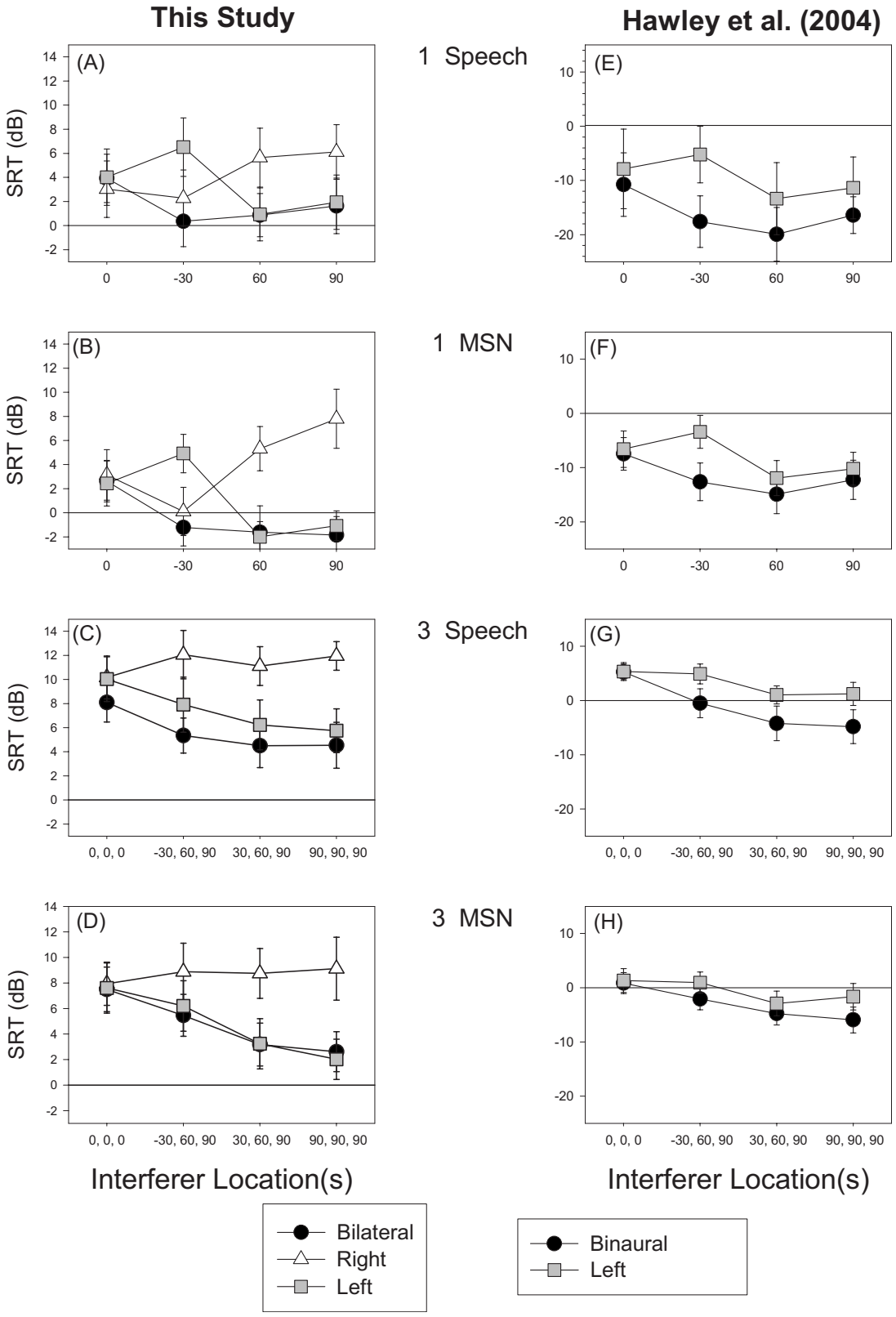


FIG. 1. [Left panels (A)–(D)] Mean SRT values obtained by bilateral CI users in various spatial configurations with different number of interferers. Data for speech and modulated speech-shaped noise interferers are shown separately. [Right panels (E)–(H)] Mean SRT values obtained by NH listeners (Hawley *et al.*, 2004) in the same conditions. Note that the CI data are plotted using a different *y*-axis range for better visual clarity. Error bars indicate standard deviations.

$p=0.037$] when the interferers were at 60° compared to 0°. The interaction between listening mode and interference location was due to the fact that performance improved significantly as the interferers moved away from the target in the bilateral and left-implant conditions but not in the right-implant conditions. This was not surprising since the interferers moved closer to the right implant (i.e., the ear with the lowest SNR). *Post hoc* (Scheffe) tests revealed that the per-

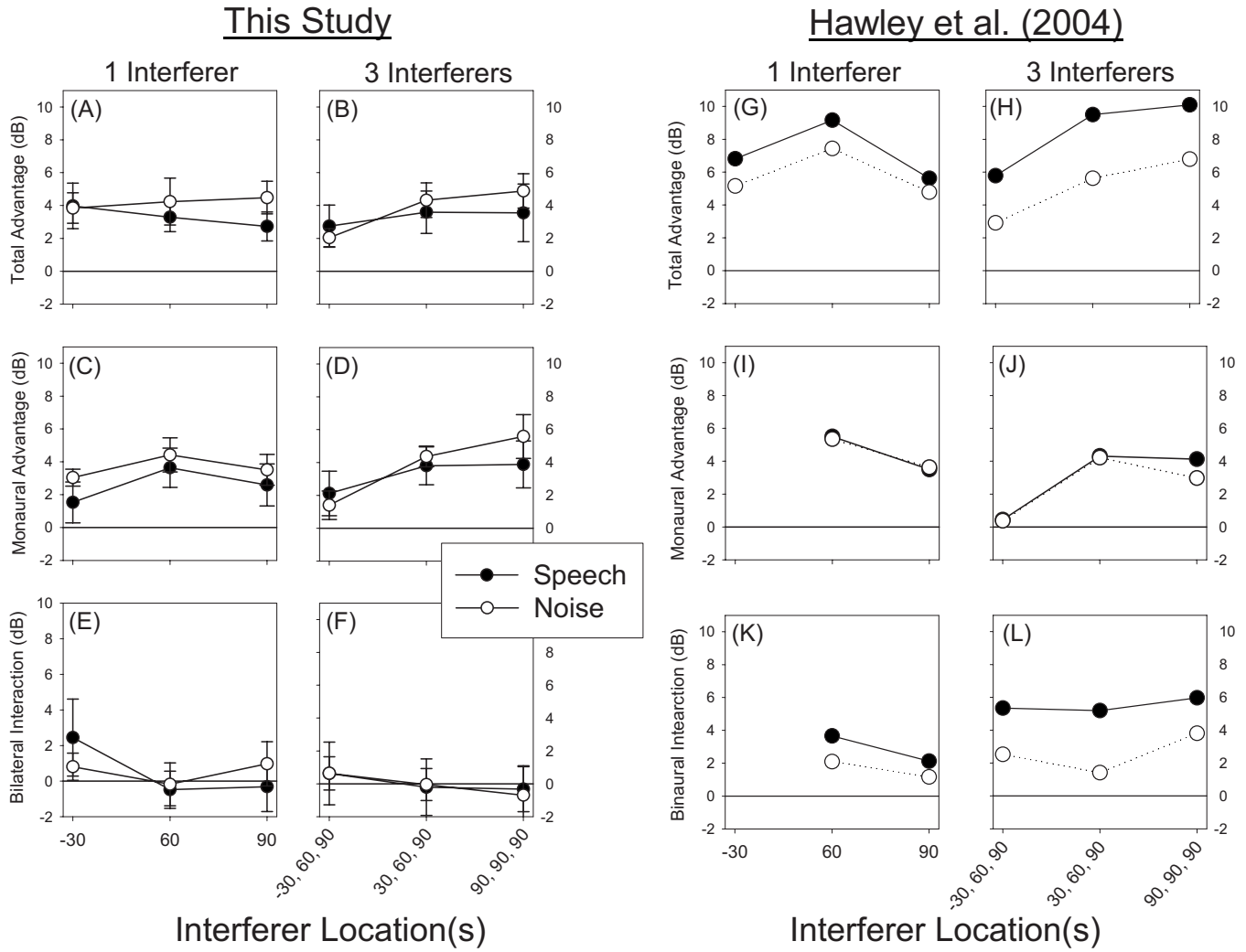


FIG. 2. [Left panels (A)–(F)] Binaural benefits, in terms of total advantage, monaural advantage, and binaural-interaction advantage (see text for details) all measured in decibels, obtained by bilateral CI users. [Right panels (G)–(L)] Binaural benefits obtained by NH listeners (Hawley *et al.*, 2004) in the same conditions. Error bars indicate standard errors of the mean.

formance with interferers at 90° was significantly ($p < 0.05$) better than performance with interferers at 0° with either the bilateral or left-implant condition.

Figure 1 (top two rows) contrasts the SRT values obtained in this study against those obtained by NH listeners in the study by Hawley *et al.* (2004) [see Fig. 1, panels (E) and (F)]. The overall pattern in bilateral performance is the same as that obtained by NH listeners in that performance improves (lower SRT values) as the interferer(s) move away from the target. The absolute SRT values obtained by NH listeners, however, are notably lower, by about 10 dB in the noise-interferer condition and by about 15–20 dB in the speech interferer condition [compare Fig. 1, panels (A) versus (E), and panels (B) versus (F)]. The pattern, however, obtained with the left implant alone does not follow the pattern followed by NH listeners when presented with the stimuli monaurally via the left ear. For the CI users, the unilateral SRT values obtained with the left implant are nearly identical to the SRT values obtained bilaterally [see Fig. 1, panels (A) and (B)]. This outcome reflects the absence of binaural advantage (more on this in Sec. III A 2), since the bilateral SRT values are not better (lower) than those of the better-ear (which is the left implant for interferers at 60° and 90°) SRTs. In contrast, the binaural SRT values obtained by NH listeners [Fig. 1, panels (E) and (F)] were always lower (more so in speech interferers) than the SRT values obtained with the left ear monaurally, reflecting a binaural advantage.

2. Advantages of separation

Figure 2 (leftmost column) quantifies the mean advantage of separation in terms of total advantage [Fig. 2, panel (A)], monaural advantage [Fig. 2, panel (C)], and binaural-interaction advantage [Fig. 2, panel (E)]. The mean (across subjects) total advantage of target-interferer separation was 3–4 dB in all conditions [Fig. 2, panel (A)]. A two-factor ANOVA (3 interferer locations × 2 interferer types) revealed no significant effect [$F(2,10)=0.1$, $p=0.9$] of interferer location, no significant effect of interferer type [$F(1,5)=0.4$, $p=0.5$], and no significant interactions [$F(2,10)=0.4$, $p=0.6$]. This suggests that the total advantage received by bilateral CI users under the conditions tested here was not affected by the location and type of interferer.

The mean monaural advantage [Fig. 2, panel (C)] due to better-ear listening, was around 3–4 dB for the noise interferer (open symbols) and 1–3 dB for the female talker interferer (filled symbol). A two-factor ANOVA (3 interferer locations × 2 interferer types) revealed no significant effects, suggesting that the monaural advantage was not affected by the location or type of interferer. With the exception of one condition (female interferer at 90°), the mean monaural advantage was significantly above zero in all conditions ($p < 0.05$, one-tail t-tests).

The mean binaural advantage [Fig. 2, panel (E)] was smaller than 1 dB in all but one condition. Two-factor ANOVA (3 interferer locations × 2 interferer types) revealed no significant effects. The mean binaural advantage was not significantly ($p > 0.05$) above zero in any condition.

The bilateral implant users' data are contrasted in Fig. 2 [panels (G), (I), and (K)] with the data reported in Hawley *et al.* (2004) with NH listeners. Note that the data for −30° azimuth are missing in Fig. 2 [panels (I) and (K)] because Hawley *et al.* (2004) did not test the right ear monaurally. The total advantage [Fig. 2, panel (G)] seen in NH listeners with the speech interferer is nearly double (7–10 dB) of that obtained by bilateral CI users, but the total advantage received by NH listeners with the noise interferer was lower (5–7 dB) and more similar to that obtained by bilateral CI users. The monaural advantage [Fig. 2, panel (I)] observed in NH listeners for the noise interferer was about 4–6 dB and similar to that received by bilateral CI users. The monaural advantage received by NH listeners for the speech interferer was about 2–3 dB higher than that obtained by bilateral CI users. Similarly, the binaural advantage [Fig. 2, panel (K)] observed in NH listeners was about 2–4 dB higher than that observed in bilateral implant users.

B. Three interferers

The SRT results for three interferers are shown in Fig. 1 (column 2) and results for advantage of separation are shown in Fig. 2 (column 2).

1. Raw SRTs

Figure 1 [panels (C) and (D)] shows the mean raw SRT values obtained with three interferers. SRTs decreased in the bilateral condition as the interferer moved away from the target (located at 0°) for both interferer types. Overall, the mean SRT values with three interferers were 4–6 dB higher than the corresponding SRT values with one interferer in the bilateral condition [compare panels (A) and (C) or panels (B) and (D) in Fig. 1]. A similar, albeit larger, increase in SRT values was also observed with NH listeners (Hawley *et al.*, 2004) with three interferers in the binaural condition [Fig. 1, panels (G) and (H)]. A three-way ANOVA (2 types of interferer × 3 listening modes × 4 interferer locations) revealed a significant effect [$F(2,8)=21.9$, $p<0.001$] of listening mode (bilateral versus unilateral left and unilateral right), a significant effect [$F(3,12)=10.9$, $p<0.001$] of interferer location, a nonsignificant effect of interferer type, a significant interaction [$F(6,24)=4.8$, $p<0.05$] between listening mode and interference location, and a nonsignificant interaction between interferer type and listening mode. As indicated by the above ANOVA, performance was similar to the noise and female talker interferers. *Post hoc* tests indicated that the bilateral performance was significantly [$F(2,4)=22.7$, $p=0.006$] better (i.e., lower SRT values) than the performance obtained with the left-implant alone [Fig. 1, panel (C)], which is the implant with the better SNR in most conditions. This was true for the speech interferers but not for the noise interferers [see Fig. 1, panel (C)]. Performance with the left-implant or bilateral implants at (30°,60°,90°) and (90°,90°,90°) was significantly ($p<0.05$) better than corresponding performance at (0°,0°,0°). The interaction between listening mode and interference location was due to the fact that performance improved significantly as the interferers moved away from the target in the bilateral and left-implant conditions but not in the right-implant conditions [see Fig. 1, panels (C) and (D)]. This was not surprising since the right implant was ipsilateral to the location of the interferers in most conditions.

2. Advantages of separation

Figure 2 (column 2) quantifies the mean advantage of separation in terms of total advantage [Fig. 2, panel (B)], monaural advantage [Fig. 2, panel (D)], and binaural advantage [Fig. 2, panel (F)]. The mean total advantage of target-interferer separation ranged from 2 to 5 dB across conditions (see Fig. 2, panel (B)]. A two-factor ANOVA (3 interferer locations × 2 interferer types) revealed a significant effect [$F(2,12)=6.4$, $p=0.013$] of the interferer's location, but no significant effect of interferer type [$F(1,6)=0.07$, $p=0.8$] and no significant interactions [$F(2,12)=2.1$, $p=0.2$]. *Post hoc* tests indicated that the significant effect of the interferer's location was due to the significant ($p=0.005$) difference in performance between the (−30°,60°,90°) and (90°,90°,90°) configurations in the noise-interferer conditions. There was no significant difference in the size of total advantage with the speech interferers [Fig. 2, panel (B), filled symbols] across the various spatial configurations.

The mean monaural advantage [Fig. 2, panel (D)] was in the range of 1–6 dB for the noise interferers and 2–4 dB for the female talker interferers. A two-factor ANOVA (3 interferer locations × 2 interferer types) revealed a significant effect [$F(2,8)=6.7$, $p=0.02$] of interferers' location, no significant effect of interferer type [$F(1,4)=0.7$, $p=0.4$], and a significant interaction [$F(2,8)=5.3$, $p=0.03$] between interferer type and location. *Post hoc* tests indicated that the interferer's location affected the size of the monaural advantage significantly in the condition with noise interference but not in the condition with the female talker interferer. Monaural advantage was largest when the three interferers were located on the right (90°,90°,90°) and was significantly ($p=0.012$) larger than in the condition in which the interferers were asymmetrically placed around the listener (i.e., −30°,60°,90°) [see Fig. 2, panel (D)].

The mean binaural advantage [Fig. 2, panel (F)] was near 0 dB in all conditions and for both types of interferers. A two-factor ANOVA (3 interferer locations × 2 interferer types) revealed no significant effects, confirming the absence of binaural advantage in all conditions.

The data obtained in Hawley *et al.* (2004) with NH listeners are contrasted in Fig. 2 (column 4). The total advantage [Fig. 2, panel (H)] received by NH listeners with the female interferers is nearly double (7–10 dB) than seen in bilateral CI users (3–4 dB) [see Fig. 2, panel (B)]. In contrast, the total advantage observed in NH listeners with the noise interferers was comparable (only 1–2 dB higher) to that observed in bilateral CI users [compare open symbols in panels (B) and (H) in Fig. 2]. The monaural advantage observed in NH listeners [see Fig. 2, panel (J)] for the noise-interferer conditions was 1–2 dB higher (at most) than that observed by bilateral CI users [see panels (D) and (J), open symbols in Fig. 2], but the monaural advantage observed in NH listeners with the speech interferers was roughly the same as that observed in the bilateral CI users [see panels (D) and (J), filled symbols in Fig. 2]. Finally, the binaural advantage observed in NH listeners [see Fig. 2, panel (L)] was about 2–6 dB higher than that observed in bilateral users [see Fig. 2, panel (F)]. In all, it is clear that the total advantage observed in bilateral users due to the interferer-target spatial separation is dominated by the unilateral benefit, i.e., access to the implant with a more favorable SNR. This benefit persists even when multiple (three) interferers are present [see Fig. 2, panel (D)].

IV. GENERAL DISCUSSION

The present study was intended to assess the performance of bilateral CI users in more complex and realistic listening environments than previously studied. This was achieved by measuring SRTs both unilaterally (one implant alone) and bilaterally in four different spatial configurations with one and three interferers. Experiments were designed to answer the important question of how performance of bilateral CI users is affected by multiple versus single interferers, and whether there are notable effects of the type of interferers on the ability of the listeners to benefit from spatial separation of target speech and interferers. The extent to which performance was affected by the number, type, and location of interferers was directly compared with data from NH listeners, who were presented with the same test material and the same simulated anechoic space (Hawley *et al.*, 2004). The data were analyzed in terms of monaural and binaural effects, with the intent of isolating the individual contributions of the monaural advantage (i.e., better-ear listening) and binaural advantage. The data analysis revealed a number of effects that are discussed next.

A. Monaural advantage

The monaural advantage (i.e., better-ear listening) received by bilateral CI users was significantly better than 0 dB, ranging from 2 to 6 dB, and was largest when the interferer(s) was (were) mostly energetic (speech-shaped noise). As shown earlier, the monaural advantage was not affected by the type of interferer (speech versus noise) used. Nonetheless, the monaural advantage was found to be robust as it was maintained even when three interferers were present. It was comparable to that obtained by NH listeners (Hawley *et al.*, 2004) with the same test material and the same HRTFs (i.e., the same simulated anechoic environment). This suggests that in real-world settings, bilateral users receive significant benefit owing to having access to an implant with a more favorable SNR.

The significant monaural advantage found in the present study is consistent with that found in other studies investigating similar issues in bilateral CI users (van Hoesel and Tyler, 2003; Muller *et al.*, 2002; Buss *et al.*, 2008; Schleich *et al.*, 2004). It should be pointed out that all other studies reported head-shadow advantage, which is measured differently (see Sec. II G), but nonetheless assessed the intelligibility benefit incurred by better-ear listening. A 4 dB head-shadow advantage was found in the study by van Hoesel and Tyler (2003) for a single interferer (speech-shaped noise) presented either to the left or right of the listener. Schleich *et al.* (2004) reported a 6.8 dB head-shadow benefit in bilateral users of the Med-El Combi 40/40+ CI when presented with a single interferer (speech-shaped noise).

B. Binaural advantage

There is ample evidence in the binaural hearing literature suggesting that when both ears are available, NH listeners are able to receive a 3–5 dB binaural advantage (Zurek, 1993; Hawley *et al.*, 2004). Much of this advantage is attributed to good ITD sensitivity, particularly in the low frequencies (Bronkhorst and Plomp, 1988). No such binaural advantage was found, however, with bilateral CI users in the present study. This is consistent with previous reports in which the binaural-interaction effect was found to be very small (1–2 dB) and marginally or nonsignificant (van Hoesel and Tyler, 2003; Muller *et al.*, 2002; Schleich *et al.*, 2004; Buss *et al.*, 2008). Specifically, a 2 dB binaural-interaction benefit was reported for five bilaterally implanted Nucleus CI users in the study by van Hoesel and Tyler (2003). That benefit, however, was found to be only weakly significant ($p=0.04$) and was measured with a single interfering steady-state noise source presented to the left (−90°) or to the right (90°) of the listeners. Similarly, Muller *et al.* (2002) reported a small, but statistically significant, contribution of the binaural-interaction effect of 10.7 percentage points for sentences presented in speech-shaped noise (at a fixed SNR level of 10 dB) from either 90° or −90° azimuth to nine Med-El bilateral users. A 0.9 dB benefit of binaural-interaction effect was reported by Schleich *et al.* (2004) with 21 bilateral Med-El users when presented with a single interferer (speech-shaped noise). Litovsky *et al.* (2006b) reported an average binaural-interaction effect of 1.95 dB. Although this effect was overall statistically significant, there was relatively large intersubject variability (SD=3.3 dB). Examination of individual subject performance suggested that the overall group effect can be attributed to a few individuals with larger effect sizes with some of the subjects also showing a decrement in the bilateral listening conditions. In summary, the binaural-interaction benefit reported in most studies is quite small, often not significant, and variable across subjects.

The lack of binaural interaction (binaural advantage) in bilateral CIs can be attributed to several factors including poor ITD sensitivity (van Hoesel and Tyler, 2003; Grantham *et al.*, 2007), poor spectral resolution (small effective number of frequency channels), and difference/asymmetry in the state of the binaural auditory pathways (Long *et al.*, 2006; Litovsky *et al.*, 2006b). Grantham *et al.* (2007) reported that the best ITD threshold (among 11 bilateral Med-El users) was about 400 μs, with only 5 of 11 subjects achieving thresholds <1000 μs. Moderate ITD threshold values around 100–150 μs were obtained by five nucleus bilateral users in the study by van Hoesel and Tyler (2003), but only at low stimulation rates (<200 pps). These ITD threshold values are still appreciably larger than the ITD values ($\sim$70 μs) achievable by naive NH listeners (Wright and Fitzgerald, 2001) and more than an order of magnitude larger than the sensitivity of 10–20 μs reported in highly trained listeners (Durlach and Colburn, 1978). The fact that bilateral CI users may not achieve benefits from binaural hearing may be due to the fact that the etiology of hearing loss might differ in the two ears. This is further complicated by possible differences in electrode insertion depth in the two ears. Some might argue that such a mismatch in insertion depth might be beneficial in terms of providing complementary information and contributing to a binaural summation effect (Schleich *et al.*, 2004; Buss *et al.*, 2008), but can be quite harmful to the mechanisms involved in processing ITD information (Long *et al.*, 2003). Finally, the lack of synchronization of the two (independently run) speech processors worn by bilateral CI users can affect the coding of ITD information in the fine structure of the signal (Majdak *et al.*, 2006), at least for pulse rates as high as 800 pulses/s. The outcome of the present study, as well as those of others, highlights the importance of developing strategies capable of preserving ITD information in a way that will be perceived by bilateral CI users.

C. Informational and energetic masking

Energetic masking is typically present in noise interferers and is generally accounted for by peripherally based models of the auditory periphery that take into account spectral overlap of the target and the interferer. Unlike the noise interferers, however, the speech interferers (e.g., competing talkers) produce both energetic and nonenergetic components of masking. The nonenergetic masking, often called informational masking, is attributed to confusion that may be caused by content similarity between the target and the interferer. In complex listening situations informational masking is thought to be at least partly responsible for the difficulty that listeners experience in teasing apart the content carried by the target in the presence of the interferer (Brungart, 2001).

There is evidence to suggest that informational masking is reduced considerably when the target and interferer signals are spatially separated, and the benefit of spatial separation can be significantly larger in the presence of speech interferers compared with noise interferers (Peissig and Kollmeier, 1997; Kidd *et al.*, 1998; Hawley *et al.*, 2004). That was not found to be the case in the present study with bilateral CI users. The benefit from spatial separation was roughly the same with either noise or speech interferers (Fig. 2).

It is interesting to note that the total advantage as well as the monaural advantage received by bilateral users were comparable (within 1–2 dB) to that of NH listeners (Hawley *et al.*, 2004) in nearly all noise-interferer conditions for both single and multiple interferers. In contrast, a large difference (4–7 dB) was observed in nearly all speech interferer conditions between bilateral CI users and NH listeners. We cannot attribute this disparity to differences in the way that the target was "glimpsed" during momentary dips in the amplitude of the interferer since both speech and (modulated) noise interferers contained "envelope dips" which provided occasional favorable SNR segments. Rather, we consider the possibility that the difference in performance with speech interferers reflects the poorer ability of bilateral CI users to take advantage of directional cues under conditions of informational masking. In NH listeners (e.g., Hawley *et al.*, 2004), binaural cues are relied on more heavily to segregate target and interfering sounds particularly when other cues for source segregation are not available. That is, when the target and interferers can be more easily confused with one another, as is the case when the interferers consists of speech rather than noise, binaural cues that provide differential spatial information for the target and interferers become particularly salient (see also Freyman *et al.*, 2001, 2007). The bilateral CI users tested here did not demonstrate such release, suggesting that their weaker ability to integrate binaural cues reduced their experience of spatial advantage under conditions of informational masking. In fact, although there were no statistically significant effects regarding advantage of spatial separation and interferer type, there was a slight trend for a larger advantage in the presence of noise interferers rather than speech.

Another indicator of informational masking would have been higher SRTs in the conditions with speech interferers compared with noise interferers. Several unilateral implant studies (e.g., Stickney *et al.*, 2004; Nelson *et al.*, 2003; Nelson and Jin, 2004) have shown that CI users generally perform better with noise interferers (modulated or nonmodulated) than with speech interferers, suggesting the presence of informational masking. However, a different outcome emerged in the present study. There was no significant difference in performance (in terms of absolute SRT values) with the noise and speech interferers when either single or multiple interferers were present. One factor that could account for the difference in outcome between prior studies (Stickney *et al.*, 2004; Nelson and Jin, 2004) and the present study is that here we used a different-sex talker for the interferer, which may have reduced the extent of informational masking (Brungart, 2001). CI users in this study operated at SNR levels generally above 0 dB (see Fig. 1), a range over which it may be unlikely for them to have successfully extracted information from the interferer, i.e., confused the interferer at low SNRs. Thus, it is possible that the speech and noise interferers resulted in similar degradation of information in the target and that the large SRTs for the CI users abolish the potential differences in interferer type as far as energetic versus informational masking effects are con-

cerned. Alternatively, in the present study, informational masking may have occurred in the presence of both speech and noise interferers. Given that degraded spectral information occurs readily in CI users, it is possible that listeners confounded the target and interferers just as readily in both conditions and that the advantage of spatial separation seen here was due to release from a combination of energetic and informational masking regardless of the type of interferer.

In addition, although persons fitted with bilateral CIs demonstrate measurable benefits from having a second CI compared with listening unilaterally, the SNR at which they are able to hear speech in the presence of interferers is markedly higher than the levels at which NH listeners are able to cope with in the same challenging situation. In the presence of a single speech interferer [Fig. 1, panels (A)–(D)], bilateral CI users consistently require target speech that is several decibels higher than the interferer, compared with NH listeners who can perform the task at negative SNRs [Fig. 1, panels (E)–(H)]. This result is consistent with anecdotal reports by CI users that everyday noisy situations are challenging even with a second CI.

V. CONCLUSIONS

Unlike previous bilateral studies (van Hoesel and Tyler, 2003; Schleich *et al.*, 2004; Buss *et al.*, 2008) which considered only one interferer (steady-state noise) emanating from a single location in space, the present study considered more realistic listening situations wherein multiple interferers were present, and in some cases originating from both hemifields. Aside from noise interferers, speech interferers which are known to contain informational masking were also considered. This was done to examine whether bilateral CI users receive any release of informational masking when the target and interferer are spatially separated, as found in the NH literature (e.g., Kidd *et al.*, 1998). SRTs were measured both unilaterally (one implant alone) and bilaterally in four different spatial configurations with one and three interferers. The data were analyzed in terms of binaural benefits including better-ear listening and binaural advantage (binaural interaction). After comparing the present data with those by NH listeners (Hawley *et al.*, 2004) who were presented with the same test material in the same listening environment, we can draw the following conclusions.

- The SRT values obtained by bilateral CI users are significantly higher (worse), by about 10 dB in the noise-interferer condition (one interferer) and by about 15–20 dB in the speech interferer condition (one interferer), than those obtained by NH listeners in the same listening conditions. This may have rendered the speech and noise maskers equally difficult to ignore for this population.
- The difference between NH and CI users in terms of the overall spatial release of masking (total advantage) was considerably smaller than the differences between groups in raw SRTs.
- The overall total advantage (overall spatial release of masking) of target-interferer separation ranged from 2 to 5 dB across all conditions. This advantage was maintained even when multiple interferers were present. A larger overall advantage (5–10 dB) was observed with NH listeners (Hawley *et al.*, 2004), particularly with speech interferers.
- The monaural advantage (i.e., better-ear listening) received by bilateral CI users was large ranging from 1 to 6 dB and was largest when the interferers were mostly energetic. This benefit was comparable (within 1–2 dB) to that obtained by NH listeners (Hawley *et al.*, 2004) in nearly all conditions.
- No binaural advantage (binaural interaction) was found in the present study with either type of interferer (speech or noise).
- The total advantage as well as the monaural advantage received by bilateral users were comparable (within 1–2 dB) to that of NH listeners (Hawley *et al.*, 2004) in nearly all noise-interferer conditions for both single and multiple interferers. In contrast, a large difference (4–7 dB) was noted in nearly all speech interferer conditions between bilateral CI users and NH listeners. This difference is due to the fact that there was no effect of interferer type for the CI users, and suggests that bilateral users are less capable of taking advantage of binaural cues for source segregation under conditions of informational masking compared with NH listeners. In fact, there is little evidence that bilateral users experience informational masking in a way that is akin to that experienced by NH listeners. This outcome also indicates that the use of steady-state noise interferers (which are utilized extensively in bilateral studies) does not adequately reflect the difficulties bilateral implant users experience in real-life noisy situations.

The present study extended the findings of prior bilateral studies to complex listening settings (cocktail party) and showed that bilateral implants can yield substantial benefit when the target and interferers are spatially separated. This benefit is dominated for the most part by better-ear listening, i.e., access to an implant with a favorable SNR. Compared to NH listeners who receive a moderate benefit (3–5 dB) from binaural interaction (Zurek, 1993), bilateral users do not receive such benefit. A highly plausible reason for the lack of binaural interaction is the poor ITD sensitivity of bilateral users (van Hoesel and Tyler, 2003; Grantham *et al.*, 2007), particularly at high (>1000 pulses/s) stimulation rates commonly used in commercial implant devices. Further research is warranted to develop signal processing strategies that preserve ITD information (even at high stimulation rates). Such strategies will hold promise for introducing binaural advantage in bilateral CIs.

ACKNOWLEDGMENTS

The authors would like to thank the two anonymous reviewers for their valuable comments and suggestions. Research was supported by Grant No. R01 DC007527 to P. C. Loizou and Grant No. R01 DC003083 to R. Litovsky from the National Institute of Deafness and other Communication Disorders, NIH.

[1]Following the procedure used in Hawley *et al.* (2004), the SRT was computed as the average of the SNR levels on trials 4 through, and including, 11. There was no 11th trial, but we used instead the level that would have been presented by the result of the 10th trial.

Blauert, J., Brueggen, M., Bronkhorst, A. W., Drullman, R., Reynaud, G., Pellieux, L., Krebber, W., and Sottek, R. (**1998**). "The AUDIS catalog of human HRTFs," J. Acoust. Soc. Am. **103**, 3082.

Bronkhorst, A. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acta Acust. **86**, 117–128.

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Brungart, D. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Buss, E., Pillsbury, H., Buchman, C., Pillsbury, C., and Clark, M. (**2008**). "Multicenter U.S. bilateral Med-El cochlear implantation study: Speech perception over the first year of use," Ear Hear. **29**, 20–32.

Cherry, E. (**1953**). "Some experiments on the recognition of speech with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.

Durlach, N., and Colburn, S. (**1978**). "Binaural Phenomena," in *Handbook of Perception*, edited by E. Carterette and M. Friedman (Academic, New York), Vol. **IV**, pp. 365–466.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception SRT for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Helfer, K. S., and Balakrishnan, U. (**2007**). "Variability and uncertainty in masking by competing speech," J. Acoust. Soc. Am. **121**, 1040–1046.

Gantz, B. J., Tyler, R. S., Rubinstein, J. T., Wolaver, A., Lowder, M., Abbas, P., Brown, C., Hughes, M., and Preece, J. P. (**2002**). "Binaural cochlear implants placed during the same operation," Otol. Neurotol. **23**, 169–180.

Grantham, D., Ashmead, D., Ricketts, T., Haynes, D., and Labadie, R. (**2007**). "Interaural time and level difference thresholds for acoustically presented signals in post-lingually deafened adults fitted with bilateral cochlear implants using CIS+processing," Ear Hear. **29**, 33–44.

Hawley M., Litovsky R., and Colburn H. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**(6), 3436–3448.

Hawley, M., Litovsky, R., and Culling, J. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

IEEE Subcommittee (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **AU-17**, 225–246.

Kidd, G., Jr., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Litovsky, R., Johstone, P., and Godar, S. (**2006a**). "Benefits of bilateral cochlear implants and/or hearing aids in children," Int. J. Audiol. **45**, 78–91.

Litovsky, R. Y., Parkinson, A., Arcaroli, J., and Sammath, C. (**2006b**). "Clinical study of simultaneous bilateral cochlear implantation in adults: A multicenter study," Ear Hear. **27**, 714–731.

Long, C., Carlyon, R., Litovsky, R., and Downs, D. (**2006**). "Binaural unmasking with bilateral cochlear implants," J. Assoc. Res. Otolaryngol. **7**, 352–360.

Long, C., Eddington, D., Colburn, S., and Rabinowitz, W. (**2003**). "Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user," J. Acoust. Soc. Am. **114**, 1565–1574.

Majdak, P., Laback, B., and Baumgartner, W. (**2006**). "Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing," J. Acoust. Soc. Am. **120**, 2190–2201.

Muller, J., Schon, F., and Helms, J. (**2002**). "Speech understanding in quiet and noise in bilateral users of the Med-El Combi 40/401 cochlear implant system," Ear Hear. **23**, 198–206.

Nelson, P. B., and Jin, S.-H. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.

Nelson, P. B., Jin, S.-H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.

Peissig, J., and Kollmeier, B. (**1997**). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660–1670.

Plomp, R. (**1986**). "A signal-to-noise ratio method for the speech-reception SRT of the hearing impaired," J. Speech Hear. Res. **29**, 146–154.

Qin, M., and Oxenham, A. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Schleich, P., Nopp, P., and D'Haese, P. (**2004**). "Head shadow, squelch and summation effects in bilateral users of the Med-El Combi 40/40 +cochlear implant," Ear Hear. **25**, 197–204.

Stickney, G., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Tyler, R. S., Gantz, B. J., Rubinstein, J. T., Wilson, B. S., Parkinson, A. J., Wolaver, A., Preece, J. P., Witt, S., and Lowder, M. W. (**2002**). "Three month results with bilateral cochlear implants," Ear Hear. **23**, 80S–89S.

van Hoesel, R., and Tyler, R. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.

Vandali, A., Whitford, L., Plant, K., and Clark, G. (**2000**). "Speech perception as a function of electrical stimulation rate using the Nucleus 24 cochlear implant system," Ear Hear. **21**, 608–624.

Wright, B. A., and Fitzgerald, M. B. (**2001**). "Different patterns of human discrimination learning for two interaural cues to sound-source location," Proc. Natl. Acad. Sci. U.S.A. **98**, 12307–12312.

Yost, W., Dye, R., and Sheft, S. (**1996**). "A simulated "cocktail party" with up to three sound sources," Percept. Psychophys. **58**, 1026–1036.

Zurek, P. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255–276.

Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate

K. G. Munhall[a)]
Department of Psychology and Department of Otolaryngology, Queen's University, Humphrey Hall, 62 Arch Street, Kingston, Ontario K7L 3N6, Canada

E. N. MacDonald
Department of Psychology and Department of Electrical and Computer Engineering, Queen's University, Humphrey Hall, 62 Arch Street, Kingston, Ontario K7L 3N6, Canada

S. K. Byrne and I. Johnsrude
Department of Psychology Queen's University, Humphrey Hall, 62 Arch Street, Kingston, Ontario K7L 3N6, Canada

(Received 29 July 2008; revised 20 October 2008; accepted 30 October 2008)

Talkers show sensitivity to a range of perturbations of auditory feedback (e.g., manipulation of vocal amplitude, fundamental frequency and formant frequency). Here, 50 subjects spoke a monosyllable ("head"), and the formants in their speech were shifted in real time using a custom signal processing system that provided feedback over headphones. First and second formants were altered so that the auditory feedback matched subjects' production of "had." Three different instructions were tested: (1) *control*, in which subjects were naïve about the feedback manipulation, (2) *ignore headphones*, in which subjects were told that their voice might sound different and to ignore what they heard in the headphones, and (3) *avoid compensation*, in which subjects were informed in detail about the manipulation and were told not to compensate. Despite explicit instruction to ignore the feedback changes, subjects produced a robust compensation in all conditions. There were no differences in the magnitudes of the first or second formant changes between groups. In general, subjects altered their vowel formant values in a direction opposite to the perturbation, as if to cancel its effects. These results suggest that compensation in the face of formant perturbation is relatively automatic, and the response is not easily modified by conscious strategy.
 [DOI: 10.1121/1.3035829]

PACS number(s): 43.70.Aj, 43.70.Mn, 43.70.Bk, 43.70.Fq [AL] Pages: 384–390

I. INTRODUCTION

Human speech and animal vocalizations are dramatically influenced by the sounds that the speakers hear themselves producing (Smotherman, 2007). Both clinical and laboratory studies demonstrate that this auditory feedback effect occurs because vocal motor control is normally dependent on the sensory consequences of talking. Hearing-impaired individuals show characteristic patterns of distortion and increased speech variability in the absence of normal feedback (Cowie and Douglas-Cowie, 1992), and a range of experimental perturbations of acoustic feedback produce rapid compensations in subsequent productions (Burnett *et al.*, 1998; Houde and Jordan, 1998; Kawahara, 1995; Lane and Tranel, 1971; Purcell and Munhall, 2006; Villacorta *et al.*, 2007). Similar phenomena can be demonstrated in a variety of species ranging from songbirds (Brainard and Doupe, 2000) to beluga whales (Scheifele *et al.*, 2005).

One of the most common questions about the way in which auditory feedback affects speech production is whether subjects are aware that the feedback is being manipulated. An implication of this concern is that subjects' compensations might be under conscious control or result from a response strategy. This is a complex question that invokes an old controversy in the neuroscience of behavior—the idea of a reflex and the conflict between the ideas of voluntary or consciously controlled action and involuntary or automatic movements.

This controversy between volitional and automatic control is at the core of both philosophical and neurophysiological debates about the control of action [see Prochazka *et al.* (2000) for a discussion], and its resolution is beyond the scope of this paper. However, empirical contributions to the question of voluntary/involuntary responses can make the discussion more explicit and well defined. Three types of data bear on the reflexive nature of response to sensory stimulation. First, one can investigate the timing of motor responses: more rapid responses can be viewed as more automatic. Second, the influence of training can be investigated in order to determine whether a response is modifiable. Finally, the influence of instructions or task on the stereotypy of the response can be assessed. If the response is unchanged by training, instructions, or task, it is more likely to be automatic and to operate independently of conscious control.

Each of these general approaches is evident in the sensorimotor control of speech literature. Rapid motor responses

[a)]Author to whom correspondence should be addressed. Electronic mail: kevin.munhall@queensu.ca

to mechanical perturbations have been frequently reported [e.g., see Gracco and Abbs (1985)], and the responses exhibit distinct patterns depending on the timing of the perturbation. Responses to auditory feedback perturbations also depend on the time from the onset of the perturbation. Rapid responses to perturbations of vocal amplitude (Bauer *et al.*, 2006) and vocal pitch (Burnett *et al.*, 1998; Hain *et al.*, 2000) have been reported, but a later response can be observed in some cases (Hain *et al.*, 2000; Kawahara, 1995). A distinction between automatic rapid responses (<200 ms) and a more voluntary slower response between 300 and 700 ms after perturbation onset has been suggested [e.g., see Hain *et al.* (2000)].

When subjects were given experience with perturbed auditory feedback, there is some evidence of an increased independence from the auditory feedback signal and thus an ability to resist compensatory behavior. Pick *et al.* (1989) exposed subjects to an increased noise level with visual feedback about their speaking level. This visual feedback was effective in helping people resist the effects of background noise level, but their learning seemed to be associated with a strategy to reduce the overall speaking level under all auditory conditions. Zarate and Zatorre (2008) found that compared to nonmusicians, trained singers are much better able to ignore pitch perturbations. However, even the singers made small compensations in response to the voice feedback pitch shifts.

Instructions or task orientation has been found to produce limited modifications of responses under certain conditions. Pick *et al.* (1989) found that there was little influence of instructions to ignore the presence of background noise. Their subjects exhibited the Lombard effect and increased their vocal amplitude to match the background noise even when they had been explicitly instructed not to. With the same instructions, Hain *et al.* (2000) found that subjects always produced changes in vocal pitch when fundamental frequency was perturbed. However, when instructed to raise or lower their pitch in response to a perturbation or move in the opposite direction to the perturbation, subjects did show the ability to make changes in the timing and magnitudes of compensation.

In the present study, we focus on the response to perturbations of vowel formant frequency. Rapid signal processing systems now allow the frequency of one or more formants to be shifted up or down in frequency in real time. In response to changes in auditory feedback, talkers adjust the frequency of their produced formants in the opposite direction in frequency, presumably in order to compensate for the perturbations (Houde and Jordan, 1998; Purcell and Munhall, 2006; Villacorta *et al.*, 2007). These compensations persist when feedback is returned to normal, suggesting that some type of learning has taken place. However, the extent to which such compensations are relatively automatic is uncertain.

Many of the initial studies of speech compensation used a very gradual perturbation in which the discrepancy between the actual formant frequency and the modified feedback was changed by small increments trial by trial. The subjects in these studies often seem to be unaware that their speech was being modified and thus have no particular knowledge of the nature of the manipulation [e.g., see Purcell and Munhall (2006)]. However, the same feedback perturbations can be carried out more abruptly by changing the formant frequencies in larger increments (step changes). These abrupt step changes can be noticeable and thus introduce the possibility of a more explicit strategic response.

Here, three groups of subjects are tested under different instructional conditions using this sudden large perturbation paradigm: (1) subjects who are naive to the purposes of the experiment and are not told about the feedback perturbation, (2) subjects who are told that their speech heard from the headphones may sound wrong and that they should ignore this feedback, and (3) subjects who are briefed in detail about the perturbation paradigm and instructed to not compensate for the perturbation. The aim is to examine the influence of such instructions on the pattern of formant compensation. If the second or third instruction condition reduces or eliminates the compensatory response, then we may conclude that the role of auditory feedback in speech motor control is not mandatory and is instead open to cognitive intervention. If, on the other hand, the same pattern of response is evident across all three instructional conditions, this would suggest that the maintenance of formant frequency is a more automatic response customarily tuned by auditory feedback. If so, we may conclude that studies of formant perturbation, even when the changes are large and abrupt, are minimally influenced by strategic efforts of subjects since the standard response is difficult to suppress.

II. METHODS

A. Subjects

Fifty-four female participants (mean age=20.1 yr, range: 17–25 yr) were tested in a single session. Since our experiment involves tracking individual formants, we chose to test females exclusively in order to minimize the variability in frequency of the first and second formants across participants. All subjects spoke English as their first language and reported no speech or language impairments. Hearing thresholds were assessed over a range of 500–4000 Hz Three subjects were eliminated because of heightened hearing thresholds in some frequency bands (>20 dB HL). Data from one additional subject were lost due to experimenter error.

B. Equipment and real-time formant shifting

Equipment was the same as that previously reported in Purcell and Munhall (2006). Subjects' speech was recorded using a headset microphone (Shure WH20). The signal was amplified using a Tucker-Davis Technologies MA3 microphone amplifier and low-pass filtered at a cutoff frequency of 4500 Hz (Frequency Devices 901 filter). This signal was digitized at a 10 kHz sampling rate and was filtered in real time to produce formant shifts using a National Instruments PXI-8176 controller. Noise was added using a Madsen Midimate 622 audiometer, and the voice signal and noise were presented to the subject using headphones (Sennheiser HD 265).

Detection of voicing and shifting of formants was performed as previously described in Purcell and Munhall

(2006). Briefly, the manipulation of auditory feedback was achieved by filtering the voice in real time. Voicing was detected using a statistical amplitude threshold technique. Formants in the speech were determined using an iterative Burg algorithm (Orfandidis, 1988). The formant estimates were used to calculate the filter coefficients so that a pair of spectral zeroes was positioned at the location of the existing formant frequency and a pair of spectral poles was positioned at the desired frequency of the new formant. This filtering reduced the spectral energy in the region of the produced formants and emphasized the energy in the region of the desired formants. The filtering and thus the formant shifts were implemented as soon as voicing was detected. The formant frequency estimate and new filter coefficients were computed every 900 μs.

C. Procedure and experimental conditions

Testing was performed in an Industrial Acoustics Company sound insulated room. Prior to data collection, a screening procedure was carried out to determine the best autoregressive model order for formant tracking. Subjects produced seven English vowels spaced across the vowel space in an /hVd/ context five times in a random order. During the experiment subjects produced 95 repetitions of "head" at a natural rate and speaking level with timing controlled by a visual prompt on a monitor.

The subjects were randomly assigned to one of three conditions in which the experimental instructions were manipulated: (1) *control* ($N=18$), in which subjects were naïve about the feedback manipulation; (2) *ignore headphones* ($N=15$), in which subjects were told that their voice might sound different and to ignore what they heard in the headphones; (3) *avoid compensation* ($N=17$), in which subjects were informed in detail about the manipulation and were told not to compensate. This group was given the suggestion that focusing on their kinesthetic feedback might help them avoid compensation. All three groups produced the English word "head" repeatedly in the following experimental phases (see Fig. 1): (1) *baseline*. Fifteen repetitions were spoken with normal feedback (i.e., amplified and with noise added but no shift in formant frequency) to assess baseline F1 and F2 values. In this and subsequent conditions, subjects were encouraged to speak at a natural rate and speaking level with timing controlled by a prompt on a monitor. Each prompt lasted 2.5 s, and the intertrial interval was approximately 1.5 s. (2) *Perturbation*. Forty repetitions of the utterance "head" were produced with F1 and F2 values shifted in frequency to match the formant values for each subject's production of the vowel /æ/ as in "had." (3) *Return to normal feedback*. Forty repetitions of the utterance were produced with normal feedback (i.e., the formant shift was abruptly turned off).

D. Offline formant analysis

In order to examine the extent to which the shifting of formants affected the acoustics of produced vowels on subsequent trials, segmentation boundaries for the vowel in each trial were first calculated using an automated process that examined the harmonicity of the power spectrum. These boundaries were then inspected visually and corrected if required. Once segmented, offline estimates of the formant frequencies were calculated by sliding a 25 ms analysis window for each estimate, one ms at a time (shifts of ten speech samples), and applying a similar algorithm to that used in online shifting. For each trial, a single "steady-state" F1 value was determined by averaging 40% of the F1 estimates starting almost halfway through the vowel (i.e., from 40% of the way to 80% of the way through the vowel). Single steady-state values for F2 and F3 were calculated in the same manner. Prior to data collection, a screening procedure was conducted to select the best autoregressive model order (a parameter used in the real-time formant tracking algorithm) for each talker. This reduced gross errors in formant tracking. However, for some participants, occasionally, one of the formants would be misinterpreted as another (e.g., F2 being misinterpreted as F1). These misinterpreted estimates were found and corrected by visually examining a plot with all of the steady-state F1, F2, and F3 estimates for each individual.

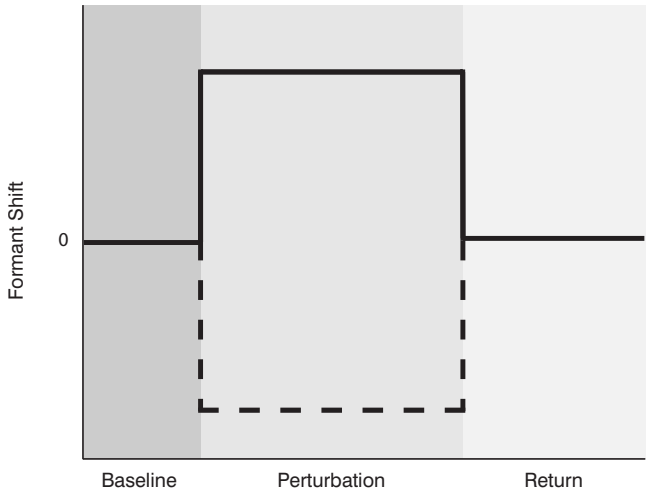


FIG. 1. Feedback shift for the three phases of the experiment. In the baseline condition (15 utterances), the feedback was normal (unperturbed). For the perturbed condition (40 utterances), the first formant was increased (solid line) and the second formant was decreased (dashed line). The size of the shift for each formant was calculated as the difference between the average frequency for that formant in "head" and in "had" separately for each individual. For the return condition (40 trials), the feedback was returned to normal.

III. RESULTS

While the frequency shifts applied to F1 and F2 in the perturbed condition varied for each individual, the average frequencies of the feedback shifts were similar across the three instruction groups. This was confirmed by an analysis of variance (ANOVA) with no significant main effect of instruction group for either the F1 [$F(2,47)=0.696$, $p=0.50$] or the F2 [$F(2,47)=0.115$, $p=0.89$] frequency shift. See Table I for the mean perturbed feedback shifts.

In order to determine the magnitude of subjects' compensations to these shifts, the formant estimates for their produced vowels were normalized for each individual by subtracting their baseline mean, defined as the mean of the estimates for trials 6–15. These trials correspond to the last ten utterances in the baseline condition. For each trial, the

TABLE I. Mean formant feedback shift and compensation in Hz for F1 and F2 for the three instruction conditions. Standard errors of the means are shown in parentheses.

	Control	Ignore headphones	Avoid compensation
F1 shift	183.2(15.7)	209.2(26.3)	221.1(28.6)
F1 compensation	−56.6(6.0)	−61.0(13.7)	−55.2(5.8)
F2 shift	−235.2(18.8)	−216.8(28.7)	−224.9(33.0)
F2 compensation	68.3(11.3)	54.9(14.6)	55.0(16.2)

TABLE II. Mean formant frequency in Hz for the baseline, perturbation, and return phases for the three instruction conditions. Standard errors of the means are shown in parentheses.

Conditions		F1	F2
Baseline:	Avoid	740.0(13.6)	2078.1(46.4)
	Ignore	732.2(18.1)	2071.6(31.8)
	Control	762.3(10.6)	2062.2(25.8)
Perturbation:	Avoid	686.4(10.9)	2126.4(50.7)
	Ignore	671.5(17.9)	2127.0(38.0)
	Control	711.2(9.5)	2122.4(21.7)
Return:	Avoid	723.6(15.2)	2088.6(48.6)
	Ignore	733.2(15.2)	2053.5(33.5)
	Control	750.8(9.7)	2062.3(25.0)

normalized F1 and F2 estimates were averaged for each group and are plotted in Fig. 2. From the figure, it is clear that on average all three groups changed their production of F1 and F2 in a direction opposite to the manipulation.

To quantify the change in production, three intervals were defined based on the last ten trials in each of the experimental conditions (trials 6–15 for the Baseline, trials 46–55 for the perturbation phase, and trials 86–95 for the return phase). In these three intervals, it is assumed that formant production has reached a steady state. The nonnormalized F1 and F2 estimates in each interval were averaged for each individual (Table II). A repeated measure ANOVA with interval as a within-subjects factor and instruction group as a between-subjects factor confirmed a significant effect of interval for both F1 [$F(1.7,94)=76.52$, $p<0.001$] and F2 [$F(2,94)=32.60$, $p<0.001$]. Multiple pairwise comparisons using Bonferroni correction confirmed that the results for the perturbation phase were significantly different from both the baseline and return phases. (See Table I for the mean compensation magnitude computed by subtracting the baseline mean from the perturbation mean for each subject.) The difference between the baseline and return phases was not significant. No significant effect of instruction group was found for either F1 [$F(2,47)=1.61$, $p=0.21$] or F2 [$F(2,47)=0.06$, $p=0.95$]. Also, no significant interaction of interval and instruction group was found for either F1 [$F(3.4,94)=1.33$, $p=0.27$] or F2 [$F(4,94)=0.96$, $p=0.43$]. This confirms that the instructional set did not affect the magnitude of the compensation in either F1 or F2; subjects modified their productions even when instructed explicitly not to.

Although the averaged results presented in Fig. 2 show consistent compensation, it is important to note that the individual responses vary greatly. The compensation in both F1 and F2 of each individual, defined by measuring the difference in the average formant frequency between the perturbation and baseline intervals used above, is plotted in Fig. 3. From the figure, it is clear that there is a wide range of compensation across individuals in both F1 and F2 for all three groups. For each of the three instruction conditions, there was also at least one subject who compensated by changing their formants in the direction of the perturbation. Following behavior similar to this has been observed in other auditory perturbation experiments [e.g., Burnett *et al.* (1998)]. Despite this large intersubject variability, it is apparent that there is consistent compensation for most subjects and a broad similarity across conditions. We also observe a

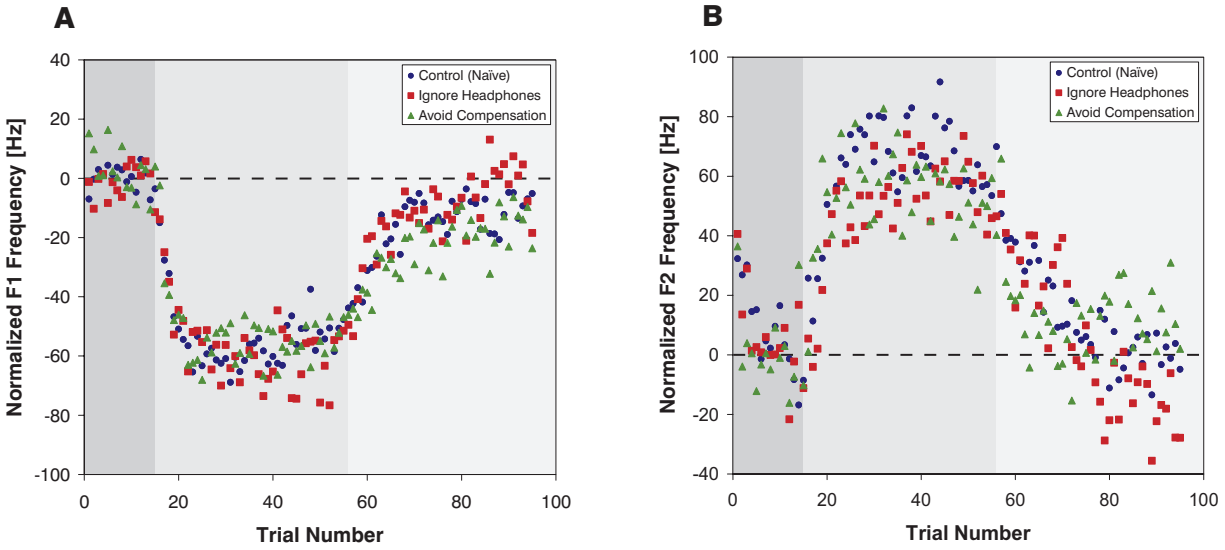


FIG. 2. Average normalized F1 and F2 frequencies for each trial for the control (circles), ignore headphones (squares), and avoid compensation (triangles) groups. (a) and (b) show the results in each trial for F1 and F2, respectively. The three phases of the experiment (baseline, perturbed, and return) are indicated with shading of increasing lightness.

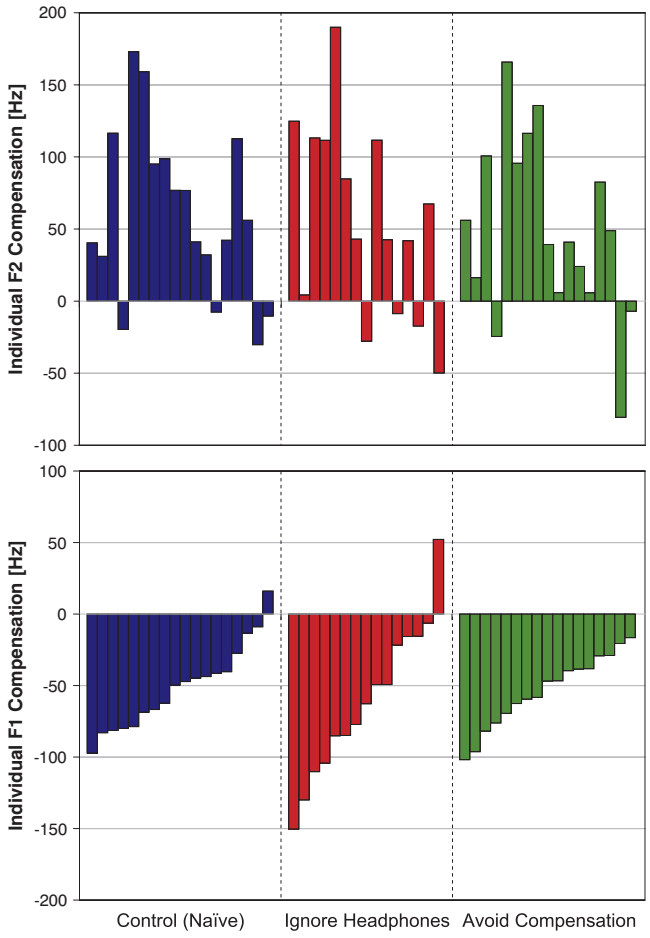


FIG. 3. Individual compensation magnitudes measured as the difference in the average formant frequency between the perturbation and baseline intervals for F2 (top panel) and F1 (bottom panel) for subjects in the control ($N=18$), the ignore headphones ($N=15$), and the avoid compensation ($N=17$) groups. Each bar represents a different subject; homologous bars in the top and bottom panels are from the same subject.

small correlation between the magnitudes of compensation in F1 and F2, $r(48)=-0.44$ and $p<0.001$. This correlation suggests that a common underlying factor might influence compensation for perturbation of both formants.

IV. DISCUSSION

Robust and consistent compensations were observed in all instruction conditions when the first two formants of auditory feedback during speech production were perturbed in real time. The perturbations were individually determined and modified the feedback from the characteristic formant frequencies for each individual's /ɛ/ vowel to the formant values normally expected for their own /æ/ vowel. To compensate, subjects almost always produced formants that were shifted in the opposite direction from the perturbation. F1 values were shifted downward in response to a perturbation that raised the frequency, whereas the F2 frequency was increased in response to a perturbation that lowered the F2 feedback frequency. Subjects rapidly modified their formant values over the course of fewer than ten trials to reach their maximum compensation value. However, on average the maximum adjustment was only a partial compensation, amounting to less than 30% of the perturbation magnitude.

The results support the view (Pick *et al.* 1989) that the auditory concomitant of speech production is used in a control system to feed back and modify vowel production. This auditory feedback is part of the standard operational principles underlying vowel production. The present results extend this conclusion to the control of formant frequency. Our findings are consistent with data from perturbations to other facets of acoustic speech feedback (Burnett *et al.*, 1998; Hain *et al.*, 2000; Pick *et al.*, 1989) as well as with data from the study of other overlearned motor behaviors [e.g., postural responses; see Weerdesteyn *et al.*, (2008)]. Compensatory behavior is not easily eliminated by instructions.

In the study of voluntary control of the Lombard effect (Pick *et al.*, 1989), it was shown that subjects were not spontaneously able to inhibit increasing their vocal amplitude as background noise level increased, but they could be trained to do so if provided with sufficient feedback and strategies. However, even under these special conditions, the subjects appear to have gained this control using a strategy of generally reducing vocal amplitude in all background noise levels. Skilled singers have experienced extensive training of voice level and pitch control. Even with this extensive training, although the magnitude of compensatory responses to vocal frequency perturbations is reduced, compensation is not completely eliminated (Zarate and Zatorre, 2008). None of the talkers in the current study had experienced formant-shifted feedback before participating in the experiment. Thus, it is possible that through training, the magnitude of compensation might be reduced.

The compensation to altered auditory feedback observed in the present experiment might serve many useful functions. Ongoing learning might be needed to stabilize a representation of the speech motor system that is used in a predictive fashion to control rapid movements. This is consistent with proposals for the role of "internal models" as part of motor planning and control (Wolpert and Kawato, 1998). Plasticity is not a necessary feature of this proposal, but some type of corrective mechanism is. Another possible function is necessitated by the fact that the vocal tract changes in morphology continuously over the lifespan (Fitch and Giedd, 1999; Vorperian *et al.*, 2005). While the major morphological changes happen before the age of 20, many structures and cavities continue to change later in life, and therefore we must possess some mechanism that is able to modify articulatory goals. Such adaptation may be part of a more general plasticity in motor learning observed for even the simplest sensorimotor behaviors (Wolpaw, 2007) and which supports the learning of new skilled activity and recovery from injury. Finally, these auditory-vocal adjustments may serve to tune speech production to ongoing changes in background acoustic conditions to ensure intelligibility.

The functions and mechanisms of such plasticity in speech production are unclear in part because it is still not known at what level of the speech motor system such changes occur. In a review of visual-motor perturbation studies that spanned more than 50 years, Epstein (1967) identified six possible alternative sites of adaptation in those studies. Of these alternatives, five are readily adaptable to the auditory feedback context. First, the adaptive changes could

be strictly auditory. Speech sound categories have been shown to be modifiable in a variety of learning contexts [e.g., see Kraljic *et al.* (2008), Maye *et al.* (2008), Norris *et al.* (2003), van Linden and Vroomen (2007), and Davis *et al.* (2005)]. The diversity of conditions that produce such changes in speech perception (lexical status, visual speech, acoustic experience, pragmatic information, etc.) indicates an auditory speech perception system that is inherently dynamic. The relationship of this system to the auditory processes supporting speech production is less well understood with only a few demonstrations of an association between the two types of perception [e.g., see Cooper *et al.* (1976) and Newman (2003)]. Second, the adaptive changes could be strictly proprioceptive. Ostry and co-workers (Tremblay *et al.*, 2003; Nassir and Ostry, 2006) showed that jaw movements during speech adapt to changes in the dynamic force field. This learning can occur without the presence of acoustic changes in the speech and presumably involves a representation of speech movement at the proprioceptive level. Cutaneous, joint, and muscle receptors provide a rich sensory representation of articulator states. Third, the adaptive changes could be strictly motoric. The tongue muscles act synergistically during vowel production to control the shape and position of the vocal tract constriction (Perkell, 1996), and any changes in resonance properties caused by the feedback perturbation would necessitate a modification of this organization. Fourth, while these specific sensory and motor changes are possible, perhaps the most likely account is that the learning is of a more multimodal nature. One multimodal change could involve sensory-motor "recorrelation," as has been suggested for prism adaptation studies [e.g., see Epstein (1967)] and some auditory-motor speech learning studies [e.g., see Purcell and Munhall (2006)]. For the study under discussion here, this would mean learning a new mapping between vocal tract shape and speech acoustics during the perturbation phase of the experiment and then relearning the old mapping again during the return phase.

The fifth and final alternative is that the observed changes could involve "conscious correction." The present study addressed this option, and the results suggest that this possibility is unlikely; instead, formant compensation is robust under a variety of different instructional sets. We do not think that our data are evidence of a fixed response system that cannot be changed with practice or strategies. The tendency to describe compensatory behavior of the kind observed here as reflexive or automatic must be tempered by the growing recognition that even the most widely accepted examples of "reflexes" such as the tendon jerk can be modified or conditioned (Wolpaw, 2007). Rather, the present results are evidence that compensatory responses to vowel perturbation are not simply overt strategic responses to detecting manipulated feedback. The presence of aftereffects that persist beyond the perturbation is itself a strong argument that strategic compensation is an unlikely explanation for these data (Epstein, 1967). These aftereffects can be seen in Fig. 2. The formant values do not immediately return to baseline levels when normal feedback is provided.

Inference of mechanism from behavioral data is limited since the same observed behavior could be accomplished in different ways. For example, the same behavior could be accomplished by relying on different neural substrates or on the same substrate but in varying amounts. Zarate and Zatorre (2008) reported that when singers and nonsingers were instructed to compensate for pitch perturbations, both performed this task equally well. However, neuroimaging revealed that singers exhibited more activity in the anterior cingulate cortex, superior temporal sulcus, and putamen than did nonsingers. In a study of the stereotypical balance-recovery response in humans, Weerdesteyn *et al.* (2008) found that subjects could voluntarily inhibit a stepping response for balance recovery when instructed to do so. However, electromyographic records showed that both the balance-recovery trials and the trials in which the subjects were instructed to inhibit the response and fall forward had similar compensatory muscle activation patterns, with similar latencies but dramatically different activation amplitudes. Weerdersteyn *et al.* (2008) suggested that a consistent balance-recovery response is always generated but that the magnitude of the response can be regulated for different goals. Thus, a more complete understanding of the mechanisms supporting the compensatory speech response will require a more detailed physiological investigation.

In summary, the data presented here indicate that motor planning and control of vowel production must incorporate the auditory consequences of the movements as feedback. Modifications of this acoustic signal that result in an error relative to expected sensory feedback initiate compensatory behavior even when subjects are aware of the manipulation. That such compensation appears obligatory suggests that, in everyday life, compensation happens automatically in response to changing acoustic conditions contributing to optimal intelligibility even while the talker is unaware of the process.

ACKNOWLEDGMENTS

This research was supported by the National Institute of Deafness and Communicative Disorders Grant No. DC-08092 and the Natural Sciences and Engineering Research Council of Canada. Bryan Burt assisted in data collection.

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.

Brainard, M. S., and Doupe, A. J. (**2000**). "Auditory feedback in learning and maintenance of vocal behaviour," Nat. Rev. Neurosci. **1**, 31–40.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Cooper, W. E., Billings, D. and Cole, R. A. (**1976**). "Articulatory effects on speech perception: A second report," J. Phonetics **4**, 219–232.

Cowie, R., and Douglas-Cowie, E. (**1992**). *Postlingually Acquired Deafness: Speech Deterioration and the Wider Consequences* (Mouton de Gruyter, New York).

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Learning to understand noise-vocoded speech. Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.

Epstein, W. (**1967**). *Varieties of Perceptual Learning* (McGraw-Hill, New York).

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**(3), 1511–1522.
Gracco, V. L. and Abbs, J. H. (**1985**). "Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes," J. Neurophysiol. **54**, 418–432.
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**, 133–41.
Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.
Kawahara, H. (**1995**). "Transformed auditory feedback: The collection of data from 1993.1 to 1994.12 by a new set of analysis procedures," TRH-120, ATR Human Information Processing Research Laboratories, Kyoto, pp. 1–52.
Kraljic, T., Samuel, A. G., and Brennan, S. E. (**2008**). "First impressions and last resorts: How listeners adjust to speaker variability," Psychol. Sci. **19**, 332-338.
Lane, H., and Tranel, B. (**1971**). "Lombard Sign and Role of Hearing in Speech," J. Speech Hear. Res. **14**, 677–709.
Maye, J., Weiss, D. J., and Aslin, R. N. (**2008**). "Statistical phonetic learning in infants: Facilitation and feature generalization," Dev. Sci. **11**, 122–134.
Nassir, S. M., and Ostry, D. J. (**2006**). "Somatosensory precision in speech production," Curr. Biol. **16**, 1918–1923.
Newman, R. (**2003**). "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report," J. Acoust. Soc. Am. **113**, 2850–2860.
Norris, D., McQueen, J. M., and Cutler, A. (**2003**) "Perceptual learning in speech," Cogn. Psychol. **47**, 204–238.
Orfandidis, S. J. (**1988**). *Optimum Signal Processing: An Introduction* (MacMillan, New York).
Perkell, J. S. (**1996**). "Properties of the tongue help to define vowel categories: Hypotheses based on physiologically-oriented modeling," J. Phonetics **24**, 3–22.
Pick, H. L., Jr., Siegel, G. M., Fox, P. W., Garber, S. R., and Kearney, J. K. (**1989**). "Inhibiting the Lombard effect," J. Acoust. Soc. Am. **85**, 894–900.
Prochazka, A., Clarac, F., Loeb, G. E., Rothwell, J. C., and Wolpaw, J. R.. (**2000**). "What do reflex and voluntary mean? Modern views on an ancient debate," Exp. Brain Res. **130**, 417–432.
Purcell, D. W., and Munhall, K. G. (**2006**). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," J. Acoust. Soc. Am. **120**, 966–977.
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.
Smotherman, M. (**2007**). "Sensory feedback control of mammalian vocalizations," Behav. Brain Res. **182**, 315–326.
Tremblay, S., Shiller, D. M. and Ostry, D. J. (**2003**). "Somatosensory basis of speech production," Nature (London) **423**, 866–869.
Van Linden, S., and Vroomen, J. (**2007**). "Recalibration of phonetic categories by lipread speech versus lexical information," J. Exp. Psychol. Hum. Percept. Perform. **33**, 1483–1494.
Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (**2007**). "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," J. Acoust. Soc. Am. **122**, 2306–2319.
Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (**2005**). "Development of vocal tract length during early childhood: A magnetic resonance imaging study," J. Acoust. Soc. Am. **117**, 338–350.
Weerdesteyn, V., Laing, A. C., and Robinovitch, S. N. (**2008**). "Automated postural responses are modified in a functional manner by instruction," Exp. Brain Res. **186**, 571–580.
Wolpaw, J. R. (**2007**). "Spinal cord plasticity in acquisition and maintenance of motor skills," Acta Physiol. **189**, 155–169.
Wolpert, D. M., and Kawato, M. (**1998**). "Multiple paired forward and inverse models for motor control," Neural Networks **11**, 1317–1329.
Zarate, J. M., and Zatorre, R. J. (**2008**). "Experience-dependent neural substrates involved in vocal pitch regulation during singing," Neuroimage **40**, 1871–1887.

Low-dimensional models of the glottal flow incorporating viscous-inviscid interaction

Tokihiko Kaburagi
Department of Acoustic Design, Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

Yosuke Tanabe
Mechanical Engineering Research Laboratory, Hitachi Ltd., 832-2 Horiguchi, Hitachinaka-shi, Ibaragi 312-0034, Japan

(Received 6 February 2008; revised 14 October 2008; accepted 16 October 2008)

The behavior of glottal flow can, to a large extent, be characterized by development and separation of the boundary layer. The point of flow separation is known to vary during the phonatory cycle due to change in channel configuration. To take the movable nature of the separation point into account, the boundary-layer equation is solved numerically, and the values of the characteristic quantities are determined as well as the point of separation. Development of the boundary layer in general reduces the effective size of the channel, and, therefore, increases the core flow velocity, which, in turn provides the boundary condition of the boundary-layer equation. The interaction between the viscous (boundary layer) and inviscid (core flow) parts of the glottal flow is, therefore, strongly indicated. To apply this viscous-inviscid interaction, the expression of the core flow is derived for a two-dimensional flow field, and is solved jointly with the boundary-layer equation. Numerical results are shown to examine the effect of the Reynolds number and glottal configuration, with special emphasis on the comparison of flow models developed for one- and two-dimensional flow fields. Numerical results are also quantitatively compared with data obtained from flow measurement experiments. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3021436]

PACS number(s): 43.70.Bk, 43.70.Aj [BHS] Pages: 391–404

I. INTRODUCTION

The glottal volume flow serves as the primary sound source of speech. It drives the vocal tract and generates acoustic signals containing linguistic information. The periodicity of this sound source determines the fundamental frequency of speech, and is used to express stress, emotion, individuality, and so on. The spectral and temporal characteristics of the glottal sound source have been studied in relation to types of phonation (Rosenberg, 1971; Klatt and Klatt, 1990; Childers and Lee, 1991). Furthermore, to understand the physical mechanism of voice production, the behavior of glottal flow has been studied extensively using flow measurements (van den Berg *et al.*, 1957; Scherer *et al.*, 1983, 2001; Alipour and Scherer, 2001; Scherer *et al.*, 2002; Hofmans *et al.*, 2003; Triep *et al.*, 2005), computer simulations (Ishizaka and Flanagan, 1972; Broad, 1979; Pelorson *et al.*, 1994; Story and Titze, 1995; Lous *et al.*, 1998), numerical solution of the conservation equations (Zhao *et al.*, 2002; Alipour and Scherer, 2004), and theoretical studies (Howe and McGowan, 2007).

van den Berg, the pioneer in this research field, used a rectangular duct as a model of the glottis, and applied the Bernoulli relation to explain the pressure distribution along the length of the duct (van den Berg *et al.*, 1957). He also pointed out the possible role of the negative pressure in the glottis that may pull the vocal folds and contribute to closing the glottis. The basic idea of flow separation was then introduced (Ishizaka and Flanagan, 1972; Broad, 1979) using a model of free jet discharged from a nozzle (Batchelor, 2000, pp. 373–376) with which the pressure loss and glottal resistance are also determined. The shape of the flow channel in this formulation is similar to van den Berg *et al.* (1957), and the constriction (glottis) has a constant sectional area and the channel expands abruptly at the exit of the constriction where the flow separates from the wall. In the upstream region of this separation point, the flow energy is maintained and the Bernoulli relation is applicable when the frictional loss can be ignored. However, the separation of flow releases vortices into the channel, forming a jet and its wake in the downstream region. Flow energy is then lost and the Bernoulli relation is no longer valid. Instead, the momentum conservation law is used to predict the flow by assuming its reattachment to the wall. The separation point changes according to the glottal shape, and it tends to shift in the upstream direction for the divergent glottal channel (Alipour and Scherer, 2004). For a better understanding of the flow behavior, therefore, the separation point should be estimated.

To perform the simulation of phonation more accurately by taking the movable separation point into account, Pelorson *et al.* (1994) employed the boundary-layer approximation. Pelorson *et al.* (1994) argued that the fluid motion in the boundary layer can be expressed by the momentum-integral equation developed by von Kármán (Schlichting and Gersten, 1999), and that it can be solved numerically using the Pohlhausen method; however, they combined known solutions of the equation for a cylinder and a flat plate to obtain computational stability and efficiency. Change in the separation point has also been considered in the body-cover model

of the vocal fold (Story and Titze, 1995). Possible points of flow separation are very limited, however, since the shape of the vocal fold is represented by a combination of rectangles, and no solution of the boundary-layer equation is used. In contrast, Lous *et al.* (1998) expressed the separation point by a simple function of the distance between both vocal folds, each of which was modeled by concatenating three line segments. As an alternative to the Pohlhausen method, Vilain *et al.* (2004) showed the usefulness of the Thwaites method with steady and unsteady flow measurement experiments.

The results of these studies indicate that the estimation of the separation point has been the central issue in the simulation of phonation. The fact that the pressure difference between the sub- and supraglottal regions is dependent on the separation point when the volume flow rate or the flow Reynolds number is constant (Pelorson *et al.*, 1994) emphasizes its importance. In addition, if the flow is estimated for a specific subglottal pressure, the volume flow rate is approximately proportional to the channel area at the separation point. This can be shown easily [Eq. (22) of the work by Kaburagi (2008)] by ignoring the unsteady term and viscous friction loss ($a=b=0$) and assuming that the sectional area at the separation point is sufficiently smaller than the area at the supraglottal section ($N=0$). In the closing phase of the vocal fold vibration, the glottal area at the separation point is smaller than the area at the glottal outlet, since the channel shape is divergent and the separation point is located upstream of the outlet. If the separation point is unreasonably fixed at the glottal exit (Ishizaka and Flanagan, 1972; Broad, 1979), the volume flow rate will then be overestimated. This flow estimation error results in an increase in the core flow velocity and an excessive negative pressure. When the negative pressure is overestimated, the vocal folds are pulled more intensely and the glottis is closed more quickly. We suppose that this is the cause of the pitch difference in the results of voice generation simulations with and without the separation point estimation reported in Pelorson *et al.* (1994).

When the mechanical motion of the vocal fold is represented using a distributed mass (Titze, 1973) or continuum model (Berry and Titze, 1996; Gunter, 2003) and combined with a model of glottal flow, the separation point should be estimated for an arbitrary configuration of the glottis. To solve such a general problem, the Pohlhausen method mentioned above is useful. In the method, the momentum-integral equation of the boundary layer is solved on the basis of the similarity of the velocity profiles. The parameters that determine the actual velocity profile are called the shape factor, and the integral equation is transformed to an ordinal differential equation of the shape factor, which is more convenient to solve numerically. The transformed equation can be solved by simply specifying the velocity of the core flow, which is strongly dependent on the channel configuration, as its boundary condition.

When the Reynolds number is sufficiently large, the thickness of the boundary layer is thin and the core flow velocity is determined by the actual channel configuration. Based on the assumption, Tanabe *et al.* (2006) gave the boundary condition of the boundary-layer equation by the core flow velocity just along the surface of the glottal wall. Although the flow Reynolds number is typically of the order of 10^3 when the glottis is open, it decreases to only several hundred when the channel narrows (Scherer *et al.*, 1983; Alipour and Scherer, 2001). The thickness of the boundary layer should, therefore, be taken into consideration in the calculation of the core flow velocity since the boundary layer can influence the effective size of the channel. Inversely, a change in the core flow velocity affects the solution of the boundary-layer equation in the Kármán-Pohlhausen framework. This strongly suggests that the viscous (boundary layer) and inviscid (core flow) parts of the glottal flow can interact (Kalse *et al.*, 2003; Lagrée *et al.*, 2007).

To build an effective analysis method of the glottal flow that takes the movable separation point and the viscous-inviscid interaction into account, Kaburagi (2008) presented a flow model formulated using one-dimensional flow field that jointly solved the equations for the core flow and boundary layer. In this paper, the spatial dimensionality is increased to two, and the effectiveness of the new flow model is examined through numerical experiments. By assuming that the flow is quasi-steady, incompressible, and irrotational, the behavior of the core flow is governed by a Laplace equation. The method of potential analysis is then applicable, and we use the conformal mapping technique to obtain an analytic solution to the Laplace equation. The core flow velocity is expressed as a function of the effective shape of the channel, which is obtained by subtracting the displacement thickness of the boundary layer from the nominal channel space bounded by the vocal folds. The boundary-layer equation is solved on the basis of the ordinal Kármán-Pohlhausen framework. Our model is, thus, capable of predicting the flow behavior for an arbitrary configuration of the glottis, when it is smoothly shaped. In addition, expressions of the layer thickness, core flow velocity, and flow separation point are coupled; therefore, the effect of the flow Reynolds number is correctly taken into consideration along with the configuration of the glottis.

This paper is organized as follows. Section II provides mathematical explanations of the flow analysis method. Numerical results are shown in Sec. III to examine the effect of the Reynolds number and the configuration of the glottal channel on the flow behavior. In addition, the predicted results are compared with those for the one-dimensional flow model (Kaburagi, 2008) and flow measurement experiments (Scherer *et al.*, 2001, 2002), and the accuracy of the proposed model is demonstrated. Finally, the conclusions of this work are given in Sec. IV.

II. FLOW MODEL WITH INTERACTIVE BOUNDARY-LAYER PROBLEM

In this section, the interactive boundary-layer problem is presented for a two-dimensional flow field. For a high Reynolds number flow, the core flow outside the boundary layer is treated as inviscid. The viscosity is thought to be important in the vicinity of the vocal fold surface and in the wake. As such, the boundary-layer approximation permits a constitutive representation. The boundary-layer equation is solved in this study by using the Pohlhausen method, while the inter-

action between the viscous and inviscid flow regions is taken into consideration. Besides the boundary-layer approximation, the glottal flow is thought to be quasi-steady and incompressible, since the typical Strouhal number is of the order of 10^{-2} and the flow velocity is sufficiently lower than sound velocity with a Mach number less than 0.1 (Pelorson *et al.*, 1994). In addition, the glottal height is smaller than its length, indicating that the flow field can be considered two dimensional. Because the Pohlhausen method is applicable to the steady, two-dimensional momentum-integral equation of the boundary layer (Schlichting and Gersten, 1999), these assumptions are very important. Although the quasi-steady assumption is valid for the glottal flow, specifically for the core flow, it might be noted that the separation of a boundary layer is a rather slow process (Hirschberg, 1992).

Equations of the boundary layer and the core flow are explained in subsections A and B. The complex variable analysis technique is applied for the core flow (Tanabe *et al.*, 2006) from its incompressible, irrotational nature. In subsection C, it is shown how the interactive boundary-layer problem is solved for specified volume flow rate (equivalently, the flow Reynolds number) and effective quantities, such as core flow velocity, boundary-layer thickness, and separation point are obtained. If the transglottal pressure is specified, the volume flow rate needs to be estimated with the boundary-layer problem, as presented in D.

A. The boundary-layer equation

The momentum-integral equation of the boundary layer is expressed as:

$$\frac{d}{dx}\{v^2\delta_2\} + \delta_1 v\frac{d}{dx}v = \frac{\tau}{\rho}, \tag{1}$$

for two-dimensional steady flow (Schlichting and Gersten, 1999), where δ_1 is the displacement thickness, δ_2 is the momentum thickness, τ is the wall shear stress, and ρ is the density of the air. A curvilinear coordinate system is used, with the x-axis following the surface of the vocal fold (Fig. 1). $v=v(x,y)$ is the velocity of the core flow along the outer edge of the boundary layer, giving the boundary condition of the equation, where the y-axis is orthogonal to the x-axis. The characteristic quantities of the boundary layer, δ_1, δ_2, and τ, are a function of x. The flow velocity in the boundary-layer has been integrated out in Eq. (1) and is implicitly related via the definition of these characteristic quantities. Continuity requires that the internal velocity agrees with the core flow velocity at the outer edge of the boundary layer.

The Pohlhausen method can compute the solution of the boundary-layer equation efficiently assuming similarity of velocity profiles inside the layer. It can be represented using the polynomial (Pelorson *et al.*, 1994), or the solution of the single-parameter Falkner-Skan equation, i.e., the Hartree profile (Kalse *et al.*, 2003; Lagrée *et al.*, 2007). When the equation is solved, the internal velocity of the layer and the values of the characteristic quantities are determined. After that, the point of flow separation is obtained by searching for the point along the surface of the vocal folds at which the wall shear stress becomes zero.

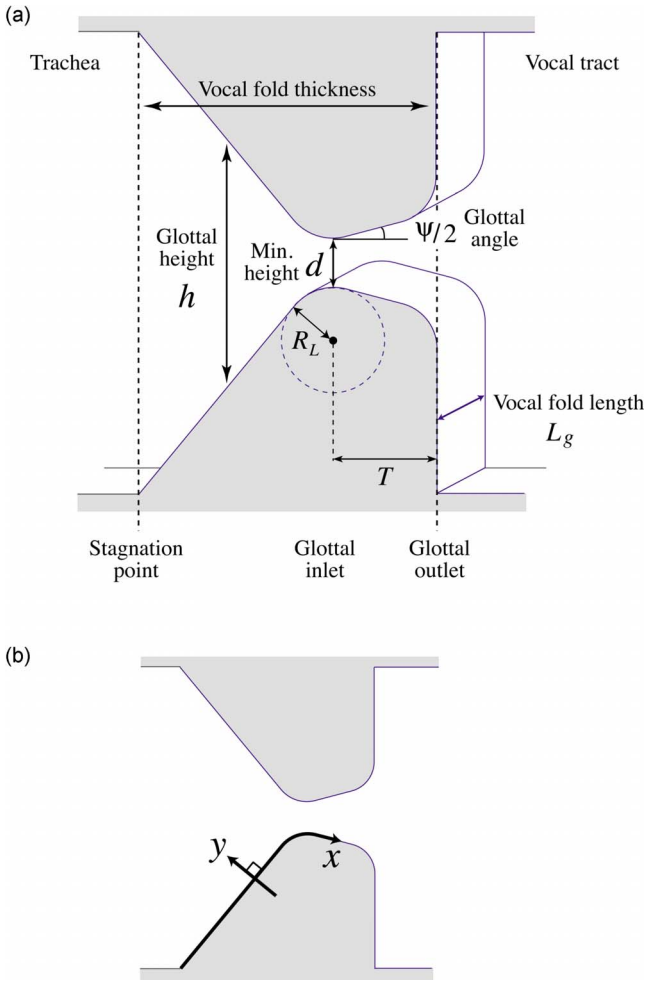


FIG. 1. (Color online) The glottal channel is represented using the model proposed by Scherer *et al.* (2001). The model is mainly controlled by two parameters representing the minimum height of the channel (d) and the glottal angle (ψ) as shown in (a). L_g, the length of the vocal fold, and T are fixed parameters. h represents the height of the channel. The x axis of the flow model is taken along the surface of the vocal fold, and the y axis perpendicular to it as shown in (b).

B. Equation of the core flow velocity

From the boundary-layer approximation, the core flow is treated as incompressible, inviscid, and irrotational. The velocity field is governed by the Laplace equation (Batchelor, 2000, Chap. 6) as $\nabla\cdot\mathbf{v}=\nabla^2\phi=0$, where $\mathbf{v}=(v_X,v_Y)$ is the flow velocity and ϕ is the velocity potential. The two-dimensional potential flow is analyzed conveniently by applying the complex variable theory (Batchelor, 2000, pp. 106–108; Tanabe *et al.*, 2006). Let $z=X+iY$ denotes a complex argument, where X and Y are the axes of an orthogonal coordinate system. Note that they are different from the x and y of the coordinate system for the glottis. In addition, the complex potential function is defined as $f=\phi+i\varphi$, where ϕ is the velocity potential and φ is the stream function, and they satisfy the Cauchy-Riemann conditions. It follows that $(df)/(dz)=v_X-iv_Y$, and the absolute flow velocity is $v=|(df)/(dz)|$.

The complex potential function is known for some specific cases. As a trivial example, $f=(U-iV)z$ represents uniform flow in the unbounded free field. Obviously, the X com-

ponent of flow velocity is $v_X=U$ and the Y component is $v_Y=V$. However, when the configuration of the glottis is specified, it is generally difficult to derive the potential function of the flow passing through that channel. We, therefore, adopt the conformal transformation technique, i.e., the Schwartz-Christoffel formula (Driscoll and Trefethen, 2002), in which the potential function is derived conveniently when the shape of the channel wall is approximated using polygonal lines (Tanabe *et al.*, 2006). In the method, the physical domain (flow region with the actual boundary shape) is connected to the canonical domain (the infinite strip in our study) via the conformal mapping. In addition, because the potential function of the uniform flow is known in the canonical domain, that for the physical domain is obtained using the mapping function. Therefore, the velocity pattern of the core flow is determined for any glottal configuration.

Let $z=X+iY$ and $\zeta=\xi+i\eta$ are complex arguments representing the coordinate systems of the physical and canonical domains, respectively. Further, g is the mapping function that connects both domains as $z=g(\zeta)$. The flow velocity in the physical domain is expressed as:

$$v_X - iv_Y = \frac{df}{dz} = \frac{df}{d\zeta}\frac{d\zeta}{dz} = \frac{df}{d\zeta}\frac{1}{g'(\zeta)}, \tag{2}$$

where $g'(\zeta)=dz/d\zeta$. The potential function of the uniform flow in the canonical domain is expressed as $f[g(\zeta)]=V\zeta$, and it is used to represent the expiratory flow from the lungs. V is a constant related to the volume flow rate of the glottal flow such that $V=U_g/(2L_g)$, where L_g is the length of the vocal folds. The flow velocity in the physical domain can then be rewritten as:

$$v_X - iv_Y = \frac{V}{g'(\zeta)}. \tag{3}$$

Figure 2 is an example of Schwartz-Christoffel mapping, where the glottal channel is formed using the model in Fig. 1(a) with the parameter values of $d=0.2$ cm and $\psi=0$ degrees. The channel is symmetrical with respect to the glottal midline, and the lower half of the channel is illustrated in the figure. It is easy to show that the flow in the canonical domain has only the ξ-axis component, since $df/d\zeta=V$ and V is real. The line $\eta=0$ corresponds to the boundary of the channel in the physical domain, and the velocity perpendicular to this boundary should be zero. The line $\eta=1$ corresponds to the axis of symmetry of the glottal channel. The flow is parallel with this line in the canonical domain, and, hence, the velocity perpendicular to it is also zero.

To derive the mapping function, the shape of the glottal channel is first approximated by polygonal lines with N vertices $(z_k, k=1,2,\ldots,N)$ in the physical domain, and the corresponding vertices ζ_k in the canonical domain are then determined.[1] Following the Schwartz-Christoffel formula, the mapping function for the infinite stripe is expressed (Driscoll and Trefethen, 2002, pp. 44–47):

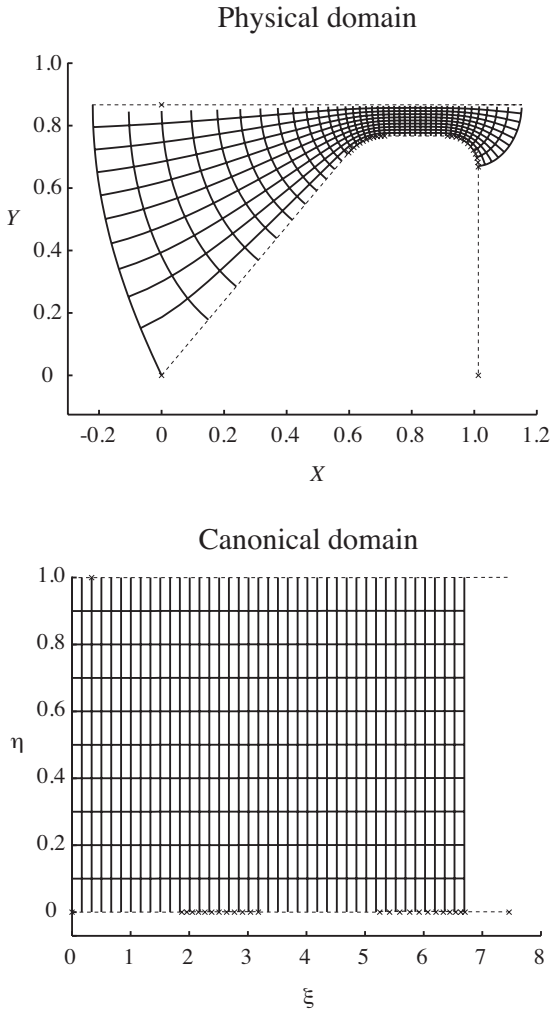


FIG. 2. An example of the glottal shape represented in the physical domain and the corresponding representation in the canonical domain. The real and imaginary axes in the physical domain are denoted as X and Y, respectively. Those in the canonical domain are ξ and η. To derive the mapping function, the glottal shape, indicated by broken lines, is first approximated by polygonal lines. Crosses show the vertices (z_k) used in this approximation. The corresponding vertices (ζ_k) are then determined in the canonical domain, and the mapping function in Eq. (4) is finally obtained. To indicate the interrelation between both domains, a grid composed of a number of horizontal and vertical lines is drawn in the canonical domain, and these lines are transformed to the physical domain using the derived mapping function.

$$g(\zeta) = C\int_0^{\zeta}\prod_{k=1}^{N}\left[\sinh\frac{\pi}{2}(\eta-\zeta_k)^{\alpha_k-1}\right]d\eta, \tag{4}$$

where $\alpha_k\pi$ is the interior angle of the kth vertex in the physical domain. This equation is different from its textbook definition (Driscoll and Trefethen, 2002, pp. 44–47) in some points. First, the term $\exp[\pi/2(\alpha_- - \alpha_+)\zeta]$ is omitted, since the divergence angles α_- and α_+ at both ends of the strip are equal. Second, the constant A is set to zero, so that the origin of the ζ-axis is mapped to that of the z-axis. The scaling factor C is set to $h/2$, half the channel height at the stagnation point and the point just downstream of the glottal outlet.[1] The flow velocity is calculated using Eq. (3) after the mapping function is derived.

C. The interactive boundary-layer problem

Next, we consider the interactive problem between core flow and the boundary layer. When a Hartree profile is assumed to represent the velocity profiles of the boundary-layer equation, Eq. (1) can be rewritten as follows (Schlichting and Gersten, 1999, Sec. 8.1):

$$\frac{dZ}{dx} + (3+2H)\frac{Z}{v_s}\frac{dv_s}{dx} = F_1, \tag{5}$$

and

$$\frac{Z}{v_s}\frac{dv_s}{dx} = \Gamma, \tag{6}$$

where

$$F_1 = \frac{2\tau\delta_2}{\rho\nu v_s}. \tag{7}$$

$v_s = v_s(x)$ is the absolute velocity of the core flow along the surface of the glottis. $Z = Z(x)$ and $\Gamma = \Gamma(x)$ are shape factors. $H = H(x)$ is the ratio of the displacement and momentum thicknesses as $H = \delta_1/\delta_2$. ν is the kinematic viscosity. From Schlichting and Gersten (1999), we obtain the relation $Z/v_s = \delta_2^2/\nu = \delta_1^2/(\nu H^2)$. In addition, from the approximation method of Kalse *et al.* (2003), it follows that $\Gamma = 1/H^2 f_1(H)$ and $F_1 = 2f_2(H)$, where $f_1(H) = -2.4\{1 - \exp[0.43(2.59 - H)]\}$ and $f_2(H) = 4/H^2 - 1/H$. Therefore, Eqs. (5) and (6) can be rewritten:

$$\frac{dZ}{dx} + \frac{3+2H}{H^2} f_1(H) = 2f_2(H), \tag{8}$$

and

$$\frac{Z}{v_s}\frac{dv_s}{dx} = \frac{1}{H^2} f_1(H). \tag{9}$$

These equations can be solved numerically, if core flow velocity v_s is given.

From the complex variable representation and conformal mapping explained in Sec. II B, v_s can be expressed:

$$v_s(x) = \left| \frac{V}{g'(\zeta)} \right| [\zeta = \zeta_s(x)], \tag{10}$$

where $\zeta_s(x)$ represents the glottal shape in the canonical domain: a straight channel. If $z(x)$ is a complex variable in the physical domain along the glottal surface, $\zeta_s(x)$ is determined as $\zeta_s(x) = g^{-1}[z(x)]$, where $\zeta = g^{-1}(z)$ is the inverse mapping of Eq. (4).[1] This constitutes the boundary-layer problem without the viscous-inviscid interaction, because the thickness of the boundary layer is ignored in the calculation of the core flow velocity.

The formulation of the interactive problem faces difficulties in the two-dimensinal case that are not present for one-dimensional flow (Kaburagi, 2008). First, the effective channel should be represented by polygonal lines to derive the mapping function. The polygonal approximation requires alignment of the vertices. This prevents a direct relationship between the configuration of the channel and the velocity of the core flow because the selection of the number of vertices and their positions is not unique. Second, the boundary-layer equation is solved by evaluating Eqs. (8) and (9) numerically with respect to the x axis to the downstream direction. This means that the boundary condition, i.e., the core flow velocity, of the downstream region does not affect the solution of the upstream region. However, Laplace's equation is solved so that the whole field is known at once, indicating that the expression of the core flow velocity does not match well with the numerical evaluation of the boundary-layer equation.

Lagrée *et al.* (2007) used a method to solve a two-dimensional interactive problem in which the equation for the core flow and that for the boundary layer are solved iteratively. Instead, a noniterative method is examined in this paper. We suppose that the effective core flow velocity v is expressed as

$$v = v_s + v_\delta = \frac{U_g}{(h - 2\delta_1)L_g} = \frac{U_g}{hL_g} + \frac{U_g}{hL_g}\frac{2\delta_1}{h - 2\delta_1}, \tag{11}$$

by following the one-dimensional treatment (Kaburagi, 2008), where v_s is the nominal flow velocity and v_δ is the change in the velocity caused by the development of the boundary layer and exclusion of the flow. The first term on the right side of Eq. (11) corresponds to the nominal velocity when the flow is approximately one dimensional. Therefore, the second term can be assumed as an expression of the velocity change for the two-dimensional case.

Following the expression in Eq. (11), the effective velocity $v(=v_s+v_\delta)$ along the glottal wall is determined as follows for the two-dimensional flow field. First, the nominal velocity v_s is calculated using Eq. (10) for the specified glottal shape. The velocity change v_δ is expressed as

$$v_\delta = \frac{U_g}{hL_g}\frac{2\delta_1}{h - 2\delta_1}, \tag{12}$$

where δ_1 is the layer thickness given as the solution for the two-dimensional boundary-layer equation. Eq. (9) is then rewritten as

$$\frac{Z}{v}\frac{dv}{dx} = \frac{Z}{(v_s + v_\delta)}\frac{d(v_s + v_\delta)}{dx} = \frac{1}{H^2} f_1(H), \tag{13}$$

and finally, the interactive boundary-layer problem is formulated as a set of equations, [Eqs. (8), (12), and (13)] with additional constraints $v = v_s + v_\delta$ and $\delta_1 = H\sqrt{\nu Z/v} = H\sqrt{\nu Z/(v_s + v_\delta)}$. If the problem is solved, the values of the characteristic quantities are then obtained as $\delta_2 = \delta_1/H$ and $\tau = \mu v f_2(H)/\delta_2$, where μ is the dynamic viscous coefficient. In addition, the flow separation point is estimated by finding the x-axis position where the wall shear stress becomes zero, indicating that $f_2(H) = 0$ and $H = 4$ at that point. Our method is computationally efficient because the interactive problem is solved just as in the one-dimensional case (Kaburagi, 2008), using the downstream marching as will be explained later. However, the expression in Eq. (12) may introduce an error; the channel height is defined vertically in terms of the axis of symmetry of the channel (Fig. 1) while the layer thickness is perpendicular to the glottal wall. Therefore, the

accuracy of the method will be checked through a number of numerical experiments.

To solve the simultaneous equations, a method similar to that for the one-dimensional flow model (Kaburagi, 2008) is used. When the nominal flow velocity is calculated from the specified volume flow rate and glottal channel, we have three unknown variables of the simultaneous equations v_δ, H, and Z. At the origin of the x-axis taken at the stagnation point (Fig. 1), the initial values of these variables are set to $v_\delta=0$, $H=2.216$, and $Z=0$. Values are then calculated for $x=\Delta x, 2\Delta x, 3\Delta x$ and so on to the downstream direction. Here, the functions to be zeroed are defined:

$$F_1(v_\delta,Z)=v_\delta-\frac{U_g}{hL_g}\frac{2\delta_1}{h-2\delta_1}, \tag{14}$$

$$F_2(v_\delta,H,Z)=\frac{Z}{v}\frac{v-v_{\Delta x}}{\Delta x}-\frac{1}{H^2}f_1(H), \tag{15}$$

and

$$F_3(v_\delta,H,Z)=\frac{Z-Z_{\Delta x}}{\Delta x}+\frac{3+2H}{H^2}f_1(H)-2f_2(H), \tag{16}$$

where $v_{\Delta x}=v(x-\Delta x)=v_s(x-\Delta x)+v_\delta(x-\Delta x)$ and $Z_{\Delta x}=Z(x-\Delta x)$. Note that F_1, F_2, and F_3 correspond to Eqs. (12), (13), and (8), respectively, and the differential operations in Eqs. (13) and (8) are replaced by the finite differences. Δx is a small quantity. The optimal solution that simultaneously satisfies $F_1=0$, $F_2=0$, and $F_3=0$ can be computed using a nonlinear programming technique such as the Newton-Raphson method (Press *et al.*, 1992).

If the unknown variables at a point of x are written in vector form as $\mathbf{s}(x)=(v_\delta,H,Z)^T$, they can be optimized iteratively such that $\mathbf{s}_{k+1}(x)=\mathbf{s}_k(x)+\Delta\mathbf{s}_k$, where k is the index of the iteration and T is the transposition. The initial value of this iteration can be set as $\mathbf{s}_0(x)=\mathbf{s}^*(x-\Delta x)$, where $\mathbf{s}^*(x-\Delta x)$ is the optimal value determined for the upstream position $x-\Delta x$. $\Delta\mathbf{s}_k$ is the solution of the linear equations $J\Delta\mathbf{s}_k=-F$, where $F=(F_1,F_2,F_3)^T$ and $J=\partial F/\partial\mathbf{s}$ is the Jacobian matrix. The Jacobian matrix may be approximated by finite differences. For example, $\partial F_1/\partial v_\delta$ is calculated as $\partial F_1/\partial v_\delta=\{F_1(v_{\delta k}+\Delta v_\delta,Z_k)-F_1(v_{\delta k},Z_k)\}/\Delta v_\delta$, where $v_{\delta k}$ and Z_k are the values at the kth iteration, respectively, and Δv_δ is a small quantity. The iteration is stopped when the sum of the absolute values of the updating quantities $|\Delta s_k(1)|+|\Delta s_k(2)|+|\Delta s_k(3)|$ is smaller than a threshold. After the values of v_δ, H, and Z are obtained, the separation point is estimated by finding the point x_s on the x axis for $H=4$. To determine the point x_s, the absolute error $|H-4|$ is calculated for $x=0,\Delta x,2\Delta x,3\Delta x$, and so on. The optimal point, at which the error is the minimum, is the estimated separation point.

D. Flow analysis from the specification of the transglottal pressure

The method in C is adequate for a flow analysis problem in which the volume flow rate or the Reynolds number is specified. On the other hand, it is useful to accommodate the method to the flow condition of transglottal pressure, since it is commonly used in flow measurement experiments (Scherer *et al.*, 2001, 2002). Here, a method is presented to estimate volume flow rate from the specification of the transglottal pressure P_d in combination with the solution of the boundary-layer problem.

When viscous friction is ignored, the pressure at the glottal outlet relative to that at the stagnation point is expressed as follows using the Bernoulli relation:

$$\Delta P=\frac{1}{2}\rho\{v_0^2-v_e^2\}, \tag{17}$$

where v_0 is the effective core flow velocity at the stagnation point, and v_e is that at the glottal outlet. Both are measured on the centerline of the glottis. This pressure difference is related to the transglottal pressure as $\Delta P=-P_d$, where P_d takes a positive value. Because the value of ΔP is determined from the solution (effective velocity) of the interactive boundary-layer problem, and the boundary-layer problem is solved for a specific value of the volume flow rate U_g, ΔP is a function of U_g and the equality constraint can be written as

$$C(U_g)=\Delta P(U_g)+P_d=0, \tag{18}$$

for the specific value of P_d. This problem can be solved by applying nonlinear programming techniques, such as the Newton-Raphson method (Press *et al.*, 1992).

Given an initial value for U_g it can be iteratively optimized with $U_g^{k+1}=U_g^k+\delta U_g^k$, where k is the index of the iteration and the incremental value is given as $\delta U_g^k=-C(U_g^k)/C'(U_g^k)$. When the flow is approximated using the one-dimensional interactive model (Kaburagi, 2008), the effective velocity on the glottal centerline is expressed as $v=(U_g)/[(h-2\delta_1)L_g]$. $C'(U_g^k)$ is then calculated as:

$$\begin{aligned}C'(U_g^k)&=\left.\frac{dC}{dU_g}\right|_{U_g=U_g^k}=\left.\frac{d}{dU_g}\{\Delta P(U_g)+P_d\}\right|_{U_g=U_g^k}\\&=\left.\frac{d}{dU_g}\left[\frac{1}{2}\rho\{v_0^2-v_e^2\}\right]\right|_{U_g=U_g^k}\\&=\frac{\rho U_g^k}{L_g^2}\left[\frac{1}{\{h(0)-2\delta_1(0)\}^2}-\frac{1}{\{h(x_e)-2\delta_1(x_e)\}^2}\right],\end{aligned} \tag{19}$$

where $\delta_1(x)$ is the solution of the boundary-layer problem when the volume flow rate is given as U_g^k. The procedure is stopped when the absolute value of δU_g^k is smaller than a threshold value. Otherwise, the volume flow rate is changed as $U_g^{k+1}=U_g^k+\delta U_g^k$ and the procedure is repeated. When the sectional area of the glottal jet is supposed to be constant for the interval between the separation point and the glottal outlet and the energy loss in the jet is ignored, the pressure and flow velocity at the separation point are maintained. The effective velocity v_e is then obtained using the value of h and δ_1 at the separation point.

III. NUMERICAL EXPERIMENTS

A. Experimental conditions

This section presents the results of numerical investigations. The shape of the flow channel is controlled using a

vocal fold model shown in Fig. 1 (Scherer *et al.*, 2001). The model has parameters representing the minimum glottal height d and the tilt angle of the vocal fold ψ. Other parameters of the model are set as $T=0.3$ cm and $L_g=1.2$ cm. The height of the trachea and the vocal tract tubes is 1.73 cm, and the sectional area of the vocal tract is set to 2.1 cm^2. To approximate the glottal shape and derive the conformal mapping, 20 vertices are placed along each curved section of the vocal fold model with equal intervals, and are connected with line segments. Two vertices are also placed at the stagnation point and the end point of the vocal fold at the glottal outlet, as well as the additional one at a point on the axis of symmetry. The air density is $\rho=1.184$ kg/m^3, and the dynamic viscous coefficient is $\mu=0.0182\times10^{-3}$ Pa s.

The boundary-layer problem is solved using the Newton-Raphson method. The value of Δx is set to 10^{-3} cm in the numerical calculation. The small value in the calculation of the Jacobian matrix, denoted by Δv_δ in Sec. II C is 10^{-3}. The iterative procedure is stopped when the absolute value of the updating quantity is smaller than 10^{-5}. With this criterion, convergence is obtained in five or six iterations for most of the experimental conditions. The same optimization method is used to estimate the volume flow rate in Sec. II D. By quitting the iterative computation when the absolute value of the updating quantity is smaller than 0.01, convergence is obtained in less than five or six iterations.

B. Effect of the Reynolds number and channel configuration

Channel configuration greatly influences the behavior of confined flow; especially in the case of the glottis, it varies from a convergent to a divergent shape during the phonation cycle. It has been revealed that the flow separates around the outlet of the glottis in the convergent or parallel shape, and moves to the entrance when the glottal angle is increased in the divergent shape (Scherer *et al.*, 2001, 2002; Alipour and Scherer, 2004). On the other hand, if the flow Reynolds number increases, the thickness of the boundary layer will be thinner and, hence, the velocity of the core flow will increase. Therefore, simulations are first performed to examine the effect of the flow Reynolds number and channel configuration on the behavior of the core flow and boundary layer. Here, the Reynolds numbers are changed at 800, 1600, 2400, and 3200. The height parameter of the vocal fold model is fixed at $d=0.08$ cm, but the glottal angle is changed from −20 to 20 deg so as to cover the angle conditions described in the literature (Scherer *et al.*, 2001; Alipour and Scherer, 2004).

Figure 3 shows computed results for each combination of the glottal angles, −20, 0, 10, and 20 deg, and the Reynolds numbers, 800, 1600, 2400, and 3200. The dotted lines indicate the actual shape of the vocal fold, and the left side of the figure corresponds to the tracheal side of the glottis and right to the outlet. The effective channel configuration is obtained by adding the displacement thickness of the boundary layer to the actual shape of the glottis, and is drawn by solid lines. The estimated point of flow separation is also shown by circles. The figure shows that the displacement thickness becomes slightly thinner when the Reynolds number increases: the effective channel becomes closer to the actual (nominal) one. Change in the separation point in terms of the glottal angle is very clear: it is located near the glottal outlet at angles of 0 and −20 deg, and shifts in the upstream direction at angles of 10 and 20 deg at which the glottis takes divergent shapes. It is interesting to see that the influence of the Reynolds number on flow separation location is most apparent at an angle of 10 deg. When it is 20 deg, however, the separation point seems to be the same for all four Reynolds numbers. The separation point is more precisely examined as a function of the glottal angle later (see Figs. 5 and 6).

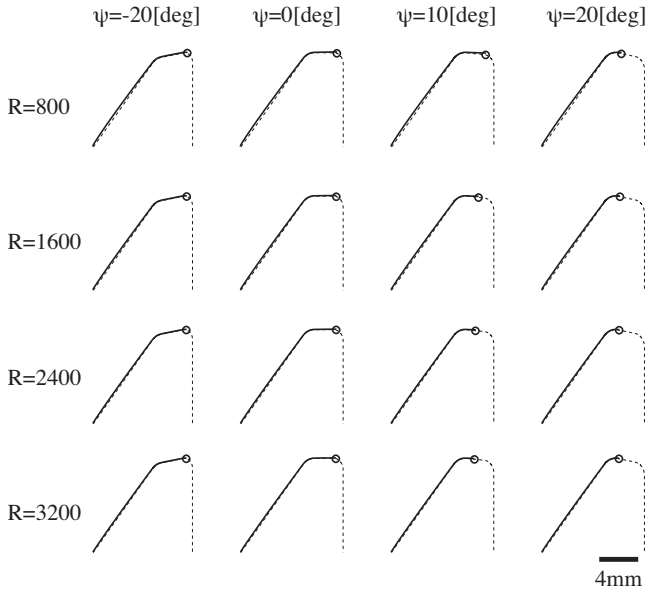


FIG. 3. Results of flow analysis based on the two-dimensional interactive boundary-layer problem. The plots are for four values of the glottal angle (ψ) and four values of the flow Reynolds number (R). The minimum glottal height (d) is 0.08 cm. The solid and broken lines represent the outlines of the displacement thickness, obtained as the solution to the boundary-layer problem, and the glottal shape, respectively. The circle on each vocal fold indicates the estimated point of flow separation.

Next, Fig. 4 shows estimation results of the core flow velocity and displacement thickness as a function of distance along the channel. The minimum height of the glottis is 0.08 cm, and the glottal angle is set to −20, 0, and 20 deg, forming convergent, parallel, and divergent channels, respectively. The effective velocity of the core flow is shown in the upper part of the figure together with the actual shape of the glottis. Here, the horizontal axis, the X axis, represents the position along the glottal centerline, which is coincident with the x axis of the one-dimensional flow model (Kaburagi, 2008). The origin of the X axis is taken at the stagnation point. The solid lines correspond to the results for the two-dimensional model presented in this paper, and broken lines to those for the one-dimensional one. In the lower part, four types of information, velocity difference (VD), displacement thickness (DT), thickness difference (TD), and displacement thickness relative to the channel height (RT), are shown. These values are calculated as $\mathrm{VD}(X)=v^1(X)-v^2(X)$, $\mathrm{DT}(X)=\delta_1^2$, $\mathrm{TD}(X)=\delta_1^1(X)-\delta_1^2(X)$, and $\mathrm{RT}(X)=100\times\delta_1^2(X)/h(X)$, where v^1 and v^2 are the effective core flow velocities and δ_1^1 and δ_1^2 are the displacement thicknesses.

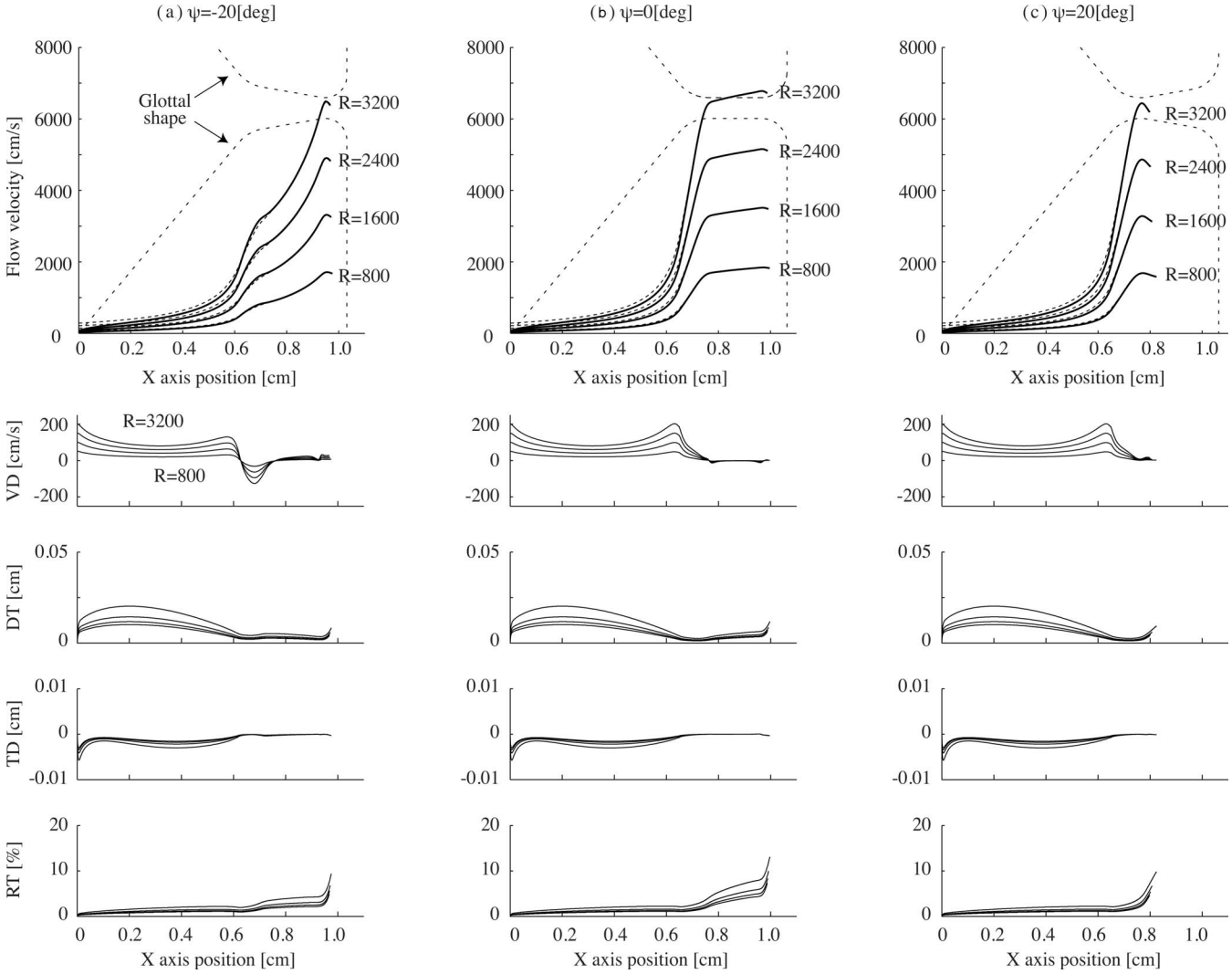


FIG. 4. Comparison of the flow analysis results between the one- and two-dimensional models for three glottal angles (ψ) and four Reynolds number values (R). In the top part of the figure, the estimated velocity of the core flow is indicated for the two-dimensional model by solid lines and the one-dimensional model by broken lines. In the lower part, each trace shows the difference of the velocity estimation between both models (VD), displacement thickness (DT), difference of the displacement thickness estimation between both models (TD), and the ratio of the displacement thickness to the height of the channel (RT). DT and RT are the results for the two-dimensional model.

The superscript indicates the spatial dimension of the flow model. For example, v^1 is the velocity calculated using the one-dimensional model. VD, DT, TD, and RT are plotted for each Reynolds number 800, 1600, 2400, and 3200.

In the interactive boundary-layer problem, the core flow velocity is determined by excluding the thickness of the boundary layer from the actual shape of the channel. For the two-dimensional model, it should be remembered that the behavior of the core flow is assumed to be approximately one-dimensional as explained in Sec. II C. Therefore, it is important to compare the results of the one- and two-dimensional models, and evaluate the validity of the assumption made for the two-dimensional one. First, the figure shows that the velocity patterns of both models agree well, and the VD and TD are small, especially in the region of about $X>0.7$. Therefore, the behavior of the core flow is inherently one dimensional in that region, when the influence of the boundary layer is taken into consideration. On the other hand, a larger difference in VD is observed for $0<X<0.7$, the section between the stagnation point and the entrance of the constriction. The figure also shows that this velocity difference is related to the increase of the absolute value of the TD, which is typically negative. When the thickness difference is negative, $\delta_1^1(X)$ is smaller than $\delta_1^2(X)$, and the effective channel in the two-dimensional model is narrower than in its one-dimensional counterpart. This agrees with the fact that the velocity difference is typically positive, indicating that $v_1^1(X)$ is larger than $v_1^2(X)$. The RT plots indicate that the thickness of the boundary layer is negligible compared with the channel height for $0<X<0.7$, and the influence of the boundary layer on core flow is also quite small. Because the behavior of the core flow is mainly determined by the actual shape of the channel, we can conclude that the velocity difference in that region is due to differences in the dimensionality of the flow field. In addition, the velocity difference is small enough compared with the peak value of the core flow velocity, implying that the velocity

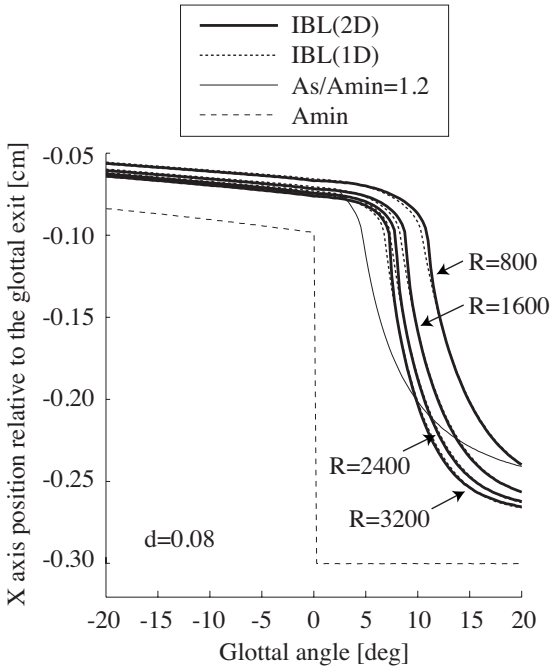


FIG. 5. The estimated point of flow separation as a function of the glottal angle. Results of the one- and two-dimensional models are, respectively, indicated by broken and solid lines. The glottal height (d) is set to 0.08 cm, and the flow Reynolds number (R) is changed. Thin lines represent the estimated separation point based on the area-based rule (Lous *et al.*, 1998) (solid) and the point of the minimum glottal area (broken).

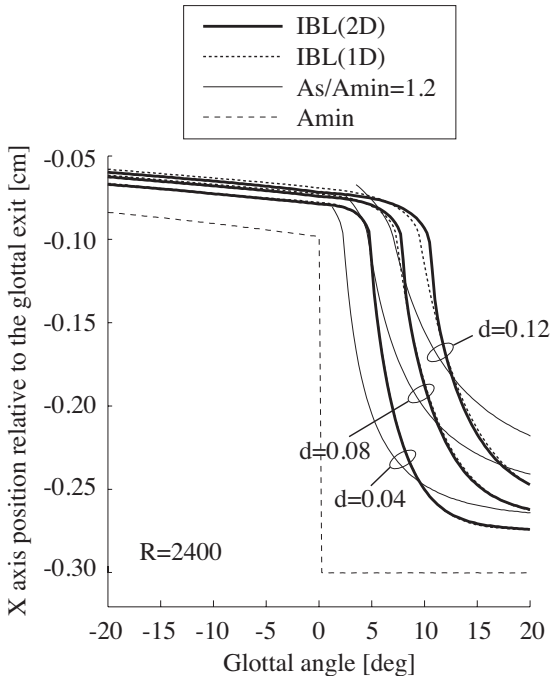


FIG. 6. The estimated point of flow separation as a function of the glottal angle. Results of the one- and two-dimensional models are, respectively, indicated by broken and solid lines. The flow Reynolds number is set to 2400, and the value of the height parameter (d) is changed. Thin lines represent the estimated separation point based on the area-based rule (Lous *et al.*, 1998) (solid) and the point of the minimum glottal area (broken).

difference between both models does not result in prediction error for the pressure distribution along the channel, which is discussed below.

Finally, Figs. 5 and 6 show the estimation results of the separation point as a function of the glottal angle. Because the thickness of the vocal fold model and the position of the glottal outlet vary according to the value of the angle parameter, the separation point is shown in the vertical axis of the figure as the X axis position relative to the glottal outlet. In Fig. 5, the minimum height is set to $d=0.08$ cm and the Reynolds number is changed as $R=800$, 1600, 2400, and 3200. In Fig. 6, on the other hand, the Reynolds number is fixed to 2400 and the height is changed to $d=0.04$, 0.08, and 0.12 cm. The solid lines indicate the results for the two-dimensional and the broken lines are for the one-dimensional model. When the glottal angle increases from −20 to 0 deg, these figures show that the separation point shifts in the upstream direction very slightly. When the glottal angle increases further, the separation point moves to the entrance of the glottis very abruptly as a function of angle. The starting angle of this sharp shift depends on the Reynolds number and the glottal height; it takes place at a smaller angle when the Reynolds number is higher or the glottal height is smaller. The estimation results of both models in general agree well. Only a slight discrepancy is observed in each experimental condition in the vicinity of the starting angle of the sharp shift.

For comparison, the separation point is also estimated using the information of the glottal area (Lous *et al.*, 1998) and the results are plotted as thin solid lines. In addition, the point corresponding to the minimum glottal area is shown by thin broken lines because it was used to determine the separation point in Story and Titze (1995). The condition $h_2=sh_1$ in the literature is replaced in this study by a condition for the glottal angle, $\psi_L=\tan^{-1} d/2$, where h_1 and h_2 are the channel heights for each mass and s is the separation constant used by Lous *et al.* (1998). When the glottal angle is greater than ψ_L, the separation point is determined so that the glottal area at that point is s times the minimum area. Estimation is not performed for angles below ψ_L because the shape of the vocal fold models is different for the outlet region where the separation takes place. It is clear that there is a significant difference between the estimation results for the above area-based methods and those for the interactive boundary-layer methods. When the glottal angle is smaller than zero, the minimum of the glottal area locates at the downstream end of the line segment of the vocal fold model. The results indicate that the flow separation takes place further downstream of the minimum area point, and it is positioned on the curved section of the model as can be seen in Fig. 3. Note that the above definition of ψ_L is derived by considering that the value for the separation constant is 1.2, the vocal fold thickness, x_2-x_1 in the literature, is 0.2 cm, and the minimum glottal height, d, agrees with h_1 for a divergent channel.

C. Quantitative evaluation of prediction accuracy

Quantitative evaluation is important to determine the validity of the flow model. The predicted results are, therefore, compared with the data obtained by the flow measurement experiments (Scherer *et al.*, 2001, 2002) in terms of the vol-

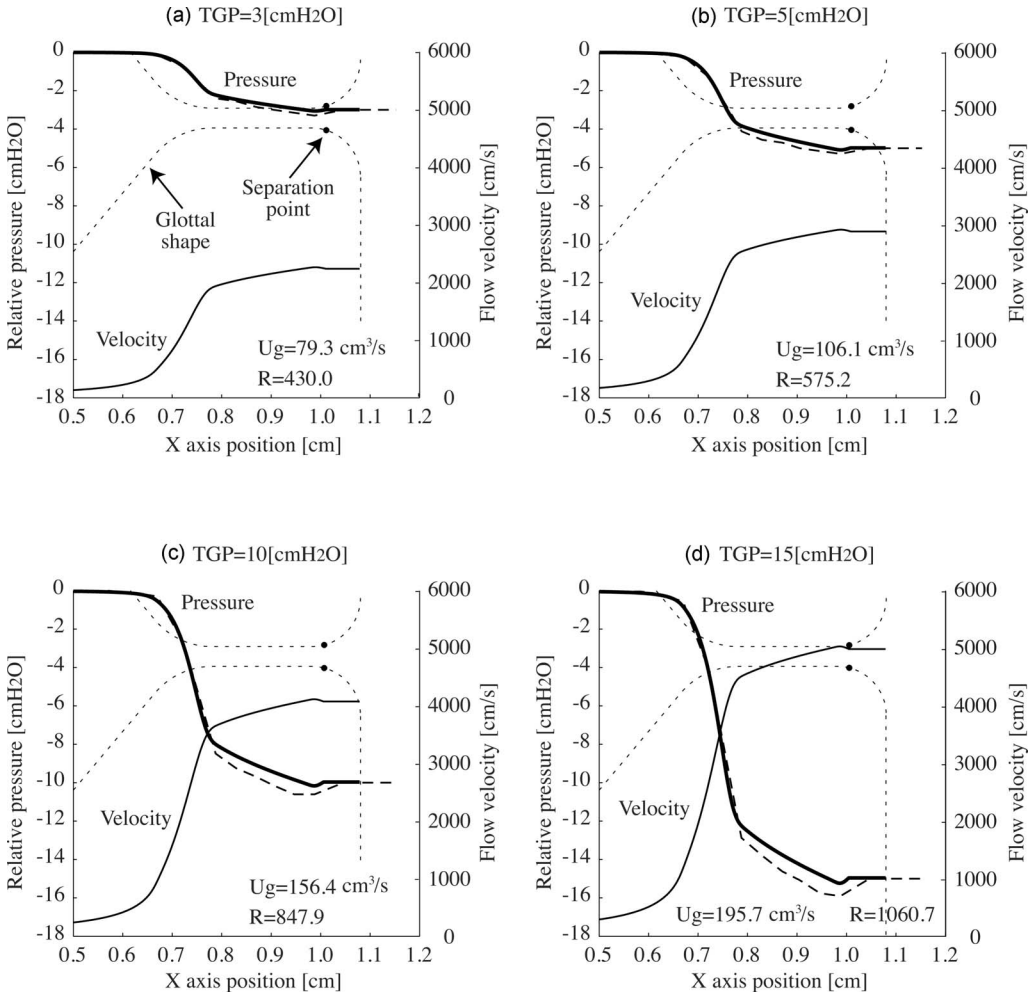


FIG. 7. Comparison of flow analysis results with flow measurements (Scherer *et al.*, 2002) for the two-dimensional model. The glottal angle is set to zero deg. TPG is the experimental value for transglottal pressure. Thick lines are the predicted (solid) and measured (broken) values of the pressure. The thin line indicates the estimated core flow velocity. Numerical values show the estimated volume flow rate (U_g) and the corresponding Reynolds number (R).

ume flow rate and the pressure distribution, where the transglottal pressure is specified as the aerodynamic condition. In these measurements, the vocal fold was fabricated using Plexiglas, and the intraglottal pressure profiles measured with a pressure transducer at 16 points on the wall of the channel under steady-state conditions. The volume flow rate was adjusted so that the desired transglottal pressure, the pressure difference between the tracheal pressure tap and the pressure tap furthest downstream, was obtained.

Figure 7 shows the results for the two-dimensional glottal flow model for glottal angles of zero deg and the height of 0.04 cm set in line with the conditions of the measurement experiments. The values for the transglottal pressure (TGP) are 3, 5, 10, and 15 cm H_2O. The profiles of the pressure and the core flow velocity are shown by thick and thin solid lines, respectively. Measurement results for the pressure distribution (Scherer *et al.*, 2002) are connected by segments of broken lines. Relative pressure, namely, the pressure drop from the trachea is plotted. This reference point was taken at the stagnation point in the flow model. The numerical values in the figure show estimated values of the volume flow rate and the Reynolds number. Figure 8 shows results of a divergent glottal channel with an angle of 10 deg (Scherer *et al.*, 2001).[2] These figures reveal that the pressure profiles of the model agree well with the measured data. From the numerical results, we also found that the difference between one- and two-dimensional models appears small, but we examine this point more precisely below. Note that we assumed here that the pressure and effective velocity at the separation point are maintained until the glottal outlet.

The accuracy of the volume flow rate estimation is summarized in Fig. 9, which compares predicted and measured values (Scherer *et al.*, 2001, 2002) for each glottal angle and transglottal pressure. Both models tend toward underestimation when ψ is zero and toward overestimation when it is 10 deg, but in general, the estimation results are fine. For example, when the transglottal pressure is 15 cm H_2O, the absolute difference is about 10 cm^3/s for $\psi=0$ and 11 cm^3/s for $\psi=10$. As shown in Table I, these errors are less than 5% of the measured value. The results show that the relative error is less than 5% for $\psi=0$, and it is about 5 to 10% for $\psi=10$.

Next, Fig. 10 shows the estimation error for pressure. Black squares correspond to the mean errors in the two-dimensional model, and gray squares to those in the one-dimensional model. The maximum error in each model is

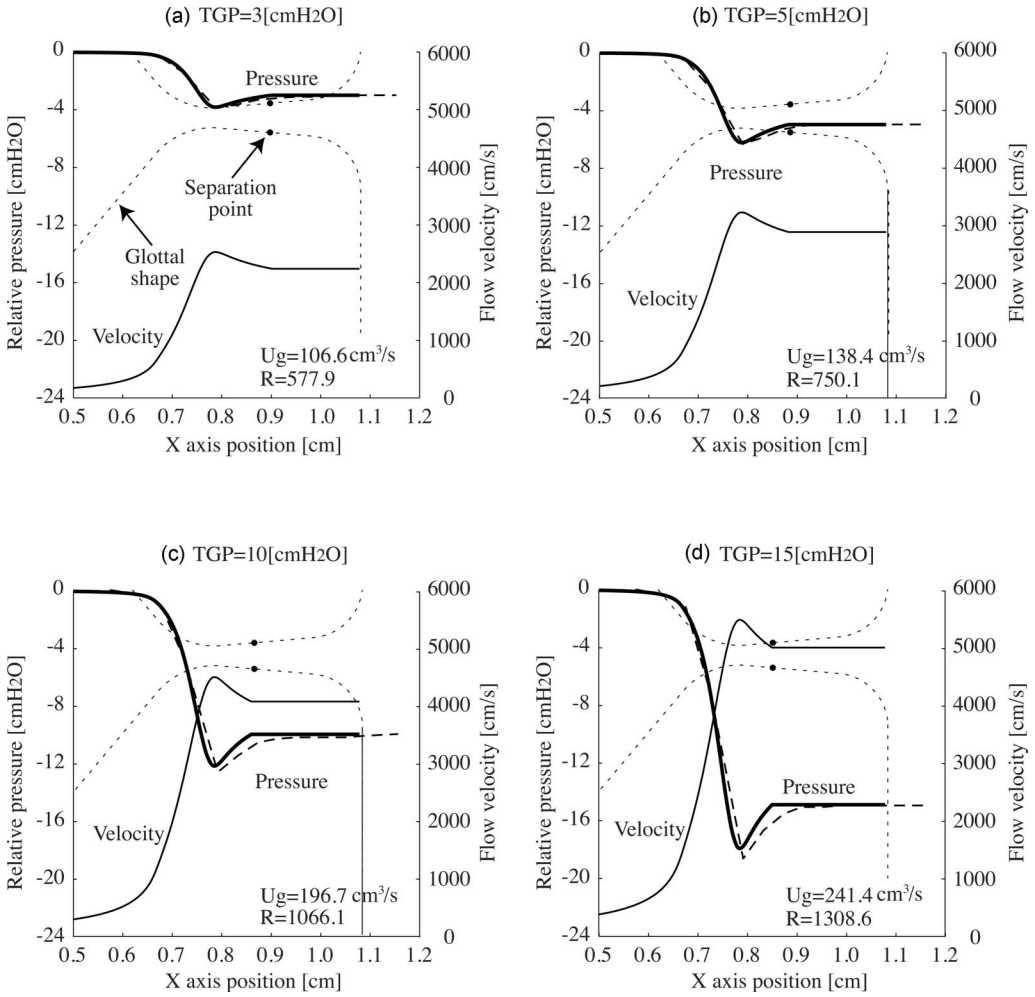


FIG. 8. Comparison of flow analysis results with flow measurements (Scherer *et al.*, 2001) for the two-dimensional model. The glottal angle is set to 10 deg.

indicated by a blank square. The mean error is calculated as $\Sigma_{k=1}^{N}|P(X_k)-P_k|/N$, where X_k is the sampling point for each error calculation and P_k is the corresponding pressure in the references (Scherer *et al.*, 2001, 2002). $P(X)$ is the predicted pressure drop from the stagnation point. The number of samples N is 14, because the reference value of the pressure tap furthest downstream is excluded. The maximum error is obtained as the maximum of the absolute errors $|P(X_k)-P_k|$ for $k=1,2,\ldots,N$. Note that the X axis defined in this study does not agree with the axis used in the above references. The figure shows that the error is, in general, small when compared to the value of the transglottal pressure, and the differences between the one- and two-dimensional models are negligible. In Table II, the numerical results are summarized as the relative error, indicating that the mean error is about 2 to 4% and the maximum error is about 5 to 7% of the transglottal pressure.

Finally, numerical analysis was performed to examine the accuracy of the model under a small Reynolds number

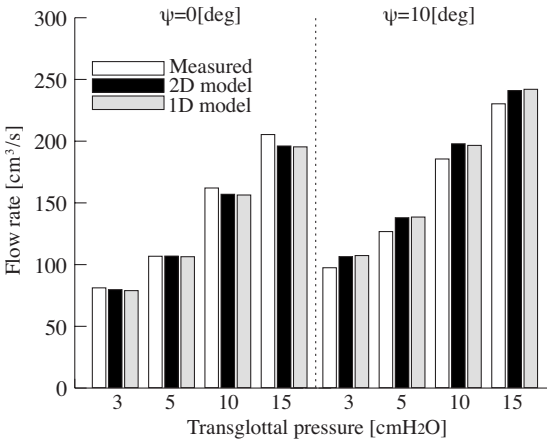


FIG. 9. Comparison of flow analysis results with flow measurements (Scherer *et al.*, 2001, 2002) in terms of volume flow rate. ψ is the glottal angle.

TABLE I. Flow rate estimation error normalized by the measured value for each condition of the glottal angle (ψ) and transglottal pressure (TGP). 2D indicates the flow model presented in this study, and 1D is the one-dimensional one (Kaburagi, 2008).

	ψ=0 (deg)		ψ=10 (deg)	
TGP (cm H_2O)	2D (%)	1D (%)	2D (%)	1D (%)
3	−1.8	−1.9	9.7	9.6
5	−0.3	−0.3	9.4	9.4
10	−3.2	−3.4	6.1	6.1
15	−4.5	−4.6	4.9	4.8

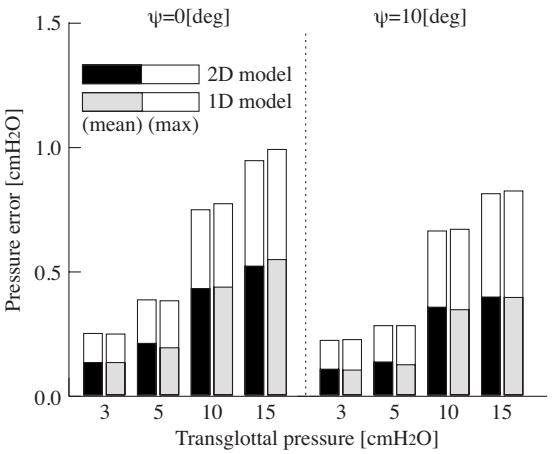


FIG. 10. Comparison of flow analysis results with flow measurements (Scherer *et al.*, 2001, 2002) in terms of the pressure profile in the glottis. ψ is the glottal angle.

condition. Following Deverge *et al.* (2003), each vocal fold was represented as a semicircle with the radius of $r=1$ cm, and its length was set to $L_g=3$ cm. The transglottal pressure was fixed to TGP=1 kPa in line with the steady flow measurement reported in the literature. The minimum distance between the semicircles d was the parameter to control the flow Reynolds number. The channel height for the sub- and supraglottal tubes was then determined as $2r+d$. The values of the two types of the Reynolds number $R\{=U_g/(\nu L_g)\}$ and $Rd/2r$ and the value of normalized pressure difference $(p_g-p_d)/\mathrm{TGP}$ were calculated in terms of the minimum channel height, where p_g and p_d are the pressure values at the throat of the channel and at the glottal exit. Note that d, R, $Rd/2r$, and $(p_g-p_d)/\mathrm{TGP}$, respectively, correspond to h_g, Re_h, $\mathrm{Re}_h h_g/L$, and $(p_g-p_d)/(p_u-p_d)$ in the literature (Deverge *et al.*, 2003). In addition, the pressure distribution was calculated by considering the Bernoulli pressure and the effect of the viscous friction loss (Pelorson *et al.*, 1994). The results shown in Table III reveal that the normalized pressure difference agrees well with the measurement data for the rounded vocal fold model (Deverge *et al.*, 2003, Fig. 4), when the Reynolds number $Rd/2r$ is greater than 4.5. When it is smaller than 1.0, it is difficult to make a comparison between the calculation and measurement results, but at the extreme of $Rd/2r\rightarrow 0$, our model seems to underestimate the normalized pressure difference.

IV. SUMMARY AND CONCLUSIONS

One-dimensional models of glottal flow employing a separation-reattachment framework (Ishizaka and Flanagan, 1972; Broad, 1979) have made great progress toward understanding the mechanism of voice production, although they rely on a strong assumption that the flow always separates at the outlet of the glottis. The idea of a movable separation point was incorporated (Pelorson *et al.*, 1994; Lous *et al.*, 1998) based on the boundary-layer assumption resulting in a further improvement of the model. This paper sets out to extend these previous works.

First of all, the interaction of the core flow and the boundary layer (Kalse *et al.*, 2003; Lagrée *et al.*, 2007; Kaburagi, 2008) is considered in addition to the boundary-layer approximation (Schlichting and Gersten, 1999). Through the interactive framework, specific quantities of the glottal flow such as the core flow velocity, pressure profile, thickness and the wall shear stress of the boundary layer, and the flow separation point are effectively analyzed. Second, an effective expression of the two-dimensional model is derived by adopting complex variable analysis and conformal mapping techniques in line with the interactive boundary-layer problem framework. Third, several numerical procedures are shown to solve the interactive boundary-layer problem in which the volume flow rate, Reynolds number, or transglottal pressure is specified as the aerodynamic condition.

To demonstrate the effectiveness of the proposed model, numerical computations were conducted for various combinations of channel configuration and the aerodynamic conditions. We then found that the predicted results are quite similar to those for the one-dimensional interactive flow model (Kaburagi, 2008). A slight disagreement is observed in the spatial pattern of the core flow velocity just downstream of the stagnation point. This may cause an error in the boundary-layer analysis, since the core flow velocity provides the boundary condition for this problem. However, it is also found that this velocity error is less than a few percent of the peak value of the core flow velocity at the constricted portion of the glottis, and does not affect the flow analysis results to a great extent. For example, when the volume flow rate and the pressure profile are estimated from the specified transglottal pressure, the predicted results of both models agree very well.

When compared with the two-dimensional model, the one-dimensional model is beneficial in that the mathematical

TABLE II. Pressure estimation error normalized by the transglottal pressure for each condition of the glottal angle (ψ) and transglottal pressure (TGP).

	$\psi=0$ (deg)				$\psi=10$ (deg)			
	2D (%)		1D (%)		2D (%)		1D (%)	
TGP (cm H_2O)	Mean	Maximum	Mean	Maximum	Mean	Maximum	Mean	Maximum
3	4.3	7.6	4.3	7.6	3.3	7.3	3.3	7.3
5	3.6	7.2	3.8	7.4	2.6	5.6	2.4	5.6
10	4.1	7.4	4.2	7.6	3.5	6.7	3.4	6.7
15	3.4	6.2	3.6	6.4	2.6	5.4	2.6	5.5

TABLE III. Flow Reynolds numbers (R and $Rd/2r$) and the relative pressure difference $[(p_g-p_d)/\mathrm{TGP}]$ estimated in terms of the minimum glottal height (d) for a rounded vocal fold model. In the right two columns, the Reynolds number ($\mathrm{Re}_h h_g/L$) and measurement data of the relative pressure difference $[(p_g-p_d)/(p_u-p_d)]$ reported in Fig. 4 of Deverge *et al.* (2003) are shown.

d (cm)	R	$Rd/2r$	$(p_g-p_d)/\mathrm{TGP}$	$\mathrm{Re}_h h_g/L$	$(p_g-p_d)/(p_u-p_d)$
0.008	139	0.55	0.16	≈0	0.74
0.01	191	0.95	0.11	1	0.01
0.02	459	4.5	−0.03	6	−0.05
0.03	734	11.0	−0.09	14	−0.09
0.04	1011	20.2	−0.11	22	−0.12
0.05	1289	32.2	−0.13	34	−0.16
0.07	1847	64.6	−0.14	64	−0.13
0.08	2126	85.0	−0.14	80	−0.15

formulation is very much simpler and requires less computation: the computational speed of the one-dimensional model is more than 10 times faster than that of the two-dimensional model when solving the interactive boundary-layer problem. In addition, the momentum preservation law is applicable according to the separation-reattachment framework (Ishizaka and Flanagan, 1972; Broad, 1979) to relate the flow regions up- and downstream of the separation point. Although flow separation releases vortices into the channel, we can avoid direct analysis of the wake, which behaves as turbulent flow. On this basis, we conclude that the one-dimensional flow model is more useful and computationally effective in analysis of glottal flow for simulating the voice production process. On the other hand, the two-dimensional model is suited to more accurate visualization of the velocity field of the core flow and boundary layer.

Our second finding from the numerical investigations is the accuracy of our flow model: the predicted results agree well with the data obtained by flow measurement experiments performed by (Scherer *et al.* 2001, 2002) in terms of the volume flow rate and pressure distribution along the channel. When the transglottal pressure varied from 3 to 15 cm H_2O, the estimation error of the volume flow rate was less than 5% for the uniform channel with a channel height of 0.04 cm. For a divergent channel with a glottal angle of 10 deg, the error was less than 10%. In relation to the pressure profile, an obvious difference was not found between the uniform and divergent channels, and the mean and maximum errors were 2 to 4% and 5 to 7% of the transglottal pressure, respectively.

Further study is needed to combine a mechanical model of the vocal fold with the proposed flow model and perform dynamic simulations. To solve the interactive boundary-layer problem and obtain an effective solution, the vocal fold should be smoothly shaped so that the boundary condition, i.e., the core flow velocity along the outer edge of the boundary layer, becomes continuous. In addition, if the incompressible nature of the vocal fold is taken into consideration, the vocal fold should be represented as a continuum (Berry and Titze, 1996; Gunter, 2003) and its motion adequately governed by an equation of motion for an elastic body. At this point, it is noteworthy that our flow model is capable of predicting wall shear stress together with the pressure to investigate its effects on the movement of the vocal folds. If the two-dimensional flow model is combined with a discrete vortex method, it is capable of representing the dynamic motion of vortices downstream from the separation point, indicating that our model has the potential to analyze flow behavior in the process of free jet formation.

ACKNOWLEDGMENTS

This research was partly supported by the Grant-in-Aid for Scientific Research from the JSPS (Grant Nos. 18500134 and 19103003).

[1]The positions of the vertices in the physical domain should be specified by the user, but those of the corresponding vertices in the canonical domain are not known. If the vertex alignment is determined for both domains, the mapping function requires numerical integration. In addition, the flow analysis procedure requires the inverse mapping: $\zeta=g^{-1}(z)$. These numerical issues can be solved by using the techniques explained in the textbook (Driscoll and Trefethen, 2002, Chap. 3). It should be mentioned here that the authors used the SC toolbox prepared for MATLAB obtainable at the web address below to develop our flow analysis programs (the last viewed date is Oct. 14, 2008): http://www.math.udel.edu/~driscoll/SC/

[2]For a divergent glottal channel, measured pressures on the two sides of the glottis had different values. The pressure data for the stalled side are used as the reference in this study.

Alipour, F., and Scherer, R. C. (**2001**). "Effects of oscillation of a mechanical hemilarynx model on mean transglottal pressure and flows," J. Acoust. Soc. Am. **110**, 1562–1569.

Alipour, F., and Scherer, R. C. (**2004**). "Flow separation in a computational oscillating vocal fold model," J. Acoust. Soc. Am. **116**, 1710–1719.

Batchelor, G. K. (**2000**). *An Introduction to Fluid Dynamics*, Cambridge Mathematical Library edition (Cambridge University Press, New York).

Berry, D. A., and Titze, I. R. (**1996**). "Normal modes in a continuum model of vocal fold tissues," J. Acoust. Soc. Am. **100**, 3345–3354.

Broad, D. J. (**1979**). "The new theories of vocal fold vibration," in *Speech and Language: Advances in Basic Research and Practice*, Vol. **2**, edited by N. J. Lass (Academic Press, New York), pp. 203–256.

Childers, D. G., and Lee, C. K. (**1991**). "Vocal quality factors: Analysis, synthesis, and perception," J. Acoust. Soc. Am. **90**, 2394–2410.

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**, 3354–3362.

Driscoll, T. A. and Trefethen, L. N. (**2002**). *Schwartz-Christoffel Mapping* (Cambridge University Press, New York).

Gunter, H. E. (**2003**). "A mechanical mode of vocal-fold collision with high spatial and temporal resolution," J. Acoust. Soc. Am. **113**, 994–1000.

Hirschberg, A. (**1992**). "Some fluid dynamic aspects of speech," Bulletin de la Communication Parlée **2**, 7–30.

Hofmans, G. C., Groot, G., Ranucci, M., Graziani, G., and Hirschberg, A. (**2003**). "Unsteady flow through *in-vitro* models of the glottis," J. Acoust.

Soc. Am. **113**, 1658–1675.
Howe, M. S., and McGowan, R. S. (**2007**). "Sound generated by aerodynamic sources near a deformable body, with application to voiced speech," J. Fluid Mech. **592**, 367–392.
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.
Kaburagi, T. (**2008**). "On the viscous-inviscid interaction of the flow passing through the glottis," Acoust. Sci. & Tech. **29**, 167–175.
Kalse, S. G. C., Bijl, H., and van Oudheusden, B. W. (**2003**). "A one-dimensional viscous-inviscid strong interaction model for flow in indented channels with separation and reattachment," J. Biomech. Eng. **125**, 355–362.
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.
Lagrée, P. Y., Van Hirtum, A., and Pelorson, X. (**2007**). "Asymmetrical effects in a 2D stenosis," Eur. J. Mech. B/Fluids **26**, 83–92.
Lous, N. J. C., Hofmans, G. C., Veldhuis, 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," Acust. Acta Acust. **84**, 1135–1150.
Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.
Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C*, 2nd Ed. (Cambridge University Press, New York).
Rosenberg, A. E. (**1971**). "Effect of glottal pulse shape on the quality of natural vowels," J. Acoust. Soc. Am. **49**, 583–590.
Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630.
Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2002**). "Intraglottal pressure distributions for a symmetric and oblique glottis with a uniform duct (L)," J. Acoust. Soc. Am. **112**, 1253–1256.
Scherer, R. C., Titze, I. R., and Curtis, J. F. (**1983**). "Pressure-flow relationships in two models of the larynx having rectangular glottal shapes," J. Acoust. Soc. Am. **73**, 668–676.
Schlichting, H., and Gersten, K. (**1999**). *Boundary-Layer Theory*, 8th Ed. (Springer Verlag, New York).
Story, B. H., and Titze, I. R. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260.
Tanabe, Y., Kaburagi, T., and Magaribuchi, K. (**2006**). "On estimating characteristic quantities of the boundary-layer for the flow through the glottis," J. Acoust. Soc. Jpn. **62**, 662–671 (in Japanese).
Titze, I. R. (**1973**). "The human vocal cords: A mathematical model, Part I," Phonetica **28**, 129–170.
Triep, M., Brücker, C., and Schröder, W. (**2005**). "High-speed PIV measurements of the flow downstream of a dynamic mechanical model of the human vocal folds," Exp. Fluids **39**, 232–245.
van den Berg, J., Zantema, J. T., and Doornenbal, P. (**1957**). "On the air resistance and the Bernoulli effect of 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.
Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (**2002**). "Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms," J. Acoust. Soc. Am. **112**, 2134–2146.

Modeling tone and intonation in Mandarin and English as a process of target approximation

Santitham Prom-on[a)]
Computer Engineering Department, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

Yi Xu[b)]
Department of Speech, Hearing and Phonetic Sciences, University College London, London WC1N 2PF, United Kingdom

Bundit Thipakorn[c)]
Computer Engineering Department, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

(Received 14 September 2007; revised 4 October 2008; accepted 6 November 2008)

This paper reports the development of a quantitative target approximation (qTA) model for generating F_0 contours of speech. The qTA model simulates the production of tone and intonation as a process of syllable-synchronized sequential target approximation [Xu, Y. (2005). "Speech melody as articulatorily implemented communicative functions," Speech Commun. **46**, 220–251]. It adopts a set of biomechanical and linguistic assumptions about the mechanisms of speech production. The communicative functions directly modeled are lexical tone in Mandarin and lexical stress in English and focus in both languages. The qTA model is evaluated by extracting function-specific model parameters from natural speech via supervised learning (automatic analysis by synthesis) and comparing the F_0 contours generated with the extracted parameters to those of natural utterances through numerical evaluation and perceptual testing. The F_0 contours generated by the qTA model with the learned parameters were very close to the natural contours in terms of root mean square error, rate of human identification of tone, and focus and judgment of naturalness by human listeners. The results demonstrate that the qTA model is both an effective tool for research on tone and intonation and a potentially effective system for automatic synthesis of tone and intonation. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3037222]

PACS number(s): 43.70.Bk, 43.72.Ja, 43.70.Fq, 43.70.Kv [AE] Pages: 405–424

I. INTRODUCTION

Quantitative modeling is one of the most rigorous means of testing our understanding of a natural phenomenon. This is particularly true if the model is built directly on assumptions that closely reflect the contested view about the mechanisms underlying the phenomenon. Modeling can also help to improve our knowledge by forcing us to make our theoretical postulations as explicit as possible. Thus for improving our understanding of human speech, quantitative modeling is also indispensable. In the present paper we report the results of an attempt to simulate tone, stress, and focus in Mandarin and English with a quantitative model that generates surface F_0 contours through the process of target approximation (TA) (Xu and Wang, 2001). Our goal is not only to develop a robust quantitative model applicable in speech technology but also to test our understanding of tone and intonation accumulated in recent years (Xu and Wang, 2001; Xu, 2005).

There have been many attempts over the past decades to build a robust model capable of simulating various prosodic phenomena through F_0 modeling (Bailly and Holm, 2005; Fujisaki, 1983; Fujisaki *et al.*, 2005; Hirst and Espesser, 1993; Kochanski and Shih, 2003; Ni *et al.*, 2006; Pierrehumbert, 1981; Taylor, 2000; van Santen and Möbius, 2000). These approaches can be divided into two general categories, namely, those that model F_0 contours directly and those that attempt to simulate the underlying mechanisms of F_0 production. Models belonging to the first category are derived mainly based on the shape of the F_0 contours, with minimal consideration about the articulatory process of F_0 production. These include the quadratic spline model (Hirst and Espesser, 1993), the Pierrehumbert model (Pierrehumbert, 1981), the tilt model (Taylor, 2000), the linear alignment model (van Santen and Möbius, 2000), the superposition of functional contours (SFC) model (Bailly and Holm, 2005), and the tone transformation model (Ni *et al.*, 2006). The quadratic spline model interpolates peaks and valleys of F_0 contours with a quadratic spline function while the Pierrehumbert model interpolates F_0 between adjacent peaks and valleys using a linear or sagging function. The tilt model generates F_0 from the tilt parameters which describe the shapes of F_0 in each intonational event, e.g., pitch accent and boundary tone. The F_0 contour of an utterance is represented by a series of these intonational events. The linear alignment model uses curve classes as templates, warping and then combining these

[a)]Electronic mail: santitham@cpe.kmutt.ac.th
[b)]Electronic mail: yi.xu@ucl.ac.uk
[c)]Electronic mail: bundit@cpe.kmutt.ac.th

curve classes superpositionally to generate F_0 contours. The SFC model simulates intonation by superpositionally combining multiple elementary contours that are functionally defined. The tone transformation model generates F_0 by specifying tone-related turning points and connecting them with a truncated second-order response function, and then imposing a global function onto the local tonal shapes. Although models in this category can represent F_0 contours at a certain level of accuracy, they do not separate surface patterns that carry intended information from those that are due to articulatory mechanisms. As a result, they have to either ignore most of the microvariations due to articulation, as done in various stylization strategies (d'Alessandro and Mertens, 1995; 't Hart *et al.*, 1990), or simulate all surface F_0 patterns directly as just described.

Models belonging to the second category are based on assumptions about the process of F_0 production. Examples in this category are the soft-template model (Kochanski and Shih, 2003; Kochanski *et al.*, 2003) and the command-response (CR) model (Fujisaki, 1983; Fujisaki *et al.*, 2005). The soft-template markup language (Stem-ML), based on a soft-template model, describes F_0 contours as resulting from realizing underlying tonal templates with different amounts of muscle forces under the physical constraint of smoothness (Kochanski *et al.*, 2003). The smoothness constraint guarantees continuous connections between adjacent templates, and the varying muscle force determines the degree to which the shape of each template is preserved in the surface F_0 under the influence of neighboring tones that are either adjacent or far away, and either preceding or following the targeted template. Stem-ML uses the optimization modeling approach for F_0 realization which requires sophisticated and complex error minimization. Even though the assumptions of Stem-ML are motivated by physical mechanisms, it requires complex mathematical translation from articulatory constraints into effort and error constraints of optimization modeling. The CR model (Fujisaki, 1983; Fujisaki *et al.*, 2005) offers, in our view, the most plausible physiological and physical simulations of the tension control mechanisms of the vocal folds compared to other models, and thus formulates the closest approximation of natural F_0 contours to date (Fujisaki *et al.*, 2005; Gu *et al.*, 2007), although certain aspects of it are still not satisfactory. The model represents surface F_0 as the logarithmic sum of phrase components and accent or tone components. The phrase components, which are assumed to be produced by the contraction of the pars obliqua of the cricothyroid (CT) muscle, represent the global contours of the utterance and are generated by a sequence of impulse response functions. The accent components, which are assumed to be produced by the contraction of the pars recta of the CT muscle, represent the local contours of the utterance and are generated by a sequence of step response functions. The CR model is thus based on the assumption that individual muscles are controlled separately. This assumption, however, is inconsistent with the finding that muscles are controlled as functional groups rather than individually (Bernstein, 1967; Gribble and Scott, 2002; Gribble *et al.*, 2003; Kelso, 1982; Zemlin, 1988). It also leads to inefficiency in model representations. For instance, to synthesize

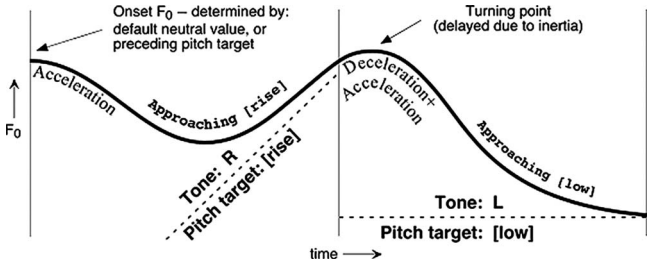


FIG. 1. Illustration of the theoretical TA model proposed in Xu and Wang (2001). In each syllable, demarcated by the vertical lines, surface F_0 (solid curve) asymptotically approaches the underlying pitch target (dashed line).

F_0 for each syllable, up to ten parameters are required. Although it is possible for the model to generate high quality resynthesis (Fujisaki *et al.*, 2005), implementing the communicative functions by summarizing these parameters together would be very complex.

The inadequacies of the existing models have motivated the proposal of the TA model (Xu and Wang, 2001). The model, as illustrated in Fig. 1, is based on the analysis of continuous acoustic data of Mandarin tone and intonation (Xu, 1997, 1998, 1999, 2001). The TA model assumes that observed F_0 contours are the outcome of implementing pitch targets which are linear functions that are either static (e.g., [low]) or dynamic (e.g., [rise]), as depicted by the dashed lines in Fig. 1. (Theoretically there may exist curvilinear targets in addition to linear ones, as explained in Xu (2005). But for the current implementation, there is no strong justification for including nonlinear targets.) The implementation of the pitch targets is synchronized with the syllable, i.e., starting at the onset of the syllable and ending at the offset of the syllable, based on evidence from acoustic data (Xu, 1998, 1999, 2001). In most cases, a tone is assumed to have only one pitch target (Xu and Wang, 2001). During each TA, the state of articulation depends not only on the discrepancy between the current state and the target but also on the final velocity and acceleration of the preceding syllable. In Fig. 1, for example, at the beginning of the second syllable, while the implementation of the [low] target has already started, F_0 is still rising due to the initial velocity and acceleration resulting from implementing the [rise] target in the first syllable. The influence of the preceding target, also known as carryover effect, would gradually decrease over time. Thus the state of articulation, as defined by F_0 height, velocity, and acceleration, is transferred from one syllable to the next at the syllable boundary. Such transfer of articulatory state is assumed to explain not only the well-known carryover assimilatory effects (Gandour *et al.*, 1994; Xu, 1997) but also the phenomenon of F_0 peak delay in both tone and nontone languages (Arvaniti and Ladd, 1995; Arvaniti *et al.*, 1998; de Jong, 1994; Ladd, 1983; Pierrehumbert and Steele, 1990; Prieto *et al.*, 1995; Silverman and Pierrehumbert, 1990; Xu, 2001).

However, the TA model alone can only describe the low-level articulatory process. A more complete framework is needed to link the articulatory mechanisms to the higher-level processes in speech. Xu (2005) proposed that TA is not limited to the realization of lexical tones but also serves as a

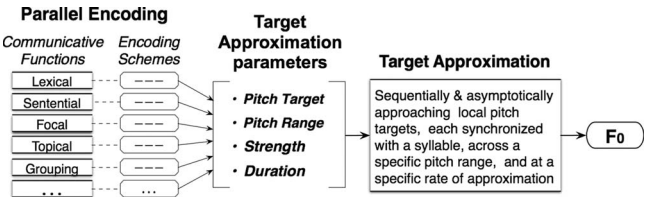


FIG. 2. A sketch of the PENTA model adapted from Xu (2005) with minor modifications. The stacked boxes on the far left represent individual communicative functions, which constitute the primary inputs to the model. They are parallel to each other with no hierarchical organizations, since the meanings they represent are independent of each other. The communicative functions are manifested through distinctive encoding schemes (second stack of boxes from left), which are either universal or language specific. The encoding schemes then specify the values of the TA parameters (middle block): pitch target, pitch range, strength, and duration, which are the control parameters of the TA model (Fig. 1) that simulates the articulatory process as syllable-synchronized sequential TA.

base mechanism for encoding other pitch related communicative meanings. That is, functions other than lexical tone also have to be encoded through the TA process, i.e., by manipulating the parameters of the process, namely, pitch range (which determines the pitch span of the targets), articulatory strength (which determines the speed of TA), and syllable duration (which assigns the amount of time for the approximation of each target). This is summarized as the parallel encoding and target approximation (PENTA) model, as illustrated in Fig. 2.

As can be seen in Fig. 2, the PENTA model assumes that speech prosody has to convey multiple communicative functions *in parallel* through individual encoding schemes. Though being abstract, these encoding schemes are always linked to specific functions. Thus it is through the TA process that a continuous link is maintained between multiple communicative functions and surface F_0 contours. Following this assumption, effective modeling of speech prosody can be achieved only if the encoding schemes of specific communicative functions are simulated. Although functional views of speech prosody are by no means new (e.g., Bailly and Holm, 2005; Bolinger, 1989; Hirschberg, 2002; Hirst, 2005; Kohler, 2005; Pierrehumbert and Hirschberg, 1990), PENTA, based on an articulatory-functional view, offers the implementational framework consisting of explicit mechanisms that directly link multiple communicative functions to the articulatory process of F_0 contour generation.

Both TA and PENTA, however, are conceptual models and thus need to be tested quantitatively. There have been previous attempts to quantify these models, including, in particular, Xu *et al.* (1999), but none of the earlier attempts has been fully satisfactory, as they are not able to simulate all the main mechanisms of target approximation revealed by previous acoustic analyses (Chen and Xu, 2006; Xu, 1997, 1998, 1999, 2001; Xu and Sun, 2002). In this paper we present the quantitative target approximation (qTA) model, which is the outcome of quantifying both TA and PENTA. We will first outline in Sec. II the basic assumptions about the articulatory mechanisms of F_0 production and how they are implemented in qTA. In Sec. III, we will explain how tone, lexical stress, and focus can be modeled based on our assumptions about the functional nature of pitch production in speech. Finally, we will report in Sec. IV the results of a series of experiments testing qTA through quantitative as well as perceptual evaluations.

II. MODELING BIOPHYSICAL MECHANISMS OF F_0 PRODUCTION

A. Background assumptions

In the following we present the background assumptions based on which the qTA model is developed. Although these assumptions are derived from recent research as just discussed, they are by no means treated as truth, but stated here so as to be explicit rather than hidden. Their validity will then be tested in the modeling experiments to be discussed subsequently. Some of the assumptions may seem to be too restrictive. But they help to reduce the degrees of freedom of the model, which is always desirable, other things being equal.

1. Vocal fold tension control as a third-order critically damped linear system

During phonation, the frequency of the vocal fold oscillation depends on the effective stiffness of the vocal folds which is directly proportional to vocal fold tension (Fujisaki, 2003; Titze, 1989) and also, though somewhat less directly, to subglottal pressure (Monsen *et al.*, 1978; Ohala, 1978; Titze, 1989). As suggested by the body-cover concept (Hirano, 1974), effective stiffness is related to the activation of the CT and the thyroarytenoid muscles, which are antagonistic to each other. The differential muscular control of the vocal folds generates two major muscle actions: increasing or decreasing their surface tension, which in turn raises or lowers F_0. Thus F_0 raising is done not only by the contraction of the CT muscle but also by the simultaneous antagonistic contraction of the thyroarytenoid muscle. Likewise, F_0 lowering is done not only by the reduction in CT contraction but also by a simultaneous thyroarytenoid contraction, which shortens the vocal folds. It is also known that the production of very low F_0 involves the extrinsic laryngeal muscles such as the sternohyoids, sternothyroids, and omohyoids (Erickson, 1976; Erickson *et al.*, 1995; Hallé, 1994). Thus the F_0 raising and lowering actions would be further aided by the contraction of these extrinsic laryngeal muscles in the lower pitch range of a speaker. A biomechanical system of the tension force controlled by antagonistic muscles that transfer energy back and forth within the system can be represented by an Nth-order linear system (Palm, 1999), where N is the number of the energy-storage elements. The qTA model is thus configured as an Nth-order linear system. Physiologically, as discussed earlier, there are at least two major antagonistic muscle forces controlling the vocal fold tension and various minor influences from the extrinsic laryngeal muscles and the subglottal pressure. Thus the model should be at least second order. This aspect of qTA is, to a large extent, similar to the CR model for F_0 control (Fujisaki, 1983; Fujisaki *et al.*, 2005) and the task dynamic (TD) model for the control of the segmental aspect of speech (Saltzman and Kelso, 1987; Saltzman and Munhall, 1989). Neverthe-

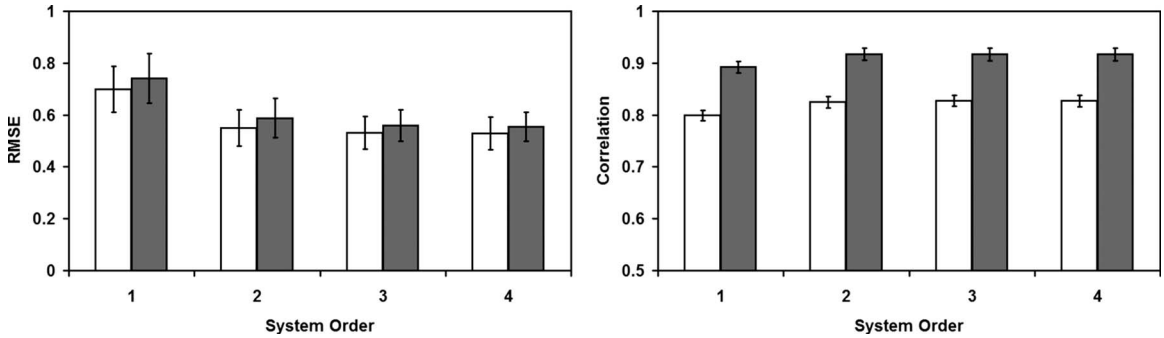


FIG. 3. Average RMSE (left) and correlation (right) of resynthesis results comparing between damping conditions and model order. White bars indicate results from the overdamped system and dark bars indicate results from the critically damped system. Vertical lines show standard errors of the mean.

less, qTA is different from CR and TD in a number of nontrivial ways, which will be highlighted as our description of qTA proceeds.

*N*th-order linear systems can have many possible damping behaviors. For instance, a second-order linear system can respond to externally imposed forces with three possible damping behaviors: underdamped, overdamped, and critically damped. Of these, critical damping can be viewed as the borderline condition between underdamping and overdamping. The responses of an underdamped second-order system have a shorter rise time compared to those of an overdamped system, but they also manifest oscillation. In contrast, the responses of an overdamped system have a longer rise time with no oscillation. In speech, F_0 movements seem to be goal oriented once the underlying functional components (e.g., tone, stress, or focus) are taken into consideration (Chen and Xu, 2006; Gandour *et al.*, 1994; Wong, 2006; Xu, 1997, 1999). In Mandarin, for example, an F_0 drop after the high tone is not for the sake of returning to a baseline or a neutral pitch register but to approach the lower pitch onset of another tone (e.g., low, rising, or the neutral tone). Similar evidence has also been found in English (Xu and Xu, 2005). This suggests that the control mechanism of F_0 regulation for speech is nonoscillatory, which means that the system is more likely to be overdamped than underdamped. An overdamped system has previously been suggested for supralaryngeal articulation (Fujisaki, 1974; Lindblom, 1983). Thus the candidate damping behaviors of the F_0 production model can be either overdamped or critically damped, although the natural process is more likely to be overdamped.

Although an overdamped linear system with an order equaling to the number of factors that antagonistically control the vocal fold tension would be a desirable choice, it is not mathematically efficient because of its complexity. Rather, a simpler model that can still generate F_0 contours with sufficient accuracy would be preferred. To determine the suitable order and damping behavior, a pilot test was conducted by resynthesizing the F_0 contours in a Mandarin corpus (Xu, 1999) and measuring the root mean square error (RMSE) and Pearson's correlation coefficient (henceforth, correlation) between the synthesized and original F_0. The detail of the corpus will be later discussed in detail in Sec. IV A while the parameter extraction method will be explained in Sec. II C. For each syllable, the model parameters, including pitch target and approximation rate, were estimated and then used to synthesize the F_0 contour. Afterward, the RMSE and correlation were calculated. The mean and confidence interval of RMSE and correlation were then derived from the mean of RMSE and correlation of each speaker. As shown in Fig. 3, although RMSE of the critically damped system is slightly lower than that of the overdamped system, their differences are not significant [$F(1,31)=0.35$, $p=0.556$, with damping condition and system order as independent factors] while the order significantly affects the error [$F(3,31)=2.80$, $p=0.047$]. In contrast, for the correlation between the synthesized and original F_0, the critically damped system is significantly better than the overdamped system [$F(1,31)=132.92$, $p<0.001$] while the effect of system order is still significant [$F(3,31)=2.90$, $p=0.042$]. Moreover, for a general *N*th-order linear system, the critically damped system has only one time-constant parameter while an overdamped system has *N* time-constant parameters (see Table I which will be explained next). Mathematically, a critically damped system is therefore simpler and possibly more accu-

TABLE I. Average rates of TA (λ, as will be explained in Sec. II B 2) and their 95% confidence intervals as functions of damping condition and model order. The number of λ values depends on the model order in an overdamped system while there is only one λ value in a critically damped system regardless of order.

Order	Overdamped (s^{-1})					Critically damped (s^{-1})
1	32.7 ± 4.3					20.5 ± 2.6
2	33.6 ± 1.3	52.3 ± 1.2				30.0 ± 2.5
3	37.9 ± 1.0	55.5 ± 1.3	68.3 ± 1.8			44.6 ± 2.9
4	35.4 ± 1.1	54.9 ± 0.9	69.1 ± 1.6	71.3 ± 1.5		49.4 ± 1.9
5	36.1 ± 1.5	59.2 ± 1.2	66.4 ± 1.5	68.6 ± 1.9	75.2 ± 1.2	50.9 ± 3.2

rate than an overdamped system. Also, as seen in Fig. 3, while the system performance is improved from second to third order in terms of both RMSE and correlation, there is little improvement from third to fourth order. Using an exponential regression to determine the lowest order at which RMSE is reduced to be within 5% from the steady state ($\lceil 3 \times \text{time constant} \rceil = \lceil 3 \times 0.8 \rceil = 3$, where $\lceil x \rceil$ is a ceiling function), we found that it is sufficient to simulate F_0 production with a third-order critically damped linear system. Table I shows the average rates of TA of both overdamped and critically damped systems with different system orders. The values in the table indicate how fast F_0 approaches the desired pitch target in terms of both slope and height when the accuracy of estimation was optimal for each system order. Last but not the least, an added advantage of a third-order system over a second-order one is that it guarantees smoothness across syllable boundaries in terms of not only F_0 level and velocity but also acceleration, which better simulates the cross-boundary state transfer assumed in the TA model.

2. Sequentiality and syllable synchronization

In the TA model, it is assumed that the most local components of tone and intonation are strictly sequential in articulation and are fully synchronized with the syllable (Xu and Wang, 2001). This assumption is based on evidence from empirical research that the F_0 movement toward an underlying target starts from syllable onset rather than from voice onset even in syllables with a voiceless initial consonant (Xu and Xu, 2003 for Mandarin; Xu and Wallace, 2004 for English; Wong and Xu, 2007 for Cantonese) and that the target approaching movement ends at syllable offset rather than at vowel offset when the syllable has a coda consonant (Xu, 1998, 2001 for Mandarin; Wong and Xu, 2007 for Cantonese). There has also been evidence that syllable-based F_0 modeling is not only feasible (Black and Hunt, 1996; Ross and Ostendorf, 1999) but also superior to accent-based modeling that ignore syllable boundaries (Sun, 2002). Furthermore, although most existing F_0 models do not hold this assumption (e.g., Fujisaki *et al.*, 2005; Kochanski and Shih, 2003; Taylor, 2002; van Santen and Möbius, 2000; but see Fujimura, 2000 for a syllable-based gestural-organization model), at least two modeling efforts have generated evidence for synchronization of the syllable with tonal units (Kochanski *et al.*, 2003) or tonal commands (Gu *et al.*, 2007). It is worth reiterating here that what is sequential and syllable synchronized is the underlying target, namely, the equivalent of the dashed lines in Fig. 1, rather than any surface F_0 event such as turning point, which is apparently delayed beyond syllable 1 in Fig. 1. This is a critical point where our assumption differs from the conclusions of many other studies that suggest variable alignment based on surface F_0 events such as peaks, valleys, and elbows (e.g., Arvaniti *et al.*, 1998; Atterer and Ladd, 2004; Chen *et al.*, 2004; Kohler, 2005).

From the perspective of modeling, sequentiality and syllable synchronization robustly reduce the degrees of freedom for the control of the TA process. That is, the implementation of a tone always starts from the onset of the syllable and ends at the offset of the syllable, as shown in Fig. 1. Nevertheless, the state of articulation, as specified by pitch level, velocity, and acceleration, is transferred from one syllable to the next at the syllable boundary. This differs from the CR model which assumes only the transfer of displacement across the executions of adjacent commands. There is no transfer of velocity or acceleration as far as we can see from the published descriptions of the model (Fujisaki, 1983, 2003; Fujisaki *et al.*, 2005).

Another important aspect of sequentiality is the assumption that all movements unidirectionally approach one target or another, with no return phases to a baseline or a neutral position. Such return phases are either obligatory or optional in other models based on a damped linear system (e.g., Fujisaki *et al.*, 2005; Saltzman and Munhall, 1989).

B. Modeling

1. Pitch target

In qTA, a pitch target is defined as the underlying goal of the local tonal or intonational component. It is a forcing function, representing the joint force of the laryngeal muscles that controls vocal fold tension. Based on the observation of the surface F_0 contours in continuous speech (Xu, 1997, 1999) and the theoretical conceptualization of the TA model (Xu and Wang, 2001), a pitch target can be represented by a simple linear equation

$$x(t) = mt + b, \tag{1}$$

where m and b denote the slope and height of the pitch target, respectively. Since the implementation of the pitch target is local to the host syllable, the time, t, is relative to the onset of the syllable.

There are two types of targets: static, e.g., [high], and dynamic, e.g., [rise]. In a static target, the slope m equals zero, while in a dynamic target m is either positive or negative. The empirical studies of Mandarin tones (Xu, 1997, 1999, 2001; Xu and Wang, 2001; Chen and Xu, 2006) suggest that high (H), low (L), and neutral (N) tones can be represented by static targets: [high], [low], and [mid], while rising (R) and falling (F) tones can be represented by dynamic targets: [rise] and [fall]. As demonstrated in Xu and Wang (2001), the slopes of the dynamic targets are essential to the dynamic tones like R and F because their F_0 variability at different speech rates cannot be adequately simulated by sequences of static targets such as [low+high] for [rise] or [high+low] for [fall]. Moreover, recent studies of English intonation suggest that an unstressed syllable may be represented by a static target [mid], while a stressed syllable may have either a static or dynamic target depending on a number of lexical and postlexical factors (Xu and Xu, 2005).

2. F_0 realization

The control of vocal fold tension in qTA is implemented through a third-order critically damped linear system. Generally, the response of the linear system consists of two parts: forced response and natural response. The forced response is the output of the system when it reaches the desired state, as is assumed in the TA model. The natural response is the

transient in the transition from the current articulatory state to the desired state represented by the pitch target. This transient effect diminishes over time. For a third-order critically damped system, the total response is in the form of

$$f_0(t) = x(t) + (c_1 + c_2 t + c_3 t^2) e^{-\lambda t}, \quad (2)$$

where the first term, $x(t)$, is the forced response which is the pitch target itself and the second term, the polynomial and the exponential, is the natural response. $f_0(t)$ is the complete form of the fundamental frequency in hertz. The model thus has three parameters, m and b which specify the pitch target and λ which represents the rate of TA. The transient coefficients c_1, c_2, and c_3 are determined jointly by the initial conditions and the target of the articulatory process. The initial conditions are the initial state of the dynamic F_0 movement, consisting of initial F_0 level, $f_0(0)$, initial velocity, $f_0'(0)$, and initial acceleration, $f_0''(0)$. Solving the systems of linear equations determined from the initial conditions, the transient coefficients can be expressed in the following formulas:

$$c_1 = f_0(0) - b, \quad (3)$$

$$c_2 = f_0'(0) + c_1\lambda - m, \quad (4)$$

$$c_3 = (f_0''(0) + 2c_2\lambda - c_1\lambda^2)/2. \quad (5)$$

Figure 4 shows the F_0 responses as each model parameter varies. Figures 4(a) and 4(b) demonstrate the F_0 contours when the first and second targets vary. Figures 4(c) and 4(d) illustrate the F_0 contours when λ varies for different combinations of dynamic and static targets. Other things being equal, F_0 approaches the target more rapidly as λ increases. As a result, the shape of the F_0 contour varies greatly with the change in the λ value. Peak delays can be clearly observed in Fig. 4(d) as λ of the second syllable varies. The approximation rate represented by λ is therefore an important parameter of the TA process, as has been suggested in previous empirical research (Chen and Xu, 2006; Xu and Xu, 2005).

C. Automatic parameter extraction

Parameter extraction was done with an automatic analysis-by-synthesis optimization algorithm, as illustrated in Fig. 5. The purpose of the automatic parameter extraction is to train the model based on available data so that we can both resynthesize the F_0 contours in the training data and synthesize novel F_0 contours. The algorithm reads the training data and parameter constraints, then automatically varies the parameter values in the specified search space, and finally adopts the parameter set with the lowest sum square error between the synthesized and original F_0 contours. The optimization is conducted one syllable at a time according to the sequentiality and syllable synchronization assumptions.

In traditional acoustic synthesis by rule, the parameters used are based on human understanding of the speech production process. These parameters can be obtained either from measurements of acoustic or articulatory data or from analysis by synthesis based on the proposed model (Klatt, 1987). For models that represent F_0 contours independent of physical mechanisms, the parameters are usually data driven based on real speech corpora. However, for models that simulate underlying mechanisms of certain processes, the analysis-by-synthesis method is typically used to estimate the model parameters (Fujisaki *et al.*, 2005; Kochanski and Shih, 2003; Mixdorff, 2000). This is because it is difficult, and sometimes impossible, to construct inverse algorithms for the production models. With an analysis-by-synthesis framework, parameter estimation can be done for any system regardless of its reversibility.

In qTA, for each syllable, the parameter λ represents its assigned level of articulatory effort. It corresponds to the rate of TA and is inversely proportional to the time constant of the approximation process. The time constant is the time relative to the onset of the current TA. Figure 6 shows computed percentages of F_0 approximation to a static target with different time constant. They are calculated by substituting t in Eq. (2) with the number that is a multiple of the time constant. A higher λ indicates that F_0 will reach the target faster. It requires more than five time constants to achieve 90% of the target. It should be noted that λ is not directly equal to the speed of F_0 movement because the same λ may result in different F_0 speed depending on b and m of the underling pitch target. In the present study, λ is allowed to vary from 0 to 120 s^{-1}, which correspond to the time to fully reach a pitch target in a very short syllable (≈70 ms). This search range is also consistent with the computed percentage of TA shown in Fig. 6.

For the other two variable parameters, namely, m and b, both specifying the pitch target, we also imposed restrictions based on our understanding of the nature of the targets as discussed earlier. For each target we specified a search space for m: zero for [high], [low], and [mid], positive for [rise], and negative for [fall]. Also for each target, we restricted the search space of b of each syllable to be within a small range (±20 Hz). The center point of this search space is around the predicted y-intercept value which is the difference between the final F_0 value and the expected excursion, i.e., target slope × syllable duration. This is because, based on previous findings, syllable offset is where surface F_0 becomes closest to the target (Xu, 1997, 1999; Xu and Wang, 2001). In cases voicing stops before the end of the syllable, the last available F_0 value and its corresponding time were used instead.

Note that these restrictions are based on previous empirical findings about tone production rather than being arbitrary, as discussed earlier in this section. Once imposed, they help to significantly reduce the possibility of the searches being stuck at local minima, thus allowing the training to be fully automatic. Table II shows the constraint violation rates and relative changes in standard deviation of estimated model parameters. The constraint violation rate is the percentage of the number of parameter estimations that fall outside the ranges of parameter constraints. The relative change in standard deviation is calculated as the relative difference between the standard deviation obtained with and without parameter constraints. The results in Table II were derived from the Mandarin dataset which will be discussed later in Sec. IV A. As shown in Table II, the most influential con-

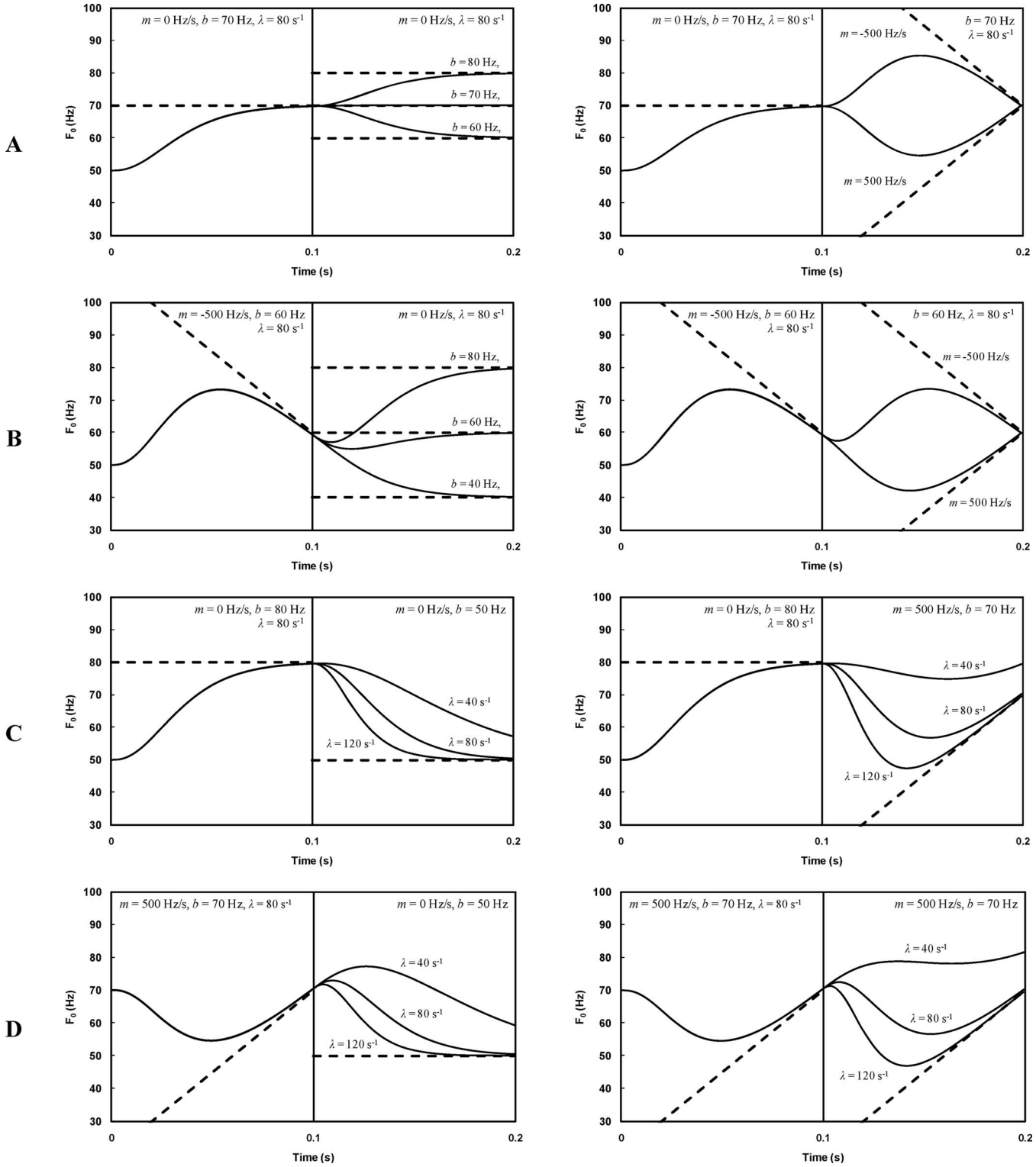


FIG. 4. Examples of F_0 contours generated by the qTA model with varying values of m, b, and λ. The dashed lines indicate the underlying pitch targets which are linear functions of m and b. The vertical lines show the syllable boundaries through which the articulatory state propagates.

straint is the m-constraint because discarding it results in the highest constraint violation rate (52.35%). Most of the m violations occurred in the H and L tones, for which the constraint $m=0$ is extremely easy to violate, as hardly any surface F_0 contour is fully flat by the end of the syllable with those tones. The right half of Table II shows that removing the constraint on m also affects the variation in λ (81.22% greater standard deviation), while the b-constraint moderately affects variations in both b and λ. The least influential is the λ-constraint which mainly affects the variation in λ itself and affects those of other parameters only slightly. Interestingly, the only parameter that has been heavily affected by constraints other than its own is λ. This indicates that the estimation of λ is quite dependent on the estimation of the pitch target parameters.

An example of the F_0 contour synthesized with param-

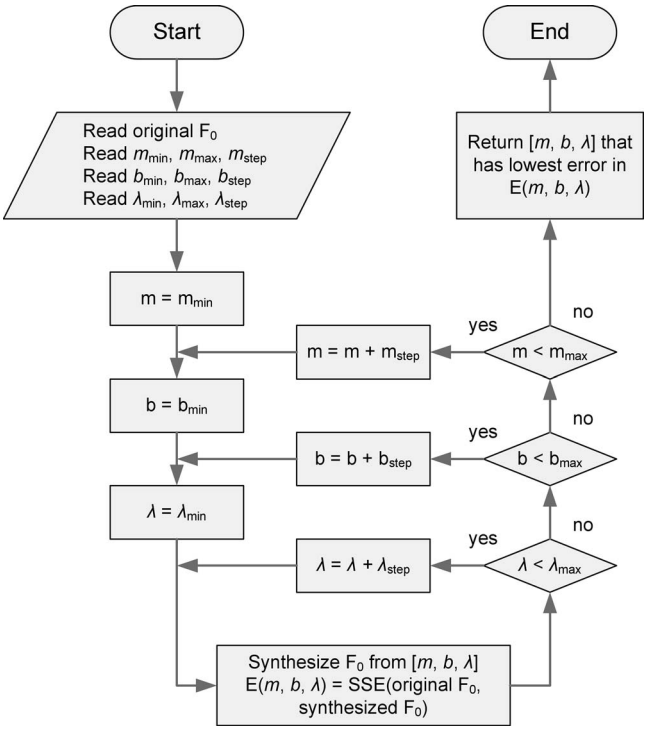


FIG. 5. A flowchart of automatic parameter extraction of qTA model. The algorithm optimizes for the suitable model parameters that, when implemented by the qTA model, generate F_0 contours that closely approximate the original ones.

eters resulting from the automatic extraction is shown in Fig. 7. The synthesized F_0 closely approximates the original one, as will be later shown in the low average RMSE in the testing on all speakers.

III. MODELING TONE AND FOCUS AS COMMUNICATIVE FUNCTIONS

As suggested by the PENTA model discussed in Sec. I, speech prosody conveys multiple communicative functions in parallel. Each function is represented by a unique encoding scheme, specified in terms of one or more of the TA parameters: pitch target, pitch range, strength, and duration, as illustrated in Fig. 2. Effective quantitative modeling of

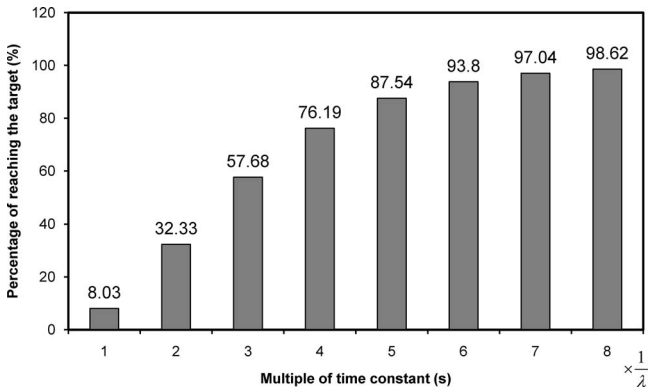


FIG. 6. Percentage of F_0 approximation to a static target as a function of time constant. These percentages were calculated by substituting t in Eq. (2) with different multiples of the time constant.

TABLE II. Constraint violation rates and relative changes in standard deviation when removing each parameter constraint in the automatic parameter extraction.

Constraint on	Constraint violation rate (%)	Relative change in standard deviation (%)		
		$\Delta\sigma_m$	$\Delta\sigma_b$	$\Delta\sigma_\lambda$
m	52.35	200.82	−0.67	81.22
b	17.15	−4.43	13.75	10.84
λ	7.21	1.74	−0.71	20.22

speech prosody, based on this view, can be achieved only if individual communicative functions are simulated.

To formalize encoding schemes of the communicative functions, we define the *parametric vector* of the jth syllable as a set of qTA model parameters, $\mathbf{p}_j=[m_j, b_j, \lambda_j]^T$. The F_0 of the sentence can be formed by consecutively executing a series of parametric vectors. A *prosodic vector* of length N represents the prosody of the sentence with N syllables. It can be expressed in the form of a series of N parametric vectors, $\mathbf{s}=\{\mathbf{p}_1, \mathbf{p}_2 \cdots \mathbf{p}_N\}$. This vector is the input to the TA process. It is possible for communicative functions to be represented by either parametric vectors or prosodic vectors, depending on the temporal domain of the function.

Three communicative functions are directly modeled in the present study, lexical tone in Mandarin, lexical stress in English, and focus in both languages, as explained next. Two other communicative functions, namely, new topic and sentence modality (Swerts, 1997; Lehiste, 1975; Liu and Xu, 2005; Wang and Xu, 2006), are only indirectly modeled as positional dependencies of the functions. Figure 8 is a block diagram of qTA and the communicative functions implemented in the present study. The input communicative functions are functional specifications which can be tone/stress symbols for the tone/stress function or focus position for the focus function. The generated parametric vector from the tone/stress function is additively combined with the focus adjustment vector from the focus function. This combination results in the focus-adjusted parametric vector which is the output of the parallel encoding process and also the input to the TA process.

A. Lexical tone and lexical stress

For Mandarin, the most local functional components are the lexical tones. Given a symbolic input tone x, the tone function generates the parametric vector $\mathbf{p}$:

$$\mathbf{p} = \text{tone}(x), \tag{6}$$

where the tone category x is N, H, R, L, or F. The tone function returns a parametric vector for tone x. Thus each tone requires one parametric vector to represent it. During training, the parameter values of a tone are derived by averaging the parametric vectors extracted from all individual occurrences of that tone.

For English, the most local pitch component is lexical stress. Unlike in Mandarin, however, pitch values related to lexical stress in English are assigned postlexically (Ladd, 1996; Liu and Xu, 2005; Pierrehumbert, 1980; Xu and Xu, 2005). Xu and Xu (2005) showed that unstressed syllables in

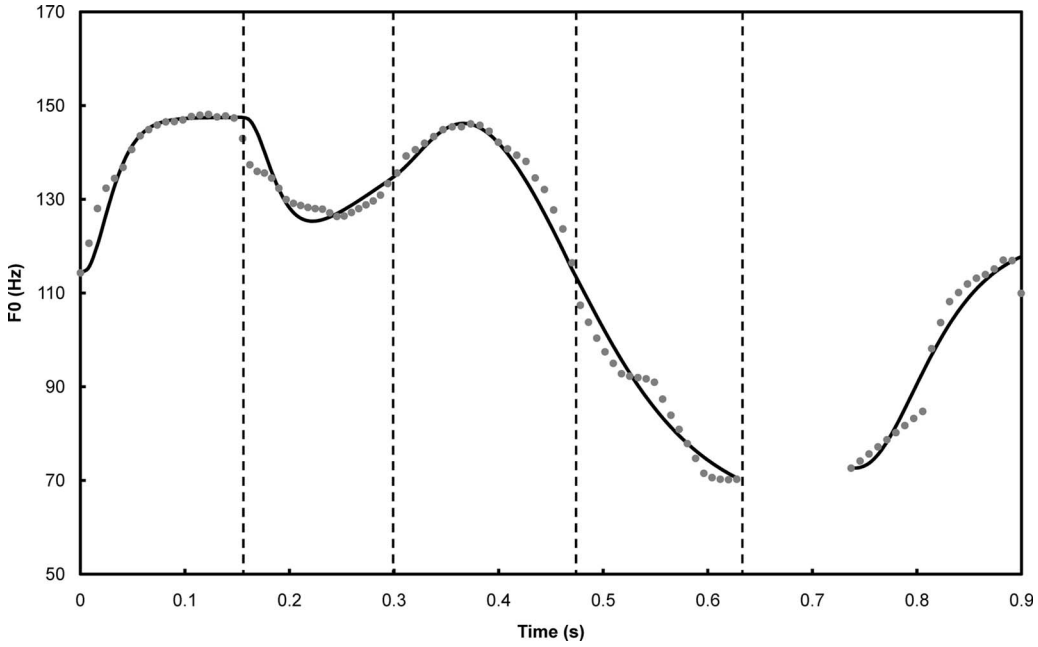


FIG. 7. An example of resynthesized F_0 for the tone sequence HRFLH. The solid line represents the synthesized F_0 contour while the gray dotted line indicates the original F_0 contour. The dashed vertical lines show the syllable boundaries. The discontinuity of F_0 at the beginning of the last syllable is due to the voiceless initial stop consonant [t] in the last syllable.

English, just like the neutral tone in Mandarin (Chen and Xu, 2006), are assigned specific pitch targets rather than being targetless. Thus for English, every syllable is assigned a pitch target regardless of its stress status. The functional representation of lexical stress is

$$\mathbf{p} = \text{stress}(x), \tag{7}$$

where x represents the stress category which is either *stressed* or *unstressed*.

B. Focus

Focus is a discourse function to highlight a particular piece of information against the background information (Bolinger, 1986; Gussenhoven, 2007; van Heuven, 1994; Ladd, 1996; Xu and Xu, 2005). In many languages, including English and Mandarin, focus expands the pitch range of the focused syllable(s) and compresses the pitch range of the postfocus syllable(s) (Botinis *et al.*, 2000; Hasegawa and Hata, 1992; Krahmer and Swerts, 2001; Mixdorff, 2004; Rump and Collier, 1996; Xu, 1999; Xu and Xu, 2005). Based on these findings, syllables in each sentence can be grouped into maximally four regions: prefocus, on focus, postfocus, and final focus, whichever is applicable. Little pitch range adjustment occurs in the prefocus region. Pitch range of the on-focus region is expanded, while that of the postfocus region, which includes all syllables up to the end of the sentence, is compressed. Final focus is treated separately because it has been found to have a pitch range larger than that of neutral focus but smaller than that of nonfinal focus in both Mandarin and English (Xu, 1999; Xu and Xu, 2005). For a sentence with no narrow focus, its entirety is treated as prefocus.

Computationally, focus is treated as an adjustment func-

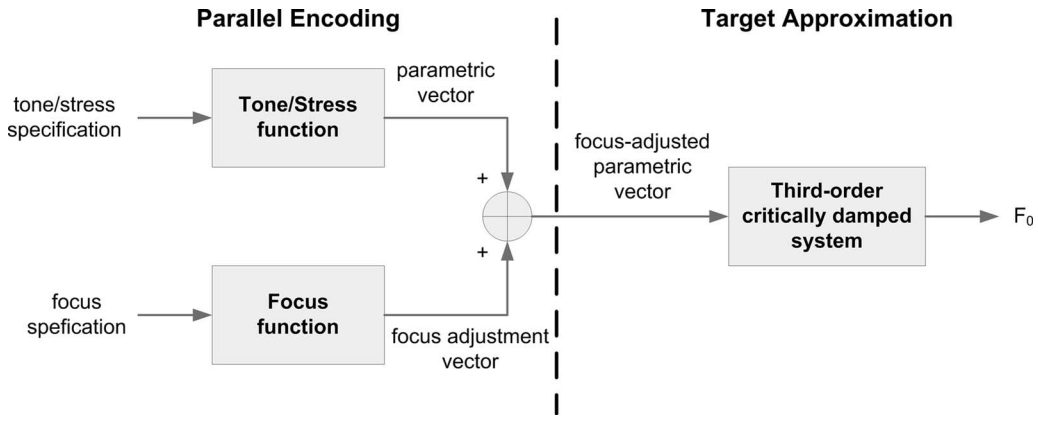


FIG. 8. A block diagram of the qTA model as the quantitative counterpart of the PENTA model. The dashed vertical line separates the parallel encoding, which mainly generates and manipulates the parametric vectors, from the TA, which implements the parametric vectors and synthesizes F_0.

tion. It maps the given prosodic vector **s** with length N and focus position **K** to the output prosodic vector $\hat{\mathbf{s}}$.

$$\hat{\mathbf{s}} = \text{focus}(\mathbf{s}, \mathbf{K}), \tag{8}$$

where **K** is a set of positive integers including one but less than or equal to N. Each element in the set **K** indicates the position of the syllables of the word under focus. If the focused word is monosyllabic, **K** has only one element. Focus encoding is done by adjusting the input prosodic vector according to the trained focus adjustment parameters. For instance, the implementation of medial focus at the Mth syllable ($\mathbf{K}=\{M\}$) would be

$$\hat{\mathbf{s}} = \{\mathbf{p}_1, \mathbf{p}_2, \ \ldots, \mathbf{p}_M + \Delta\mathbf{p}_{\text{on}}, \mathbf{p}_{M+1} + \Delta\mathbf{p}_{\text{post}}, \ \ldots, \mathbf{p}_N + \Delta\mathbf{p}_{\text{post}}\}, \tag{9}$$

where $\Delta\mathbf{p}_{\text{on}}$ and $\Delta\mathbf{p}_{\text{post}}$ denote the focus adjustment vectors of on- and postfocus regions, respectively. Here, no adjustment is made before focus. The on-focus parametric vector is changed by the on-focus adjustment which expands the pitch range of the pitch target. The remaining parametric vectors are modified by the postfocus adjustment which compresses and lowers the pitch range of the pitch targets. Similarly, the implementation of final focus at the last syllable ($\mathbf{K}=\{N\}$) would be

$$\hat{\mathbf{s}} = \{\mathbf{p}_1, \mathbf{p}_2, \ \ldots, \mathbf{p}_N + \Delta\mathbf{p}_{\text{final}}\}, \tag{10}$$

where $\Delta\mathbf{p}_{\text{final}}$ denotes the focus adjustment vectors of final-focus regions. Only the parametric vector of the last syllable is adjusted in this case. If the focused word consists of two syllables, the implementation of the initial focus of the first two syllables ($\mathbf{K}=\{1,2\}$) would be

$$\hat{\mathbf{s}} = \{\mathbf{p}_1 + \Delta\mathbf{p}_{\text{on}}, \mathbf{p}_2 + \Delta\mathbf{p}_{\text{on}}, \mathbf{p}_3 + \Delta\mathbf{p}_{\text{post}}, \ \ldots, \mathbf{p}_N + \Delta\mathbf{p}_{\text{post}}\}. \tag{11}$$

In training the focus function, the extracted parametric vectors from the training corpus are divided into four focus regions, pre-, on-, post-, and final focus. The parametric vectors in the prefocus regions are averaged together according to tone or lexical stress and treated as the reference parametric vector for further calculation in other focus regions. For other focus regions, the focus adjustment vectors are calculated by averaging the differences between the parametric vectors in on-, post-, or final-focus regions and the reference parametric vectors, respectively.

$$\bar{\mathbf{p}} = \frac{1}{N}\sum_i \mathbf{p}_i, \quad \mathbf{p}_i \in \text{prefocus set}, \quad N = \text{total number of } \mathbf{p}_i \text{ in prefocus set}, \tag{12}$$

$$\Delta\mathbf{p}_{\text{floc}} = \frac{1}{M}\sum_j (\mathbf{p}_j - \bar{\mathbf{p}}), \quad \mathbf{p}_j \in \text{floc focus set}, \quad M = \text{total number of } \mathbf{p}_j \text{ in floc focus set}, \tag{13}$$

where floc denotes three possible focus locations including on, post, and final. Thus, for example, the total number of parametric vectors required for tone and focus representation in Mandarin is 16:4 prefocus parametric vectors and 12 focus adjustment vectors. For English, the total number of parametric vectors is 8:2 prefocus parametric vectors and 6 focus adjustment vectors.

IV. EXPERIMENTAL EVALUATION

A. Corpora

Two datasets were used in the experiments for testing the ability of the qTA model to simulate tone and intonation in Mandarin and English. The Mandarin dataset was originally collected for a study of tone and focus (Xu, 1999). It consists of 3840 five-syllable utterances recorded by four male and four female Mandarin speakers. Figure 9 illustrates the cross-speaker and cross-gender variations in F_0 contours in the tone sequence HRFHH produced by eight speakers in different focus conditions. The difference between genders can be easily observed in the two nonoverlapping clusters in each panel. Within-gender variability can also be clearly seen in the spread of the contours within each cluster.

The words in the sentences and the corresponding tones in this dataset are shown in Table III. In each utterance, the first two and last two syllables are disyllabic words while the third syllable is a monosyllabic word. The first and last syllables in each sentence always have the H tone while the tones of the other syllables vary depending on the position: H, R, L, or F in the second syllable, H, R, or F in the third syllable, and H or L in the fourth syllable. Since the dataset was originally designed for studying tone and focus, each sentence has four focus conditions: no focus, initial focus, medial focus, and final focus. Thus, there are totally 96 variations in tone and focus.

The English dataset was originally collected for a study of the effect of focus on English intonation (Xu and Xu, 2005). It consists of 1176 short declarative utterances recorded by four male and four female American English speakers. A list of the sentences in this dataset is shown in Table IV. Each sentence is said in one of four focus conditions: no focus, sentence-initial focus, sentence-medial focus, and sentence-final focus, and also in one of two speaking rates: normal and fast. For each focus condition and speaking rate, the sentences were repeated seven times. The dataset is further divided into three sentence groups, which differ in the position of the target word. In total, there are 42 combinations of word, focus condition, and speaking rate.

Figure 10 shows examples of F_0 extracted from both datasets and then averaged across speakers. The top left panel shows examples of Mandarin with tonal variations at the second syllable. The bottom left panel shows the effects of focus variation on F_0 when the tones are fixed. The top right and bottom right panels show examples from two different sentences in English with different types of focus.

B. Parameter analysis

Using the automatic analysis-by-synthesis optimization algorithm described in Sec. II C, we extracted the prosodic vectors from each sentence in the two datasets. After that, we derived function-specific parametric vectors, consisting of the model parameters m, b, and λ, for different communicative functions. For Mandarin, parameters for the tone and

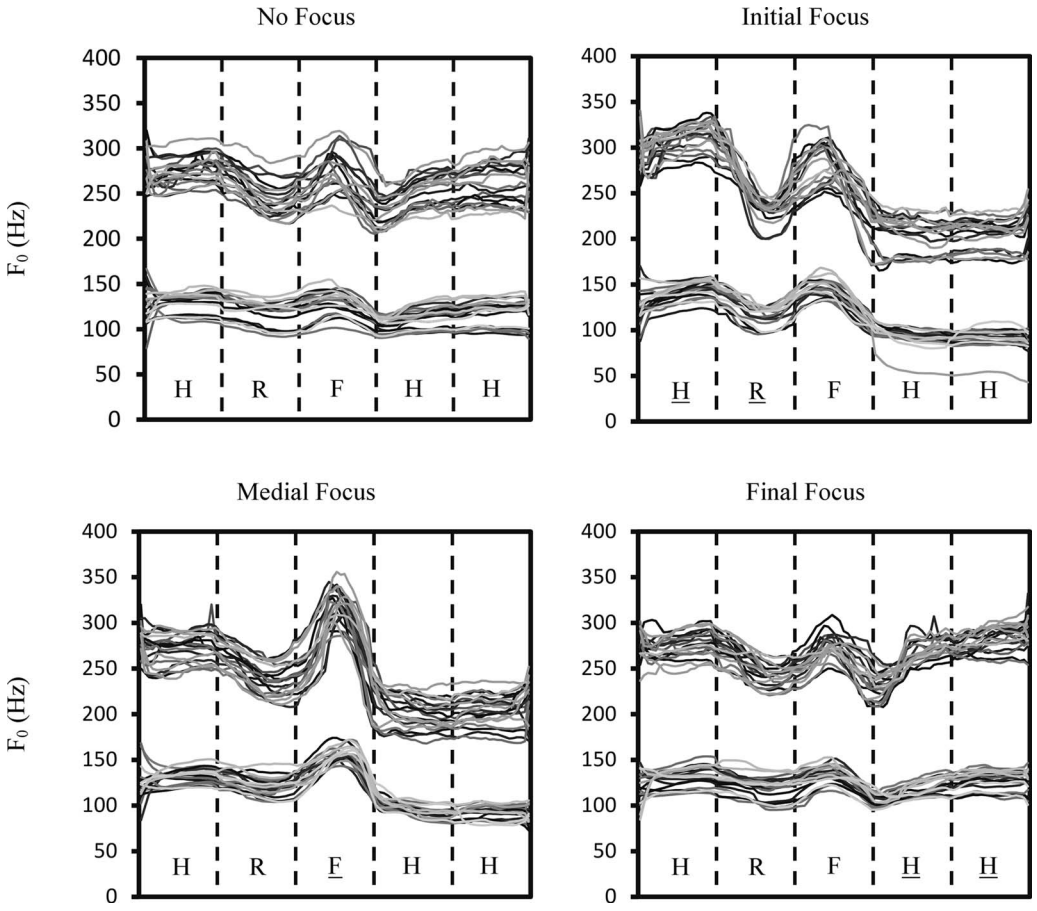


FIG. 9. F_0 contours of the HRFHH tone sequence spoken by eight speakers in different focus conditions in the Mandarin corpus. The focused syllable is denoted as the underlined letter. Each panel contains 40 F_0 contours (five repetitions by each speaker). The vertical dashed lines mark the syllable boundaries. In each panel the two clusters are from the female and male speakers, respectively.

focus functions were extracted. For English, parameters for the stress and focus functions were extracted.

In intonation modeling, F_0 estimation algorithms typically represent F_0 in hertz scale, especially those that model surface F_0 directly (Hirst and Espesser, 1993; Pierrehumbert, 1981; Taylor, 2000; van Santen and Möbius, 2000). The hertz scale is also commonly used for F_0 display in most speech analysis applications. However, there is evidence that pitch operates in speech on a logarithmic scale in both perception (Nolan, 2003) and production (Fujisaki, 2003). Thus it is inappropriate to pool data across different speakers in hertz during modeling as they may have very different pitch spans depending on their pitch ranges. The current version of qTA model thus uses the semitone (st) scale for parameter representation while the hertz scale is used only for displaying F_0 so that the surface contours more directly reflect what is commonly seen. For speaker normalization purposes, the semitone scale is used for normalizing pitch ranges across speakers, and utterance-onset F_0 is used as a baseline to normalize speaker differences in terms of average F_0 level. For cross-speaker normalization, the extracted pitch target parameters are converted from hertz scale to semitone scale before averaging within each designated communicative function.

1. Mandarin

Table V shows the means and confidence intervals of the extracted parametric vectors of the tone function. They are extracted from data in prefocus regions, which also include all the words in sentences with no narrow focus, as explained earlier. The b values are measured relative to the initial F_0 of

TABLE III. Words and corresponding tone patterns of the Mandarin dataset (Xu, 1999). H, R, L, and F represent high, rising, low, and falling tones, respectively. The numerals at the end of each syllable also represents the tones: 1, 2, 3, and 4 for H, R, L, and F, respectively.

Word 1			Word 2			Word 3		
HH	mao1 mi1	"kitty"	H	mo1	"touches"	HH	mao1 mi1	"kitty"
HR	mao1 mi2	"cat-fan"	R	na2	"takes"	LH	ma3 dao1	"sabre"
HL	mao1 mi3	"cat-rice"	F	mai4	"sells"			
HF	mao1 mi4	"cat-honey"						

TABLE IV. A list of sentences in the English dataset (Xu and Xu, 2005).

Word 1	Word 2	Word 3	Word 4	Word 5
Lee/Nina/Lamar/ Emily/Ramona	May	Know	My	Niece
Lee		Lure/mimic/ minimize		Niece
Lee		Know		Niece/nanny/ mummy

TABLE V. Means and confidence intervals of extracted parametric vectors of the tone function obtained from the Mandarin dataset. Because different speakers have different average F_0, the utterance-onset F_0 is subtracted from b, the pitch target height, so that the values of b in the table are relative to the utterance-onset F_0.

Tone	m (st/s)	b (st)	λ (s^{-1})
H	0	0.0 ± 1.0	54.5 ± 5.4
R	93.4 ± 3.4	-2.2 ± 1.1	40.7 ± 3.8
L	0	-8.9 ± 0.6	34.1 ± 5.0
F	-106.4 ± 3.0	-2.5 ± 1.3	39.3 ± 3.2

each sentence. The model parameters in Table V show clearly distinct target values for the tones for both male and female speakers. Note that the positive and negative m for the R and F tones would have to be each represented by a combination of both positive and negative commands in the CR model (Fujisaki *et al.*, 2005), which demonstrates the greater simplicity of the qTA than the CR model in representing the tone function. The smaller absolute value of m in the R tone than in the F tone is consistent with the empirical finding that the maximum speed of pitch rises is lower than that of pitch drops (Xu and Sun, 2002). It should be noted that, for Tables V–IX, the confidence intervals of m and Δm for both H and L tones equal zero because those tones are assumed to have static targets. Note also that the small confidence intervals here and in Tables VI–X are partially attributable to the restrictions placed on the search space. Some of the confidence intervals may become larger once the restrictions are removed, as shown in Table II.

Table VI shows the means and confidence intervals of the adjustment vectors of the focus function extracted from on-, post-, and final-focus regions. These adjustment parameters are expressed as differences from the parameter values in the prefocus region shown in Table V. Both m and b in the on-focus region are magnified so that b is higher for the H tone but lower for the L tone while the absolute value of m is increased for both R and F tones. In contrast to the on-focus region, the parameter values in the postfocus region are compressed and lowered. In the final-focus region, the b adjustments are very small for the H tone but slightly larger for the L tone. Interestingly, there are no significant changes in on-focus λ adjustments as can be seen from their confidence intervals. This indicates that when the syllable is focused the approximation rate remains largely constant.

2. English

For English, parametric and adjustment vectors are obtained by averaging the individual vectors according to lexical stress, position in sentence, and focus. The positional differentiation is applied only in the prefocus region, which is to indirectly model the combined effect of new topic and sentence modality. Based on the findings of Xu and Xu (2005), we assume static targets for every syllable, except for

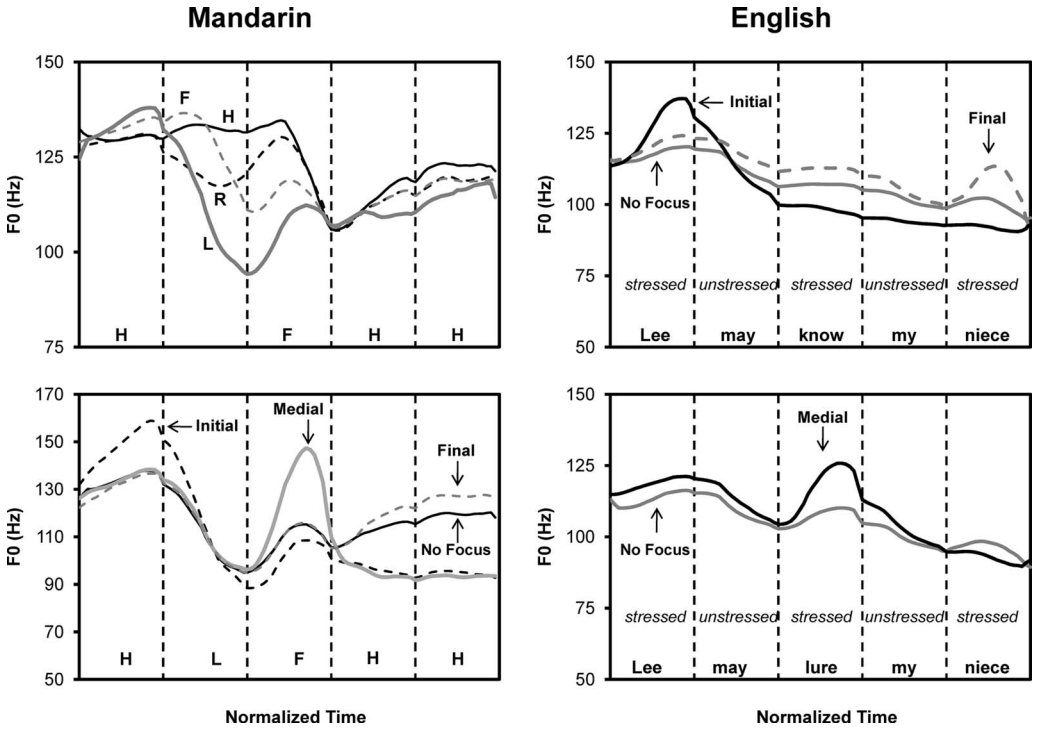


FIG. 10. Examples of naturally produced F_0 contours of Mandarin (left) and English (right) with tonal and focal variations. The vertical dashed lines mark the syllable boundaries. Adapted from Xu (1999) and Xu and Xu (2005), respectively.

TABLE VI. Means and confidence intervals of focus adjustment vectors of on-focus, postfocus, and final-focus regions obtained from the Mandarin dataset. These focus adjustment vectors are relative to the parametric vector of the tone function in Table V.

Focus location	Tone	Δm (st/s)	Δb (st)	$\Delta\lambda$ (s^{-1})
On focus	H	0	2.3 ± 1.1	-1.6 ± 3.5
	R	11.8 ± 3.6	0.6 ± 1.0	-3.9 ± 3.7
	L	0	-2.4 ± 1.8	1.0 ± 6.7
	F	-6.7 ± 2.5	1.2 ± 1.8	-1.2 ± 3.2
Postfocus	H	0	-5.6 ± 1.0	-11.2 ± 4.2
	R	-7.3 ± 2.9	-4.1 ± 0.7	7.8 ± 7.0
	L	0	-4.1 ± 1.4	-3.2 ± 5.4
	F	4.5 ± 3.3	-2.8 ± 1.5	2.5 ± 5.3
Final focus	H	0	-0.2 ± 0.8	-16.0 ± 2.5
	L	0	-2.1 ± 1.5	-4.6 ± 5.4

the sentence-final monosyllabic words and the focused word-final stressed syllables because their pitch targets in declarative sentences are likely to be [fall]. Table VII shows the means and confidence intervals of parametric vectors of the stress function in the prefocus region for each syllable position and stress condition, e.g., stress (stressed,1). The integer in the second argument of this stress function indicates syllable position. In Table VII, the b values are always higher for stressed syllables than for unstressed syllables even though focus adjustments are not yet involved in these prefocus regions. This is consistent with the finding that the effect of lexical stress on F_0 in English, though small in magnitude, is independent of focus (Xu and Xu, 2005) and audible to native listeners (Fry, 1958).

Table VIII shows the means and confidence intervals of the focus adjustment vectors obtained from on-, post-, and final-focus regions. While the values of postfocus adjustment b are all negative, the on-focus and final-focus adjustments are positive for stressed syllables but negative for unstressed syllables. This is consistent with the findings of Xu and Xu (2005). It is interesting to note that the on-focus $\Delta\lambda$ values are mostly negative. It could be partially a correction of the excessively large values for prefocus syllables seen in Table VII as compared to those in Table V for Mandarin. Speculatively, when the actual F_0 excursion size is very small, the estimation of λ tends to be larger than reality. This should be

TABLE VII. Means and confidence intervals of parametric vectors of the stress function obtained from the English dataset.

Syllable position	Stress	m (st/s)	b (st)	λ (s^{-1})
1	Unstressed	0	0.1 ± 0.8	72.8 ± 6.7
	Stressed	0	1.3 ± 0.9	48.1 ± 9.1
2	Unstressed	0	-1.2 ± 0.9	41.0 ± 5.9
3	Unstressed	0	-2.0 ± 0.7	51.4 ± 7.2
	Stressed	0	-1.2 ± 0.6	62.6 ± 11.8
4	Unstressed	0	-2.8 ± 0.6	43.3 ± 6.7
5	Unstressed	0	-5.3 ± 1.2	58.3 ± 9.9
	Stressed	0	-1.3 ± 1.2	49.8 ± 12.0

TABLE VIII. Means and confidence intervals of adjustment vectors of the focus function obtained from the English dataset. They are derived relative to the parametric vectors in Table VII.

Focus location	Stress	Δm (st/s)	Δb (st)	$\Delta\lambda$ (s^{-1})
On focus	Unstressed	0	-1.1 ± 0.7	-5.2 ± 3.7
	Stressed	0	2.9 ± 1.1	-10.6 ± 2.4
Postfocus	Unstressed	0	-1.9 ± 0.9	14.3 ± 5.0
	Stressed	0	-2.9 ± 1.0	1.4 ± 11.1
Final focus	Unstressed	0	-1.7 ± 2.7	0.9 ± 13.5
	Stressed	0	2.3 ± 1.6	-24.0 ± 8.7

examined in future research. Also, the postfocus $\Delta\lambda$ values are positive for an unstressed syllable and no change for a stressed syllable, which differs from the reduction in postfocus word strength in the Stem-ML found by Shih and Kochanski (2003). This is also an issue worth exploring in future research.

Table IX shows the parametric vectors of the stress and focus functions for the sentence-final monosyllabic words and the focused word-final stressed syllable. Here the values of m are always negative regardless of focus condition, although the search space was not limited to below zero, indicating steep falling targets (Pierrehumbert, 1980; Xu and Xu, 2005). The postfocus adjustment of m is positive, indicating a decrease in the slope due to pitch range reduction.

Overall, the parameters obtained for both Mandarin and English are consistent with the results of systematic acoustic analyses for the respective corpora (Xu, 1999; Xu and Xu, 2005). This is initial indication that the analysis-by-synthesis method employed in the present study is effective.

C. Model evaluation by assessing synthesis quality

The effectiveness of the qTA model was further evaluated in two ways: (a) numerical assessment of closeness of fit between synthesized and natural F_0 and (b) perceptual identification of tone and focus as well as judgment of naturalness by native speakers of Mandarin and English.

1. Numerical assessment

The tests were conducted using a leave-one-out cross-validation scheme. This is to assure the reliability of the evaluation by testing the robustness of the qTA model against interspeaker variability. Each time, the data of one

TABLE IX. Means and confidence intervals of parametric vectors of the stress function and the adjustment vectors of the focus function for the exceptional cases where the pitch target is dynamic. These parametric vectors are also derived from the English corpus. The symbol Δ indicates that the parameters in that row are relative to the prefocus parameters.

Stress	Focus	Δm (st/s)	Δb (st)	$\Delta\lambda$ (s^{-1})
Focused word-final stressed syllable	On focus (Δ)	-81.1 ± 14.8	3.5 ± 1.4	-20.5 ± 1.3
Sentence-final monosyllabic word	Prefocus	-90.2 ± 11.8	-4.3 ± 1.5	38.4 ± 7.9
	Postfocus (Δ)	27.1 ± 14.6	-1.0 ± 1.9	-4.6 ± 5.6
	Final focus (Δ)	-22.8 ± 9.7	0.3 ± 1.6	-11.6 ± 2.7

TABLE X. Average RMSE and correlation coefficients in the Mandarin simulations. Synthesized F_0 used in these comparisons are those generated by resynthesis, tone function, focus function, and positional effect.

Imposed function	No. of parametric vector	RMSE (st)	Correlation
Resynthesis	19 200	0.56	0.92
Tone	4	2.84	0.74
Tone+position	20	2.46	0.75
Tone+focus	16	2.24	0.77
Tone+focus+position	80	2.16	0.78

speaker were circularly selected as the test set while those of the rest of the speakers formed the training set. Using this cross-validation scheme, the experiment repeated eight times, thus maximizing the chances of detecting the worst errors. Two measurements were used to assess the closeness of fit between synthesized and natural F_0, RMSE and Pearson's correlation coefficient. RMSE measures the difference between natural and synthesized F_0 contours while correlation coefficient indicates the linear relationship between them. It should be noted that the correlation coefficients were used for evaluating F_0 contours, not the model parameters. This evaluation matrix is the same as those reported in previous modeling attempts (Taylor, 2000; Dusterhoff *et al.*, 1999; Jilka *et al.*, 1999). Positive high correlation indicates consistency between original and synthesized F_0 contours not only in height but also in contour shapes. The semitone rather than hertz scale is used for measuring RMSE so that the results from different speakers can be assessed together. In contrast, correlation coefficients are computed in hertz to maximize the F_0 discrepancy. The training phase began with automatically extracting the parametric vectors from each utterance. The resulting parametric vectors were then summarized according to the functions to be tested. In the testing phase, RMSE and correlation coefficients for each sentence were obtained for each function.

The natural F_0 contours were first compared to the resynthesized contours, i.e., those generated with the parameters extracted from each individual sentence itself. Then, F_0 contours of two synthesized communicative functions were compared to the natural F_0: tone (Mandarin) or stress (English) and focus. Also, two positional dependencies were tested: syllable position (F_0 generated with parameters averaged for each position in a sentence) and tonal context (F_0 generated with parameters averaged for each tonal context). The syllable position dependency was tested for both Mandarin and English. The tonal context dependency was tested only for Mandarin.

2. Perceptual evaluation

For the perceptual evaluation experiments, the testing datasets were generated based on the communicative functions simulated in the present study. To test the effectiveness of duration modification in the focused syllable, the duration adjustment of the focus function was also included in the perceptual test. For the Mandarin tests, the following stimuli were generated.

Tone:	mao (H) mi (H/R/L/F) mo (H/R/F) ma (L) dao (H)	$4\times3=12$.
Focus:	neutral/initial/medial/final	$12\times4=48$.
Synthesis:	original/ synthetic without focus duration/ synthetic with focus duration	$48\times3=144$

The natural speech material was recorded specifically for the present study by the second author, who was one of the speakers in Xu (1999). In this recording, unlike in Xu (1999), the third syllable was always /mo/ regardless of its tone so as to guarantee minimal tonal contrasts for reliable assessment of tone identification at this sentence location.

For the Mandarin perceptual test, qTA parameters for tone and focus were extracted from each utterance. For each stimulus sentence, F_0 was synthesized and then used to replace the F_0 of the host utterance. There were 12 host utterances, in which the tones of the second and third syllables were varied. The focus-specific F_0 contours generated for each focus condition were used to replace the original contours of these utterances using pitch-synchronous overlap and add (PSOLA) method in PRAAT (Boersma, 2001). The importance of focus duration is tested by modifying the syllable duration of the segmental data of the original speech. Nine native Mandarin listeners participated in the test. The test was set up as a web-based program. The setup of the test consisted of two main sections. In the first section, listeners were instructed to identify the tones of the second and third syllables by selecting the Chinese character with the specific tone. In the second section, they were instructed to identify the focused word (i.e., the word being emphasized) and to judge whether the utterance was naturally spoken or synthesized. For each question in the test, the natural and synthesized sounds were randomly presented to the listeners. Listeners could listen to the sample sounds as many times as they preferred.

The conditions in the English perception experiment were focus, sentence, and synthesis method. There were eight natural speech utterances, including four neutral focus utterances from four different sentences, two initial focus utterances (S1 and S2), one medial focus utterance (S3), and one final focus utterance (S4). For speech with synthesized F_0, there were 32 test utterances according to the following composition.

Focus:	neutral/initial/medial/final	$4=4$.
Sentence:	[four different sentences] S1/ S2/S3/S4	$4\times4=16$.
Duration:	with focus duration/without focus duration	$16\times2=32$.

The natural speech materials used in making the stimuli were drawn from the English dataset used in the present study. Eight natural utterances were selected from the speaker with lowest resynthesis RMSE in the numerical assessment. Thus, there were totally 40 test utterances. To synthesize the test stimuli, we used only four neutral focus ut-

TABLE XI. Average RMSE and correlation coefficients for adding context dependency in the simulation of the Mandarin dataset.

Imposed function	No. of parametric vector	RMSE (st)	Correlation
Tone	4	2.84	0.74
Tone+preceding context	16	2.57	0.76
Tone+following context	16	2.50	0.77
Tone+focus	16	2.24	0.77
Tone+focus+preceding context	68	2.17	0.78
Tone+focus+following context	68	2.21	0.77

terances from four different sentences and modified the syllable duration of the segmental data of the original speech using PRAAT. Fourteen native English listeners participated in this listening experiment: ten of them were American English speakers and four were British English speakers. During the test, participants were asked to identify the focused words and to judge whether the test utterances were naturally spoken or synthesized. The test was set up in the same way as the Mandarin perception test.

3. Evaluation results

a. Mandarin Table X shows RMSE and correlation coefficients of the resynthesis and synthesis with successively added function-specific components. The results of resynthesis of the same utterances show very low error rates and very high correlations, indicating that the qTA-regenerated F_0 contours can fit the original contours quite well. From resynthesis to synthesis with function-specific parameters there is a general increase in error rate, but note also the dramatic reduction in the number of parameter vectors used (from 19 200 to 4–80). The tone-only condition has the highest error rate and lowest correlation. There is a slight improvement when either focus or positional specification was added. It is obvious that focus is the more effective constraint, as the result of adding focus specification is better than adding position specification. Moreover, the results of adding both focus and position are not significantly different in terms of RMSE from those of adding focus only [$F(1,15)=0.03$, $p=0.858$]. The more concrete evidence is in the analysis of correlation, which indicates that only the focus function significantly affects the correlation [$F(1,31)=5.21$, $p=0.030$]. This is probably because the focus specifications have already implicitly included the positional effect when calculating the adjustment vectors for different focus regions.

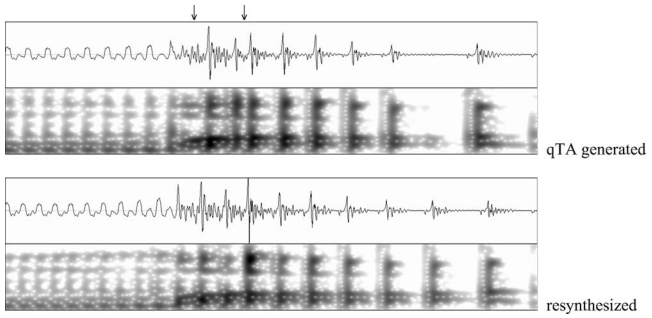


FIG. 12. Upper panel: Aperiodicity in a synthetic sentence. The arrows point to the two locations where the periods are either exceptionally long or exceptionally short. Lower panel: The resynthesized original sentence, where no strong aperiodicity is seen in the same locations.

Table XI shows the results of adding tonal context dependency to the model. The improvements are not significantly different between the preceding and following tone contexts, although they both show a slight but insignificant reduction in error rate. There are no further improvements in adding tonal context whether or not focus has been added [RMSE:$F(2,47)=0.23$, $p=0.795$; correlation: $F(2,47)=1.39$, $p=0.259$]. The focus function, however, significantly affects the correlation [$F(1,47)=4.65$, $p=0.036$]. This also indicates the importance of the focus function in determining the F_0 contours.

Figure 11 shows the results of the perceptual evaluation of natural and synthetic F_0 in Mandarin speech as described earlier. Listeners could identify the tones equally well in both syllable positions [$F(1,35)=0.54$, $p=0.470$]. There is no significant difference in identification rate between natural and synthesized F_0 [$F(1,35)=1.05$, $p=0.317$]. Moreover, there was no significant difference in focus identification between natural and synthesized F_0 whether or not natural duration adjustment was applied to the synthetic speech [$F(2,26)=2.36$, $p=0.127$]. However, naturalness perception differed between natural and synthesized F_0 [$F(2,26)=6.07$, $p=0.011$]. A close examination of the synthetic sentences that had low naturalness scores found cases of acoustic discontinuity. An example is shown in Fig. 12, where the locations of aperiodicity in the synthetic speech are indicated by the arrows. Such aperiodicity is not seen in

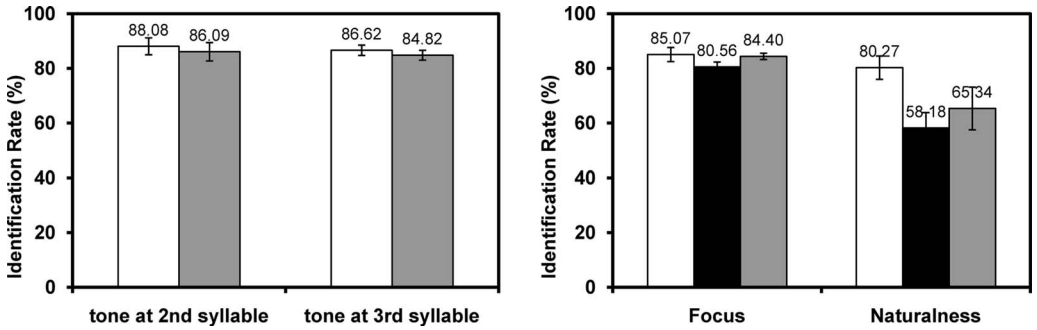


FIG. 11. Means (bars and the numbers above them) and standard errors (vertical lines) of identification rates in the Mandarin perceptual evaluation. The left graph shows averaged tone identification results while the right graph shows results of focus identification and naturalness evaluation. In the left graph, the white and gray bars indicate rate of tone identification for natural and synthetic F_0. In the right graph, the white, black, and gray bars indicate the results of focus identification and naturalness rating for natural F_0, synthetic F_0 without focus duration, and synthetic F_0 with focus duration, respectively.

TABLE XII. Averaged RMSE and correlation coefficients in the simulation of English dataset.

Imposed function	No. of parametric vector	RMSE (st)	Correlation
Resynthesis	14 224	0.32	0.83
Stress	4	1.93	0.75
Stress+position	11	1.71	0.78
Stress+focus	12	1.68	0.77
Stress+focus+position	18	1.57	0.78

the resynthesized original shown in the lower panel. This difference in synthesis quality is probably due to the ineffectiveness of the PSOLA algorithm in modifying pitch and duration at the same time.

b. English Table XII shows the results of numerical assessment for the English synthesis. For the resynthesized F_0 the errors are very low while the correlation coefficients are very high. Simulating only the stress function leads to higher errors and slightly lower correlations. This is similar to the case of the tone-only simulation in Mandarin in Table X. Similar to the Mandarin tests, there is no significant improvement by adding positional dependency when the focus function has already been used [RMSE: $F(1,15)=1.07$, $p=0.318$; correlation: $F(1,15)=0.34$, $p=0.571$]. There is, however, a significant main effect of adding positional dependency, but only in terms of correlation [$F(1,31)=4.94$, $p=0.034$]. An overall F_0 declination related to a combined effect of new topic and sentence modality, as discussed earlier, is the plausible underlying mechanism of this improvement.

Figure 13 shows the perception results of focus identification and naturalness evaluation of the natural and synthetic F_0 for the English dataset. There is a significant difference in focus identification between natural and synthetic F_0 [$F(2,41)=7.02$, $p=0.004$]. Listeners could identify focus significantly better from natural than from synthetic F_0 without duration adjustment [$F(1,27)=25.49$, $p<0.001$]. However, they could identify focus equally well from natural and synthetic F_0 with duration modification [$F(1,27)=0.62$, $p=0.444$]. Listeners also perceived a difference between synthesis methods [$F(2,41)=13.10$, $p<0.001$], but they did not recognize the difference in natural and synthetic F_0 with focus duration [$F(1,27)=1.82$, $p=0.200$].

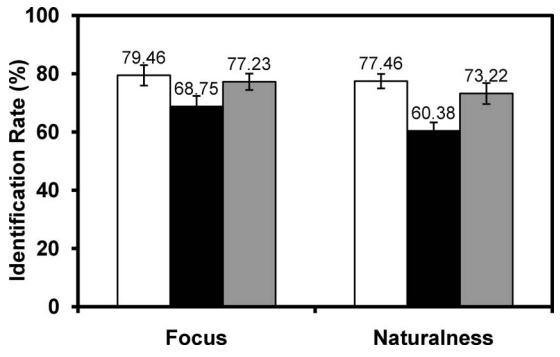


FIG. 13. Means (bars and the numbers above them) and standard errors (vertical lines) of focus identification rate and naturalness evaluation in the English perception tests. White, black, and gray bars correspond to sentences that are natural, synthetic without focus duration, and synthetic with focus duration, respectively. The vertical line in each bar indicates standard error.

V. DISCUSSION

The adequacy of any theoretical understanding of a natural phenomenon can be best measured in terms of the amount of details it can predict. This should also be true of theories about tone and intonation. Of the various tonal and intonational theories proposed so far (e.g., Bolinger, 1986; O'Connor and Arnold, 1961; Ladd, 1996; Pierrehumbert, 1980; Pike, 1945), few, if any, have been specific enough to allow full numerical testing of their adequacy. Although the autosegmental-metrical model has been used in synthesizing intonation as reported by Pierrehumbert (1981), no numerical tests were conducted to evaluate the quality of the synthesis. Meanwhile, there have also been many quantitative models of tone and intonation, as discussed in Sec. I, but none has been designed to directly test existing theories. While it is true that quantitative modeling can offer only an approximation to reality, theories without sufficient quantitative specifications can at best provide even coarser approximations to reality. The major goal of the present study is to quantify the theory about tone and intonation embodied in the PENTA model, which hitherto has not been fully numerical, and to subject it, through this quantification effort, to more rigorous testing than has been done before.

The PENTA model assumes that tone and intonation serve to convey communicative functions through specific encoding schemes that are implemented by the articulatory system, as sketched in Fig. 2. The model thus consists of two core components: an articulatory mechanism and a set of communicative functions. The articulatory mechanism assumed in PENTA is syllable-synchronized sequential TA, as depicted in Fig. 1. With this mechanism, the articulatory system asymptotically approaches the underlying targets that are either static or dynamic. The approximation is always synchronized with the syllable. But due to inertia, articulatory states are transferred onward across syllable boundaries. Although quite specific, this conceptualization needs to be quantified into a dynamic system that is biophysically plausible. The communicative functions are considered in PENTA as the driving force of the system. They are assumed to be parallel to each other, each with a unique encoding scheme in terms of one or more of the TA parameters. The uniqueness and complexity of these functions entail that each of them can be properly simulated only with sufficient knowledge about its temporal domain of operation, the TA parameters involved, and the value range of the parameters, e.g., whether the target is static or dynamic.

To develop a biomechanically plausible dynamic system, the simulation is best done at a level where the link between articulatory control and communicative functions is the most direct, although many levels of biophysical processes are apparently involved. Simulation at the level of individual muscular forces, as done in the CR model (Fujisaki *et al.*, 2005), would entail variable parameter values depending on the tonal contexts, as the individual mus-

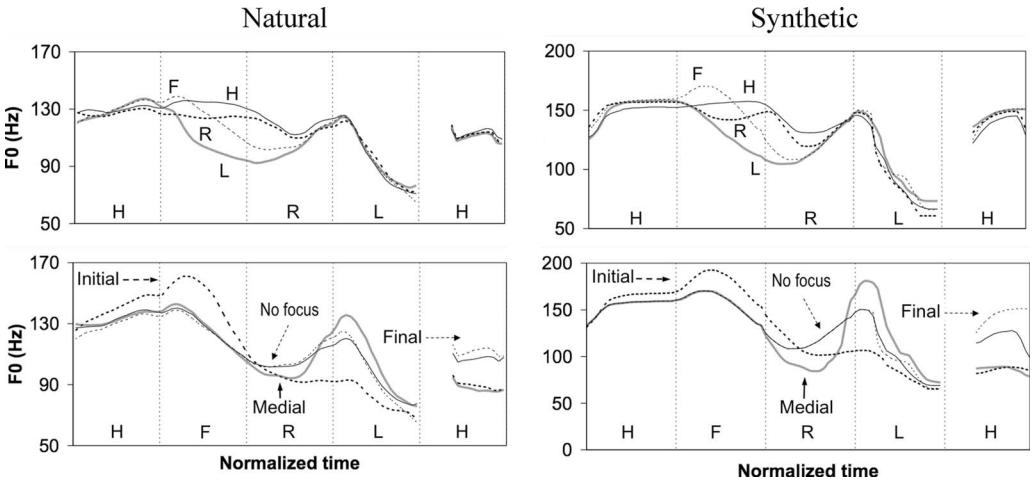


FIG. 14. Examples of natural and synthetic F_0 contours of tone and intonation in Mandarin. In the top plots, the sentences differ in the tone of the second syllable, with focus always on the last two syllables. In the bottom plots, the sentences differ in focus location: the first (disyllabic) word, the second (monosyllabic) word, the final (disyllabic) word, or none of the words.

cular forces have to be differentially adjusted according to the distance to be covered between the initial and targeted articulatory states. Simulation at a level where F_0 is controlled without the mediation of an articulatory mechanism, as done in the tilt model (Taylor, 2000) and SFC model (Bailly and Holm, 2005), would miss critical dynamic details due to articulatory constraints, such as unidirectional left-to-right assimilatory influence (Gandour *et al.*, 1994; Wong, 2006; Xu, 1997, 1999) or peak delay across syllable boundaries (Xu, 1998, 2001), both due to the physical constraint of inertia. The qTA model developed in the present study implements a third-order critically damped system, which generates F_0 contours through syllable-synchronized sequential TA. This strategy severely constrains the degrees of freedom of the model, limiting its variable control parameters to only three, each corresponding directly to a TA parameter in PENTA. Automatic parameter extraction through analysis by synthesis in the present study has found these parameters to remain largely invariant across different tonal contexts, especially the preceding tonal context which is known to cause a large amount of contextual variability (Gandour *et al.*, 1994; Wong, 2006; Xu, 1997, 1999). This can be seen in Fig. 14 (upper row). Also can be seen in Fig. 14 is that the phenomenon of peak delay (Xu, 2001) in a RL sequence is effectively simulated without using peak location as a control parameter.

The correspondence of the qTA parameters to the TA parameters in the PENTA model means that the assumptions of the PENTA model about the interorthogonality between the communicative functions and about the uniqueness of the function-specific encoding schemes can be tested through supervised learning. To that end, we tested the model with two natural speech corpora (Xu, 1999; Xu and Xu, 2005). The parameter extraction was guided by knowledge obtained from previous production studies (for English: Cooper *et al.*, 1985; Xu and Xu, 2005; for Mandarin: Liu and Xu, 2005; Xu, 1997, 1999), which was used to limit the value range of the parameters. For example, based on previous research on focus, separate parameter sets were obtained for prefocus, on-focus, postfocus, and final-focus regions. Note that restricting the parameters' range this way during training is actually risky because anything faulty in the assumptions behind the restrictions would increase the chance of generating large errors during testing. The results of the testing were nevertheless encouraging. The numerical evaluations showed that even when applied to speakers not included in the training, the error rates were comparable with those of previous studies (Fujisaki *et al.*, 2005; Hirst and Espresser, 1993; Kochanski and Shih, 2003; Pierrehumbert, 1981; Taylor, 2000). More significantly, the results of the perceptual evaluations show that not only can native listeners identify the functional categories from model-based synthesis just as well as from natural speech, but also they could not reliably distinguish utterances with synthetic F_0 from those with natural F_0 in terms of naturalness, especially in the case of English. These results not only demonstrate the effectiveness of the qTA model but also provide support for the validity of the theoretical assumption of the PENTA model that tone and intonation are in essence articulatorily encoded communicative functions.

A secondary goal of the present study is to develop an intonation generation system that is applicable in speech technology. There have been a number of attempts to apply automatic trainings to articulatory-oriented parameters in intonation synthesis (Fujisaki *et al.*, 2005; Kochanski and Shih, 2003; Mixdorff, 2000). Although they can approximate F_0 contours quite accurately, the lack of complete framework to implement the communicative functions in these models makes it difficult to effectively integrate them in speech processing systems. The simple form of underlying pitch targets consisting of only two parameters in qTA also makes the process of TA more straightforward than those that are more complex, e.g., templates and commands. Although the practicality of analysis by synthesis is limited due to a large search space, the articulatory constraints imposed on the automatic parameter extraction process effectively reduce the search space, as shown in this study. Thus for speech technology, the application of qTA in speech synthesis may improve the intonation components of the synthesizers.

Despite the overall quality of the simulation in the present study, a number of caveats need to be mentioned. The first is that the strategy used here is to simulate only functions that have been systematically studied in empirical research, and to do so by imposing specific restrictions on the search space for the model parameters based on the empirical findings. The simulation process is therefore not fully "automatic." It is our understanding, however, that our current knowledge level does not yet allow us to build a fully automatic process without sacrificing the quality of simulation. This is because there are many communicative functions that are encoded by intonation as discussed in the Introduction, and without knowing full well what they are and which of them have occurred in a particular utterance, it is difficult even for trained human labelers to label them consistently (Wightman, 2002; Xu, 2006). On the other hand, once a given specific communicative function has been recognized, it is just one step further from the current strategy of imposing *a priori* limits on the search space to one that is able to discover the TA parameters used by the function and the range of their values. This could be done by taking advantages of recent development in unsupervised learning of phonetic categories (Gauthier *et al.*, 2007; Guenther, 1994). It would be even more desirable for future development of the model to attain the ability to automatically discover the communicative functions encoded in a particular language. As far as we can see, this could be done only if we have accumulated a substantial amount of knowledge through further research.

Second, several phenomena reported for tone and intonation that are likely due to articulatory mechanisms have not yet been simulated in the current version of the qTA model. These include anticipatory dissimilation, i.e., the raising of F_0 by a following L tone (Gandour *et al.*, 1994; Laniran and Clements, 2003; Wong, 2006; Xu, 1997, 1999), post-L bouncing (Chen and Xu, 2006; Pierrehumbert, 1980), whereby F_0 is raised by a preceding L tone, and dynamic delay, whereby the initial approximation of a dynamic pitch target is delayed at a slow speech rate so as to guarantee that the slope of the target is not reached earlier than the end of the syllable (Wong, 2006; Xu, 1998, 2001). Future development of the qTA model will need to find ways to simulate these mechanisms. Because each of these mechanisms seems to affect only a particular aspect of the TA process (Xu, 2006), they will likely to be modeled as additional mechanisms added on top of the core algorithm of qTA.

Third, one of the TA parameters, namely, duration, was only partially tested in the present study. It was tested in the perceptual evaluations by comparing focus-specific syllable duration with non-focus-specific duration. The former was found to generate F_0 contours that better match natural F_0 in conveying focus. While this is an encouraging result, what is lacking in general is a theoretical link between duration and specific communicative functions, for which existing research so far has provided only a vague picture (see Shattuck-Hufnagel and Turk, 1996 for a review). Further studies aimed at revealing that the functional sources of duration patterns are needed [see Xu, (2008) for further discussion].

Finally, only a limited number of communicative functions have been modeled in the current project. Many more that have been discussed in literature (Bolinger, 1989; Hirschberg, 2002; Hirst, 2005; Kohler, 2005; Xu, 2005, 2008) such as sentence type (e.g., statement versus question), topic and turn taking, speaking style, emotion, to name just a few, have not been tested. As discussed in the Introduction, qTA, as a quantitative implementation of PENTA, was developed exactly for the purpose of modeling multiple communicative functions with a limited number of articulatory-based parameters. The present results seem to suggest that it will provide a promising tool for modeling many more communicative functions in future research.

VI. CONCLUSION

We have proposed in this paper a qTA model which implements the theoretical PENTA model and simulates F_0 contours as the process of TA. The model simulates surface F_0 by adjusting parameters of local TA, including the height and slope of the pitch targets and the rate of approaching individual targets. There are only three free parameters in qTA, and their controls are all directly linked to specific communicative functions such as lexical tone, lexical stress, focus, sentence modality, and new topic, although only the first three were explicitly modeled in the present study. We tested the model by training it with Mandarin and English speech data using an automated analysis-by-synthesis procedure. The F_0 simulations were evaluated both numerically and perceptually. The accuracy of resynthesis by the model was high, with RMSE of 0.56 st and correlation of 0.92 for Mandarin, and RMSE of 0.32 st and correlation of 0.83 for English. The quality of synthesizing specific communicative functions, including tone in Mandarin, lexical stress in English, and focus in both languages, was comparable to previous studies even when the trained parameters were applied to speakers not included during training, with RMSE of 2.24 st and correlation of 0.77 for Mandarin, and RMSE of 1.68 st and correlation of 0.77 for English. More importantly, listeners' perceptual identification of tone, lexical stress, and focus, and their judgment of the naturalness of speech is nearly identical between natural and synthetic F_0. These results demonstrate the validity of the assumptions underlying qTA and suggest that it can be used as an effective tool both for theoretical study of tone and intonation and for generating F_0 contours in automatic speech synthetic systems.

[1]Arvaniti, A., and Ladd, D. R. (1995). "Tonal alignment and the representation of accentual targets," Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm, Vol. **4**, pp. 220–223.
[2]Arvaniti, A., Ladd, D. R., and Mennen, I. (1998). "Stability of tonal alignment: The case of Greek prenuclear accents," J. Phonetics **26**, 3–25.
[3]Atterer, M., and Ladd, D. R. (2004). "On the phonetics and phonology of "segmental anchoring" of F_0: Evidence from German," J. Phonetics **32**, 177–197.
[4]Bailly, G., and Holm B. (2005). "SFC: A trainable prosodic model," Speech Commun. **46**, 348–364.
[5]Bernstein, N. (1967). *The Co-Ordination and Regulation of Movement* (Pergamon, London).
[6]Black, A. W., and Hunt, A. J. (1996). "Generating F_0 contours from ToBI labels using linear regression," Proceedings of ICSLP 1996, Philadelphia, PA, pp. 1385–1388.
[7]Boersma, P. (2001). "Praat, a system for doing phonetics by computer,"

Glot International **5**, 341–345.
[8]Bolinger, D. (1986). *Intonation and Its Parts: Melody in Spoken English* (Stanford University Press, Palo Alto, CA).
[9]Bolinger, D. (1989). *Intonation and Its Uses: Melody in Grammar and Discourse* (Stanford University Press, Stanford, Ca).
[10]Botinis, A., Bannert, R., and Tatham, M. (2000). "Contrastive tonal analysis of focus perception in Greek and Swedish,." in *Intonation: Analysis, Modelling and Technology*, edited by A. Botinis (Kluwer Academic, Boston), 97–116.
[11]Chen, A., Gussenhoven, C., and Rietveld, T. (2004). "Language-specificity in the perception of paralingistic intonational meaning," Iowa Dent. Bull. **47**, 311–349.
[12]Chen, Y., and Xu, Y. (2006). "Production of weak elements in speech: Evidence from F_0 patterns of neutral tone in standard Chinese," Phonetica **63**, 47–75.
[13]Cooper, W. E., Eady, S. J., and Mueller, P. R. (1985). "Acoustical aspects of contrastive stress in question-answer contexts," J. Acoust. Soc. Am. **77**, 2142–2156.
[14]d'Alessandro, C., and Mertens, P. (1995). "Automatic pitch contour stylization using a model of tonal perception," Comput. Speech Lang. **9**, 257–288.
[15]de Jong, K. (1994), "Initial tones and prominence in Seoul Korean," in *Working Papers in Linguistics* edited by S.-H Lee and S.-A. Jun (The Ohio State University Department of Linguistics), Vol. **43**.
[16]Dusterhoff, K. E., Black, A. W., and Taylor, P. (1999). "Using decision tree within the tilt intonation model to predict F_0 contours," Proceedings of EUROSPEECH'99, Budapest, pp. 1627–1630.
[17]Erickson, D. M. (1976). "A physiological analysis of the tones of Thai," Ph.D. thesis, University of Connecticut.
[18]Erickson, D., Honda, K., Hirai, H., and Beckman, M. E. (1995). "The production of low tones in English intonation," J. Phonetics **23**, 179–188.
[19]Fry, D. B. (1958). "Experiments in the perception of stress," Leeds Dent. J. **1**, 126–152.
[20]Fujimura, O. (2000). "Rhythmic organization and signal characteristics of speech," Proceedings of ICSLP 2000, Beijing, Vol. **1**, pp. 29–35.
[21]Fujisaki, H. (1974). "Formulation of the coarticulatory process in the formant frequency domain and its application to automatic recognition of connected vowels," Proceedings of the Speech Communication Seminar, Almqvist & Wiksell, Uppsala, Stockhom, pp. 385–392.
[22]Fujisaki, H. (1983). "Dynamic characteristics of voice fundamental frequency in speech and singing," in *The Production of Speech*, edited by P. F. MacNeilage (Springer-Verlag, New York), pp. 39–55.
[23]Fujisaki, H. (2003). "Prosody, information, and modeling: With Emphasis on Tonal Features of Speech," Proceedings of Workshop on Spoken Language Processing, Mumbai, pp. 5–14.
[24]Fujisaki, H., Wang, C., Ohno, S., and Gu, W. (2005). "Analysis and synthesis of fundamental frequency contours of standard Chinese using the command-response model," Speech Commun. **47**, 59–70.
[25]Gandour, J., Potisuk, S., and Dechongkit, S. (1994). "Tonal coarticulation in Thai," J. Phonetics **22**, 477–492.
[26]Gauthier, B., Shi, R., and Xu, Y. (2007). "Simulating the acquisition of lexical tones from continuous dynamic input," J. Acoust. Soc. Am. **121**, EL190–EL195.
[27]Gribble, P. L., Mullin, L. L., Cothros, N., and Mattar, A. (2003). "Role of cocontraction in arm movement accuracy," J. Neurophysiol. **89**, 2396–2405.
[28]Gribble, P. L., and Scott, S. H. (2002). "Overlap of internal models in motor cortex for mechanical loads during reaching," Nature (London) **417**, 938–941.
[29]Gu, W., Hirose, K., and Fujisaki, H. (2007). "Analysis of tones in Cantonese speech based on the command-response model," Phonetica **64**, 29–62.
[30]Guenther, F. H. (1994). "A neural network model of speech acquisition and motor equivalent speech production," Biol. Cybern. **72**, 43–53.
[31]Gussenhoven, C. (2007). "Types of focus in English," in *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, edited by C. Lee, M. Gordon and D. Büring (Springer, New York), pp. 83–100.
[32]Hallé, P. A. (1994). "Evidence of tone-specific activity of the sternohyoid muscle in modern standard Chinese," Leeds Dent. J. **37**, 103–123.
[33]Hasegawa, Y., and Hata, K. (1992). "Fundamental frequency as an acoustic cue to accent perception," Leeds Dent. J. **35**, 87–98.
[34]Hirano, M. (1974). "Morphological structure of the vocal cord as a vibrator and its variations," Folia Phoniatr. **26**, 89–94.
[35]Hirschberg, J. (2002). "Communication and prosody: Functional aspects of prosody," Speech Commun. **36**, 31–43.
[36]Hirst, D. (2005). "Form and function in the representation of speech prosody," Speech Commun. **46**, 334–347.
[37]Hirst, D., and Espesser, R. (1993). "Automatic modelling of fundamental frequency using a quadratic spline function," Travaux de l'Institut de Phonétique d'Aix **15**, 75–85.
[38]Jilka, M., Möhler, G., and Dogil, G. (1999). "Rules for the generation of ToBI-based American English intonation," Speech Commun. **28**, 83–108.
[39]Kelso, J. A. S. (1982). "The process approach to understanding human motor behavior: an introduction," in *Human Motor Behavior: An Introduction*, edited by J. A. S. Kelso (Erlbaum, Hillsdale, NJ), pp. 3–19.
[40]Klatt, D. H. (1987). "Review of text-to-speech conversion for English," J. Acoust. Soc. Am. **82**, 737–792.
[41]Kochanski, G., and Shih, C. (2003). "Prosody modeling with soft templates," Speech Commun. **39**, 311–352.
[42]Kochanski, G., Shih, C., and Jing, H. (2003). "Quantitative measurement of prosodic strength in Mandarin," Speech Commun. **41**, 625–645.
[43]Kohler, K. (2005). "Timing and communicative functions of pitch contours," Phonetica **62**, 88–105.
[44]Krahmer, E., and Swerts, M. (2001). "On the alleged existence of contrastive accents," Speech Commun. **34**, 391–405.
[45]Laniran, Y. O., and Clements, G. N. (2003). "Downstep and high raising: Interacting factors in Yoruba tone production," J. Phonetics **31**, 203–250.
[46]Ladd, D. R. (1983). "Phonological features of intonational peaks," Language **59**, 721–759.
[47]Ladd, D. R. (1996). *Intonational Phonology* (Cambridge University Press, Cambridge).
[48]Lehiste, I. (1975). "The phonetic structure of paragraphs," in *Structure and Process in Speech Perception*, edited by A. Cohen and S. E. G. Nooteboom (Springer-Verlag, New York), pp. 195–206.
[49]Lindblom, B. (1983) "Economy of speech gestures," in *The Production of Speech*, edited by P. MacNeilage (Springer, New York).
[50]Liu, F., and Xu, Y. (2005). "Parallel encoding of focus and interrogative meaning in Mandarin intonation," Phonetica **62**, 70–87.
[51]Mixdorff, H. (2000). "A novel approach to the fully automatic extraction of Fujisaki model parameters," Proceedings of ICASSP 2000, Istanbul, Vol. **3**, pp. 1281–1284.
[52]Mixdorff, H. (2004). "Quantitative tone and intonation modeling across languages," Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, pp. 137–142.
[53]Monsen, R. B., Engebretson, A. M., and Vemula, N. R. (1978). "Indirect assessment of the contribution of subglottal air pressure and vocal fold tension to changes in the fundamental frequency in English," J. Acoust. Soc. Am. **64**, 65–80.
[54]Ni, J., Kawai, H., and Hirose, K. (2006). "Constrained tone transformation technique for separation and combination of Mandarin tone and intonation," J. Acoust. Soc. Am. **119**, 1764–1782.
[55]Nolan, F. (2003). "Intonational equivalence: An experimental evaluation of pitch scales," Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, pp. 771–774.
[56]O'Connor, J. D., and Arnold, G. F. (1961). *Intonation of Colloquial English* (Longmans, London).
[57]Ohala, J. J. (1978). "Production of tone," in *Tone: A Linguistic Survey*, edited by V. A. Fromkin (Academic, New York), pp. 5–39.
[58]Palm, W. J., III (1999). *Modeling, Analysis, and Control of Dynamic Systems*, 2nd ed. (Wiley, New York).
[59]Pierrehumbert, J. (1980). "The phonology and phonetics of English intonation," Ph.D. thesis, MIT, Cambridge, MA.
[60]Pierrehumbert, J. (1981). "Synthesizing intonation," J. Acoust. Soc. Am. **70**, 985–995.
[61]Pierrehumbert, J., and Hirschberg, J. (1990). "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication*, edited by P. R. Cohen, J. Morgan, and M. E. Pollack (MIT Press, Cambridge, MA), pp. 271–311.
[62]Pierrehumbert, J., and Steele, S. (1990). "Categories of tonal alignment in English," Phonetica **47**, 181–196.
[63]Pike, K. (1945). *The Intonation of American English* (University of Michigan Press, Ann Harbor, MI).
[64]Prieto, P., van Santen, J., and Hirschberg, J. (1995). "Tonal alignment patterns in Spanish," J. Phonetics **23**, 429–451.
[65]Ross, K. N., and Ostendorf, M. (1999). "A dynamical system model for generating fundamental frequency for speech synthesis," IEEE Trans. Speech Audio Process. **7**, 295–309.
[66]Rump, H. H., and Collier, R. (1996). "Focus conditions and the promi-

nence of pitch-accented syllables," Leeds Dent. J. **39**, 1–17.
[67]Saltzman, E. L., and Kelso, J. A. S. (1987). "Skilled actions: A task dynamic approach," Psychol. Rev. **94**, 84–106.
[68]Saltzman, E. L., and Munhall, K. G. (1989). "A dynamical approach to gestural patterning in speech production," Ecological Psychol. **1**, 333–382.
[69]Shattuck-Hufnagel, S., and Turk, A. E. (1997). "A prosody tutorial for investigators of auditory sentence processing," J. Psycholinguist. Res. **25**, 193–247.
[70]Shih, C., and Kochanski, G. (2003). "Modeling intonation: Asking for confirmation in English," Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, pp. 551–554.
[71]Silverman, K., and Pierrehumbert, J. (1990). "The timing of prenuclear high accents in English," *Papers in Laboratory Phonology* (Cambridge University Press, Cambridge, UK), Vol. **I**, pp. 72–106.
[72]Sun, X. (2002). "The determination, analysis, and synthesis of fundamental frequency," Ph.D. thesis, Northwestern University.
[73]Swerts, M. (1997). "Prosodic features at discourse boundaries of different length," J. Acoust. Soc. Am. **101**, 514–521.
[74]'t Hart, J., Collier, R., and Cohen, A. (1990) *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody* (Cambridge University Press, Cambridge).
[75]Taylor, P. (2000). "Analysis and synthesis of intonation using the tilt model," J. Acoust. Soc. Am. **107**, 1697–1714.
[76]Titze, I. R. (1989). "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. **85**, 901–906.
[77]van Heuven, V. J. (1994). "What is the smallest prosodic domain?," in *Papers in Laboratory Phonology*, edited by P. A. Keating (Cambridge University Press, Cambridge), Vol. **3**, pp. 76–98.
[78]van Santen, J., and Möbius, B. (2000). "A quantitative model of F_0 generation and alignment," in *Intonation: Analysis, Modelling and Technology*, edited by A. Botinis, (Kluwer, Dordrecht), pp. 269–288.
[79]Wang, B., and Xu, Y. (2006). "Prosodic encoding of topic and focus in Mandarin," Proceedings of Speech Prosody 2006, Dresden, pp. 313–316.
[80]Wightman, C. W. (2002). "ToBI or not ToBI," Proceedings of Speech Prosody 2002, Aix-en-Provence, pp. 25–29.
[81]Wong, Y. W. (2006). "Contextual tonal variations and pitch targets in Cantonese," Proceedings of Speech Prosody 2006, Dresden, pp. 317–320.
[82]Wong, Y. W., and Xu, Y. (2007). "Consonantal perturbation of F_0 contours of Cantonese tones," Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, pp. 1293–1296.
[83]Xu, Y. (1997). "Contextual tonal variations in Mandarin," J. Phonetics **25**, 61–83.
[84]Xu, Y. (1998). "Consistency of tone-syllable alignment across different syllable structures and speaking rates," Phonetica **55**, 179–203.
[85]Xu, Y. (1999). "Effects of tone and focus on the formation and alignment of F_0 contours," J. Phonetics **27**, 55–105.
[86]Xu, Y. (2001). "Fundamental frequency peak delay in Mandarin," Phonetica **58**, 26–52.
[87]Xu, Y. (2005). "Speech melody as articulatorily implemented communicative functions," Speech Commun. **46**, 220–251.
[88]Xu, Y. (2006). "Principles of tone research," Proceedings of Second International Symposium on Tonal Aspects of Languages, La Rochelle, pp. 3–13.
[89]Xu, Y. (2008). "Timing and coordination in tone and intonation: An articulatory-functional perspective," Lingua (2008).
[90]Xu, Y., and Sun, X. (2002). "Maximum speed of pitch change and how it may relate to speech," J. Acoust. Soc. Am. **111**, 1399–1413.
[91]Xu, Y., and Wallace, A. (2004) "Multiple effects of consonant manner of articulation and intonation type on F_0 in English," J. Acoust. Soc. Am. **115**, 2317.
[92]Xu, Y., and Wang, Q. E. (2001). "Pitch targets and their realization: Evidence from Mandarin Chinese," Speech Commun. **33**, 319–337.
[93]Xu, C. X., and Xu, Y. (2003). "Effects of consonant aspiration on Mandarin tones," J. Int. Phonetic Assoc. **33**, 165–181.
[94]Xu, Y., and Xu, C. X. (2005). "Phonetic realization of focus in English declarative intonation," J. Phonetics **33**, 159–197.
[95]Xu, C. X., Xu, Y., and Luo, L.-S. (1999). "A pitch target approximation model for F_0 contours in Mandarin," Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, pp. 2359–2362.
[96]Zemlin, W. R. (1988). *Speech and Hearing Science: Anatomy and Physiology* (Prentice-Hall, Englewood Cliffs, NJ).

Effects of obstruent consonants on fundamental frequency at vowel onset in English[a)]

Helen M. Hanson[b)]
Department of Electrical and Computer Engineering, Union College, 807 Union Street, Schenectady, New York 12308

(Received 9 February 2007; revised 6 October 2008; accepted 13 October 2008)

When a vowel follows an obstruent, the fundamental frequency in the first few tens of milliseconds of the vowel is known to be influenced by the voicing characteristics of the consonant. This influence was re-examined in the study reported here. Stops, fricatives, and the nasal /m/ were paired with the vowels /i,ɑ/ to form CVm syllables. Target syllables were embedded in carrier sentences, and intonation was varied to produce each syllable in either a high, low, or neutral pitch environment. In a high-pitch environment, *F*0 following voiceless obstruents is significantly increased relative to the baseline /m/, but following voiced obstruents it closely traces the baseline. In a low-pitch environment, *F*0 is very slightly increased following all obstruents, voiced and unvoiced. It is suggested that for certain pitch environments a conflict can occur between gestures corresponding to the segmental feature [stiff vocal folds] and intonational elements. The results are different acoustic manifestations of [stiff] in different pitch environments. The spreading of the vocal folds that occurs during unvoiced stops in certain contexts in English is an enhancing gesture, which aids the resolution of the gestural conflict by allowing the defining segmental gesture to be weakened without losing perceptual salience.
 [DOI: 10.1121/1.3021306]

PACS number(s): 43.70.Fq, 43.70.Bk [CHS] Pages: 425–441

I. INTRODUCTION

It is generally accepted that in American English (and other languages), the fundamental frequency (*F*0) in a vowel following an obstruent consonant is influenced by the voicing characteristics of the consonant. As we review in Sec. I A, methods of study and interpretation of the results have evolved somewhat over the past 50 years. However, it is agreed that there is a voiced/voiceless dichotomy, with *F*0 at vowel onset being significantly higher following voiceless obstruents than following voiced. This dichotomy is often called "pitch skip," and we sometimes refer to it as "obstruent intrinsic *F*0" (or obstruent I*F*0) in this paper.

The pitch-skip phenomenon is worth studying for several reasons. Studies have shown that the dichotomy in *F*0 at vowel onset cues listeners to voicing characteristics of the consonant (Sec. I A 2), suggesting that pitch skip should be included in speech synthesis systems to make synthesized speech both more natural and more intelligible. Likewise, speech-recognition systems could benefit by accounting for this dichotomy. In addition to these applications, studies of pitch skip are interesting because of what they can tell us about models of speech production and sound change. In some languages, lexical tone is believed to have derived from voicing contrasts of obstruents, and pitch skip is offered as evidence for a phonetic basis of such tonogenesis [e.g., see Hombert *et al.* (1979)]. Jun (1996) argued that pitch skip has affected the implementation of phrase-level tones in Korean. In Sec. IV we propose that the interaction of pitch skip and phrase-level tones in American English provides evidence supporting a model of speech production suggested by Keyser and Stevens (2006).

In the next section, we review both production and perception studies related to the pitch-skip phenomenon. We also review theories of the source of this effect.

A. Background

1. Production data

House and Fairbanks (1953) averaged *F*0 over the vowels in symmetrical CVC syllables and found that this average *F*0 was higher when the consonant was voiceless than when it was voiced. They observed the time course of a subset of the *F*0 contours and noted that the *F*0 difference occurred at the start of voicing, rather than occurring uniformly throughout the vowel. Lehiste and Peterson (1961) measured the peak *F*0 of CVC syllables and found that the peak *F*0 in vowels following voiceless consonants was higher than that in vowels following voiced. They also noted that the time course of the *F*0 contour varied depending on whether the initial consonant was voiced or unvoiced: following voiceless consonants, the peak *F*0 in the vowel occurred immediately after the consonant, while following voiced consonants the peak occurred at about the middle of the vowel. In addition, it was found that final consonants did not have any consistent effect on the *F*0 of the preceding vowel.

While those earlier studies were on data produced by American English speakers, Mohr (1971) purposely used speakers who had mixed language backgrounds (three native speakers of Chinese, Russian, and German, with English as a

[a)]Portions of this work were presented at meetings the Acoustical Society of America held in Chicago, IL (June 2001) and Austin, TX (Nov. 2003).
[b)]Electronic mail: hansonh@union.edu

second language) in order to test the universality of the pitch skip. Mohr (1971) measured *F*0 at vowel boundaries and at minima or maxima. In this way, he was able to confirm the earlier studies, quantify their observations that the influence of the preceding consonant is limited to the early part of the vowel, and show that the phenomenon occurs in languages other than English.

Lea (1973) recorded two types of utterances from two male speakers of American English. The first were bisyllabic nonsense words *hə′*CVC, in which C and V included nearly all the consonants and vowels of American English. The second were pairs of bisyllabic words such as the noun and verb forms of "compact." Both types of utterances were recorded in isolation. For the nonsense syllables, in which the medial consonant always followed an unstressed syllable and preceded a stressed syllable, Lea (1973) observed what may be called a "rise-fall dichotomy;" that is, the *F*0 contour following a voiced obstruent rises from a lower value at voice onset to a higher value at midvowel, while the *F*0 contour following a voiceless obstruent slopes down from voice onset to midvowel. However, for the second set of recordings, in which the medial consonant could precede or follow a stressed syllable, the results were more complicated, and Lea (1973) concluded that whether the *F*0 contour rises or falls at a CV transition is an interaction of both stress and segmental context.[1]

Hombert (1978) also focused on the time course of pitch skip rather than on peak or average vowel *F*0. He obtained *F*0 contours for several speakers of American English and, just as earlier studies had shown, found that vowel *F*0 following voiceless stops was higher than that following voiced stops, with the greatest difference occurring at vowel onset. When the *F*0 contours were averaged across all speakers, he observed the rise-fall dichotomy. However, Hombert (1978) noted that there were individual differences in the details of the dichotomy. For example, the *F*0 contour following voiceless consonants did not slope down for all speakers. Hombert (1977, 1978) also observed pitch skip for two speakers of the tone language Yoruba. While the dichotomy in *F*0 following voiced v. voiceless stops was observed for all three tones (high, mid, and low) of that language, it was greater for the high tone, both in magnitude and duration.

In a study on American English, Ohde (1984) recorded data both in isolation and in carrier phrases. The stimuli were the six voiceless aspirated and voiced stops paired with five vowels (/i,e,u,o,ɑ/) to form symmetrical CVC syllables. He observed that although the *F*0 contour following voiceless stop consonants is higher than that following voiced stop consonants, it cannot be described as a rise-fall dichotomy. Rather, the *F*0 contour falls after both voiced and unvoiced consonants for his three male subjects.

Most of the studies described thus far have observed *F*0 on CVC syllables in isolation or embedded in carrier sentences, for example, "Say CVC again." That is, the target syllable is in focus and most likely carries the phrase-level prominence and a high-pitch accent. Kohler (1982) was one of the first to consider that pitch skip, which he referred to as a type of "microprosody," might interact with the phrase-level tone variations, or "macroprosody." He recorded German words with medial stop consonants embedded in falling, rising, and monotone *F*0 contours. When the *F*0 contour was rising or monotone, the vowel *F*0 preceding the stop consonant was not affected, while the vowel *F*0 following the consonant showed the expected dichotomy (being higher after voiceless stops than after voiced). However, when the *F*0 contour was falling, the usual dichotomy was observed on the preceding vowel rather than on the following. Kohler (1982) proposed that utterance-level intonation can sometimes cancel the segmental-level pitch-skip effects.

Silverman (1984) observed the effects of voiced and voiceless obstruents on both preceding and following vowels in Southern British English. The obstruents were embedded in three-syllable nonsense words (e.g., /ə′pi:pi:p/), which were embedded in carrier phrases. The subjects were instructed to place lexical stress on the middle syllable of the nonsense word. For two male subjects, Silverman (1984) observed an effect of consonant voicing on the *F*0 of vowels that both precede and follow the consonant, with stressed syllables displaying a greater effect. He also found that *F*0 falls after both voiced and voiceless obstruents, in line with the Ohde (1984) data on stops.

A likely explanation for the observation of a rise-fall dichotomy by some researchers, versus a "fall-fall" dichotomy by others was proposed by Kohler (1985). He noted that the utterance types used by Hombert (1978), for example, favor an F0 contour that rises and falls on the target word, while the utterances used by Ohde (1984), for example, will most likely be produced with an *F*0 contour that falls throughout the target word. Therefore, it is not surprising if the perturbations introduced to these globally differing contours appear to differ.

Jun (1996) reported data for American English, Korean, and French. In her study, the intonation contour was varied such that the target consonants (both voiced and unvoiced) either carried a nuclear pitch accent or occurred before or after the nuclear pitch accent. For American English, she observed a dichotomy similar to that described by Ohde (1984), with the effect being somewhat less in certain intonational contexts, confirming the assertion of Kohler (1982) that the intonational contour interacts with segmental effects on *F*0. For the French data, a dichotomy also occurred, but *F*0 tended to rise following the voiced consonants, rather than fall. Korean has three stop series (lenis, tense, and aspirated), none of which tend to be voiced (although the lenis stops are voiced in certain environments). Jun (1996) found that the *F*0 contours following tense and aspirated stops were similar to each other, while the *F*0 contour following the lenis stops was more similar to the sonorants /m,l/. In addition, the difference in *F*0 following the lenis v. tense and aspirated stops persisted throughout the vowel in certain prosodic contexts, unlike what has generally been observed in other languages. Jun (1996) saw this effect as evidence that pitch skip is being phonologized at a prosodic level.

Similarly, Jessen and Roux (2002) found that pitch skip following voiced and unvoiced stops and clicks in Xhosa persists at least to midvowel. They theorized that this *F*0 dichotomy is phonologized to cue the voicing distinction because the voiced stops typically do not manifest much actual

voicing during the stop closure, and there are no other phonetic cues to voicing. Their target words carried low lexical tone; it is possible that the pitch skip is extended to enhance the weak acoustic cues to the voicing distinction [e.g., see Keyser and Stevens (2006) and Kingston and Diehl (1994)] and may lead to an eventual tone splitting (Matisoff, 1973).

Most of the work we have summarized compared *F*0 contours for vowels following, and sometimes preceding, voiced and voiceless obstruents. Some studies have also compared *F*0 behavior following different series of voiceless obstruents that can occur in languages. Danish, for example, has two stop series, both of which are voiceless; one series is aspirated and the other is not. Jeel (1975) and Reinholt Petersen (1983) found that in Danish, the *F*0 contour following both voiceless series is usually raised compared with the *F*0 contour following sonorants and /v/, although individual speakers differ in the degree of the dichotomy. *F*0 following the aspirated series tended to be somewhat higher than that following the unaspirated series, although again, speakers varied somewhat. Ohde (1984) had a similar result for *F*0 following aspirated stops compared to *F*0 following the unaspirated voiceless stops in /s/-stop clusters in American English.

In a study of Korean stops, Han and Weitzman (1970) observed that the *F*0 contour following so-called weak stops starts at a lower frequency than that following the strong stops. These results are in line with the Jun (1996) results for lenis stops v. tense and aspirated stops, described earlier. However, neither Jun (1996) nor Han and Weitzman (1970) specified whether a consistent difference occurred between the aspirated and tense/strong stops. On the other hand, Xu and Xu (2003), in a study of Mandarin, compared *F*0 following voiceless aspirated and unaspirated stops, and found that *F*0 following the aspirated stops was *lower* than that following the unaspirated stops. Francis *et al.* (2006) found the same result for Cantonese. Thus, the results of studies comparing aspirated voiceless stops to unaspirated tend to show a dichotomy, but the direction of the observed dichotomy is not consistent.

As our literature review of production data shows, methods of studying the pitch-skip phenomenon have evolved over the years, from studies that observe the average or maximum *F*0 over an entire vowel to studies that observe the time course, and from studies that give little consideration to higher-level pitch phenomenon to studies that do consider the interaction of segmental, lexical, and phrase-level *F*0. While some details of the results of the summarized studies are not consistent, it is clear that there is often a dichotomy in the *F*0 contour on vowels following contrasting series of obstruent consonants, and this dichotomy occurs in most, if not all, languages that have been studied. The dichotomy is not so much one of the slopes of the *F*0 contour at vowel onset as much as it is one of the absolute difference in *F*0 over the first few tens of milliseconds of the vowel. Although most studies indicate that the effect of obstruent consonants on vowel *F*0 occurs for vowels that follow these consonants, some studies have found evidence suggesting that, in the right context, the *F*0 of preceding vowels can also be affected.

2. Perception data

It has been shown through perception experiments that perturbations to *F*0 associated with voicing contrasts of obstruents provide cues to a listener concerning the consonantal voicing characteristics. For example, Haggard *et al.* (1969) and Fujimura (1971) observed that listeners were more likely to perceive a synthesized stop consonant as being voiceless when it was followed by a high *F*0 than a low *F*0. Massaro and Cohen (1976) observed a similar result for synthesized fricatives. Furthermore, Whalen *et al.* (1993) reported that listeners use the *F*0 cue even when voice-onset time is unambiguous. Therefore, pitch skip is not just an interesting artifact but a cue that speakers can extend or exaggerate to strengthen the perceptual saliency of a voicing contrast.

Through careful manipulation of both the global *F*0 contour and the portion of the *F*0 contour local to stop consonants, Kohler (1985) used synthesized German words to show that *F*0 *prior* to a stop consonant can also influence whether the stop is perceived as being voiced or unvoiced.

3. Source of pitch skip

Theories as to why this dichotomy occurs fall into two camps. Kingston and Diehl (1994) suggested that speakers intentionally lower *F*0 following a voiced obstruent relative to *F*0 following voiceless obstruents to signal the [+voice] feature to listeners. On the other hand, the dichotomy may simply fall out from the physiology or aerodynamics of obstruent production. For example, Halle and Stevens (1971) suggested that *F*0 is modified by obstruents because the vocal folds are stiffened to inhibit glottal vibration during a voiceless obstruent and are slackened to facilitate glottal vibration during a voiced obstruent. These stiffening or slackening gestures then carry over into the adjacent vowel. This view gained support from Löfqvist *et al.* (1989), who found that cricothyroid (CT) muscle activity increases for voiceless consonants relative to voiced consonants. An increase in CT activity increases the longitudinal tension of the folds, which not only increases the frequency of vibration, but also inhibits voicing in certain situations [see, for example, Löfqvist *et al.* (1989)].

Hombert *et al.* (1979) discussed the possibility of an aerodynamic basis of pitch skip. Reduced transglottal pressure at stop release would result in a lowered *F*0 for voiced stops, while the increased airflow due to the spread vocal folds of voiceless unaspirated stops would change the Bernoulli effect such that *F*0 is increased. However, for English at least, the *F*0 perturbation can last up to 100 ms into the following vowel, and therefore Hombert *et al.* (1979) ruled out the possibility of an aerodynamic basis.

Kohler (1985), on the other hand, disagreed with this conclusion; in experiments with monotone recordings, he found that there was no perturbation prior to a voiceless stop consonant, but that there was a perturbation following such a consonant. Because speakers were carefully controlling *F*0 to be constant, he claimed that the perturbation following the consonant could not be due to a change in vocal-fold tension but would have to be due to an intrinsic aerodynamic property of stop releases. Note, however, that Kohler (1985) did

not deny that vocal-fold tension can play a role in pitch skip. Rather, he proposed that an effect of vocal-fold tension is superimposed on an aerodynamic component, and because the aerodynamic component is intrinsic, it always occurs. He claimed, however, that the vocal-fold tension component will be overridden in certain intonation environments, such as monotone speech production.

Based on their study of Cantonese, Francis *et al.* (2006) have also suggested that pitch skip is composed of both intrinsic and controlled components. In particular, they proposed that in Cantonese, speakers sharply reduce the extent to which vocal-fold tension in obstruents is allowed to overlap with following vowels, in order to preserve the tone of a syllable, but suggested that in English, speakers extend the influence of obstruents on following vowels as a way of enhancing a voicing contrast.

B. Current work

The work presented in this paper was undertaken as part of a project to improve the generation of *F*0 contours in rule-based speech synthesis by including physiological constraints implemented in the HLsyn speech synthesizer (Stevens and Bickley, 1991; Hanson and Stevens, 2002). In HLsyn there is a parameter dc (delta compliance) that can be manipulated to produce the effect of stiffening or slackening the vocal folds. This parameter tends to either inhibit or facilitate glottal vibration during the obstruent consonants, when there is a reduced transglottal pressure difference. A change in dc for the consonant can carry over into the following vowel, and a perturbation due to dc is superimposed on parameter *f*0 in the vowel (along with other *f*0 perturbations due to intrinsic vowel effects and subglottal pressure).

One of our goals was to quantify the shifts in *F*0 due to obstruent effects and the time course of these shifts for different speakers. Previous studies have compared *F*0 contours following voiced and voiceless obstruents but tend not to compare these to contours following a neutral segment, that is, a segment not expected to perturb the contour [but see Hombert (1978) and Löfqvist *et al.* (1995)]. Therefore, it is difficult to use these data to model precisely how much *F*0 is raised or lowered relative to its unperturbed value. A second goal was to compare the effects of stop consonants to those of fricatives; most previous studies have focused on stop consonants only. A third goal was to observe the interaction of phrasal intonation and obstruent intrinsic *F*0. Previous studies have mainly looked at syllables carrying high-pitch accents [but see Kingston and Diehl (1994) and Silverman (1986)]. However, if intrinsic obstruent *F*0 is due to CT activity during obstruent production, in certain contexts such activity would conflict with changes in vocal-fold tension necessary to vary intonation. In this case, obstruent effects on *F*0 might be more or less obvious in different *F*0 environments. On the other hand, if speakers are intentionally lowering *F*0 following voiced consonants, the obstruent effects on *F*0 should be similar in all pitch environments.

In the next section we describe the corpus, subjects, recording procedure, and signal-processing methods for the study, and in Sec. III we present the results. Section IV presents a discussion of the results and interprets them within the framework of featural representation and enhancement theory. Finally, we summarize our findings and present some ideas for future work in Sec. V.

II. DATA COLLECTION AND PROCESSING

A. Corpus

Target syllables were formed by combining the consonants and consonant clusters /m,v,t,f,s,b,d,g,p^h,t^h,k^h, sp,st,sk/ with /im/ and /ɑm/ to form 28 syllables. Shadle (1985) reported that vowel intrinsic pitch was dependent on sentence position. To test whether sentence position would influence obstruent I*F*0, the target syllables in this study were inserted into the carrier phrase "My (s)CVm is called my (s)CVm again." The target syllables in each sentence were identical. In order to get the target syllables produced in different *F*0 environments, it was intended (by the experimenter) that the subjects produce the sentences with three intonation contours and two phrase-level prominence[2] patterns:

(1) Both target words are prominent. We intended speakers to produce pitch-cued prominence on the target words, with high *F*0 on the first target word, followed by a fall in *F*0 to the second target word, followed by a rise in *F*0 through the last word ("again").
(2) As with (1), it was intended (by the experimenter) that the target words would be produced with pitch-cued prominence; however in this case *F*0 would be higher on the first "my" than on the first target word, followed by a rise in *F*0 to the second target word, followed by a fall in *F*0 through the last word ("again").
(3) Ideally, neither target word would be prominent; rather the *F*0 contour would remain flat throughout the sentence until the last word, when *F*0 would be increased, reaching a peak in the second syllable of "again," to result in pitch-cued prominence on that word.

Schematics of the intended intonation contours are illustrated in Fig. 1.

We are purposely avoiding the use of terms associated with a particular intonational labeling system, such as pitch accent, under the belief that it is not relevant whether subjects produce prominences that are judged by listeners to be pitch accented or not; what is important is that the words were produced when the subject was also producing either high, low, or neutral *F*0, for whatever reason.

In sum, the goal was that there would be 84 sentences in the corpus; that each target word would occur in either high, low, or neutral *F*0 environment; and that each target word would occur in either early or late sentence position. We say more about how subjects were prompted to produce the intended *F*0 contours and prominence patterns, and about how successful our attempts were, in Secs. II C and II D 3.

B. Subjects

There were five female and five male subjects, all adults and all native speakers of American English. Nine of the

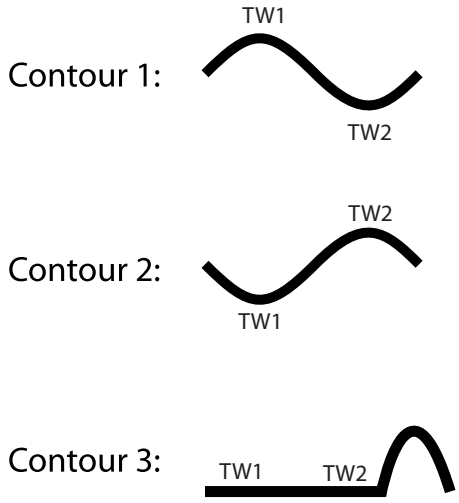


FIG. 1. Schematics of the *F*0 contours that the subjects were intended to produce. The abbreviations "TW1" and "TW2" refer to "target word 1" and "target word 2." See the text for more details.

subjects (including the author) were experienced as subjects in speech experiments (subject M5 was not experienced). Two of the subjects (F5 and M5) spent some of their early years in countries where English is not the native language but spoke English with no discernible accent. Except for the author, the subjects were paid for participating in the experiment.

C. Recording

Recordings were made in a sound-attenuated booth, either to digital audiotape (DAT) (48 000 samples/s) or directly to a hard drive (16 000 samples/s). Subjects were seated about 8 in. (20.3 cm) from an omnidirectional microphone. Prior to the recording, the experimenter demonstrated the intonation contours, and the subjects practiced producing a subset of the sentences until they were comfortable with the intonation contours and the target syllables. During the recording, each of the 84 sentences in the corpus was produced six times, yielding 504 utterances. To make the recording session easier for the subjects, blocks of 28 utterances were produced with a single intonation contour. The experimenter demonstrated the intonation contour at the beginning of a block, and the subject practiced it a few times before restarting the recording. Subjects could view a schematic of the desired contour throughout the block. Overall, subjects found contours 1 and 2 to be easy to produce, while contour 3 was somewhat unnatural. Subjects did not produce the same contour for consecutive blocks. Sentences were randomized within each block, and the randomizations were different for each subject. Care was taken so that a given target word occurred no more than once in the first or last position of a block. Subjects took a short break between each block and were offered the chance to take longer breaks as needed.

D. Processing

The data were downsampled to 10 000 samples/s using MATLAB software. Further processing was performed as follows.

1. Computation of *F*0 contours

The PRAAT software was used to label pitch periods. An autocorrelation-based algorithm included with Praat was used to estimate the pitch periods. These estimates were then corrected manually. The labels were converted to *F*0 contours and smoothed using software written in the C programming language (Xu, 1999). The smoothing algorithm compared each point in the *F*0 contour to its two neighbors; if the point was either higher than or lower than both of its neighbors (by 0.1 Hz or more), it was replaced by the average of the two neighbors. This method is effective for removing spikes that sometimes occur in *F*0 contours at the transition from one segment to the next. The smoothed F0 contours were then resampled at 5-ms intervals using linear interpolation.

2. Labeling of vowel onset

For each utterance, the starting time t_0 of the vowel in the target syllables was identified and tabulated. For the nasal and the voiced fricatives, t_0 was aligned with the release of the supraglottal constriction as identified by observations on the waveform and spectrogram. Specifically, for nasals we looked for changes in the waveform and spectrogram that indicated that formants higher than F1 were being more strongly excited; in the waveform, this stronger excitation manifests itself as higher frequency modulations on the F1 oscillations, and in the spectrogram it manifests as stronger energy at high frequencies. The CV transition for a nasal to a vowel was generally quite easy to label. For the voiced fricatives, we looked for points in the waveform where amplitude began to increase rapidly and the waveform became more periodic and less noisy. In the spectrogram these same properties were a rapid increase in energy and a switch from noise to harmonic excitation at high frequencies. In some cases the voicing ceased partway through the voiced fricatives. In these cases, t_0 was labeled at voice onset. For the stops and the unvoiced fricatives, t_0 was labeled at the onset of voicing. For voiced stops that were voiced throughout (meaning that voicing continued from the closure right through the release and into the vowel), t_0 was labeled at the point following the burst where the amplitude of voicing began to increase rapidly (similar to the way that voiced fricatives were labeled when they were voiced throughout).

3. Contour analysis

The *F*0 contours were evaluated to determine if subjects were successful in producing the target words in the desired *F*0 environments, and if they were consistent in doing so. This evaluation was purely visual and was performed by the author and some undergraduate students who assisted in the data processing.

The *F*0 contour for each utterance was examined and compared to the *F*0 contours for other tokens of the same sentence. Subjects were remarkably consistent with regard to utterance duration and shape of the *F*0 contour. Figure 2 shows the individual tokens for contours 2 and 3 of subject *F*1, target syllable "pom," limited to the target words in late sentence position. For contour 3, the target word was pro-

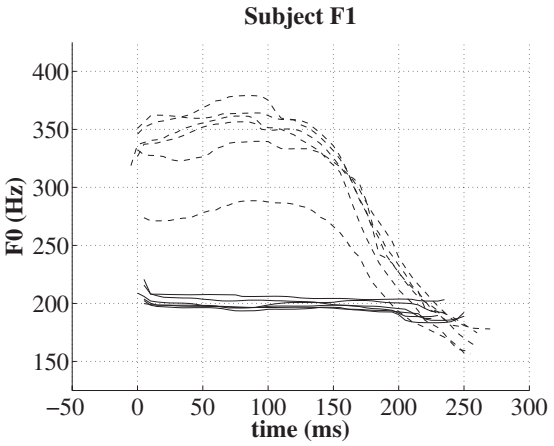


FIG. 2. *F*0 contours for the target syllable "pom" in late sentence position. Solid lines indicate contour 2 and dashed lines indicate contour 3. The speaker is subject F1.

duced in a neutral pitch environment, and it can be seen that there is little variation in average *F*0 among the tokens. For contour 2, the target word was produced in a high-pitch environment; here we observe a similar shape of the contours but great variability as to the maximum *F*0 at midvowel. This variability in maximum *F*0 was common for the words produced in high-pitch environment for both contours 1 and 2, and for all speakers. Note, however, that for this speaker, five of the contours cluster relatively closely together, while one of them deviates quite a bit from the others. Such outliers were not included in the analysis. Closer analysis revealed that token 1 was frequently an outlier, suggesting that the first round through each intonation contour was a kind of warmup for the speakers. In the later rounds the subjects were more comfortable with the target words and the intonation contours, and were therefore more consistent and natural in their productions. For this reason, the first repetition of each sentence (i.e., the first 84 sentences recorded) was not used for analysis, unless one of the later tokens (tokens 2–6) of the sentence was not usable. In the latter case, token 1 was used only if it did not deviate markedly from the other tokens. We note again that these judgments were purely subjective, based on visual inspection, but we are confident that we avoided the inclusion of outlier tokens.

Figure 3 compares intonation contour 1 for each speaker. Token 5 of the sentence "My mom is called my mom again" was used to generate the *F*0 contours. As intended, the first target word is produced with an *F*0 that is higher than that of the second target word. There is a great deal of variability among the subjects in the shape of the peak, the maximum *F*0, and the total *F*0 range. However, all the subjects seem to align their maximum *F*0 at or near the end of the vowel. Thus, all subjects produce a rising *F*0 on the vowel, but the degree of the rise varies greatly. Also, as intended, the second target word is produced with an *F*0 that is low in each speaker's pitch range, and that is lower than the word "again."

A similar plot for contour 2 is presented in Fig. 4. The *F*0 of the first target word is produced low in a speaker's pitch range, and *F*0 of the second target word is high in their range. Again, the maximum *F*0 is aligned at or near the end of the vowel, and there is great variety among the speakers as to the *F*0 range and shape of the peak. Finally, Fig. 5 shows examples of contour 3 for each speaker. The speakers all produced a level *F*0 contour up to the last word of the sentence, as hoped, but comparison with Figs. 3 and 4 show that *F*0 on the target words is about the same as the *F*0 produced in the low-*F*0 environment. Thus, we were not successful in getting an *F*0 that was intermediate between the high and low *F*0.

Subjects F5 and M5 were the least experienced at speech production experiments, but based on the contours in Figs. 3–5 their *F*0 contours are not grossly different from the more experienced subjects. In fact F5's contours are quite similar to those of F2, who is quite experienced. Similarly, M5's contours closely resemble those of M2, another well-experienced subject.

4. Averaging of F0 contours

For each subject, the ***F***0 contours for tokens 2–6 of each sentence were aligned at the sample occurring closest to t_0 (recall that the *F*0 contours were sampled at 5-ms intervals, while t_0 is essentially a continuous variable). The contours were averaged at points for which more than half of the contours had an *F*0 estimate (at other points the average *F*0 was discarded; these points were generally during obstruent closures when voicing ceased, but also sometimes at the ends of voiced segments when one token might be longer than others). Note that this alignment and averaging was done twice for each sentence: once for the first target syllable and once for the second target syllable. Informal observations of the averaged data for each sentence indicated that the data could be further reduced within a given intonation contour by averaging across place of constriction and across vowel. That is, while future work may show that there are significant differences in the magnitude of pitch skip due to vowel intrinsic pitch or to place of articulation, the general effect appears to be the same despite vowel or place, and we focus on that general effect in this paper. Averages across subject were also obtained, once for male subjects only and once for female subjects. For both of these additional types of average *F*0 contours—place/vowel and subject—averages were computed at points for which more than half of the contours had an *F*0 estimate. Thus, points in an average *F*0 contour across place and vowel include 5–10 data points for /m/, 10–20 for the fricatives, and 15–30 for the stops. Each point in an average *F*0 contour across subject includes 25–50 data points for /m/, 50–100 for the fricatives, and 75–150 for the stops.

5. Normalization of F0 contours

While *F*0 values at midvowel were not always observed to be the same across the voiced and unvoiced obstruents, they also did not seem to have systematic differences, as has been reported for some languages [e.g., see Jun (1996) and Cho *et al.* (2002)]. This observation leads us to believe that in general, English speakers have the same target *F*0 at midvowel for vowels following voiced and unvoiced obstruents. To prevent small differences in *F*0 at midvowel from obscuring or exaggerating obstruent effects at vowel onset, the *F*0

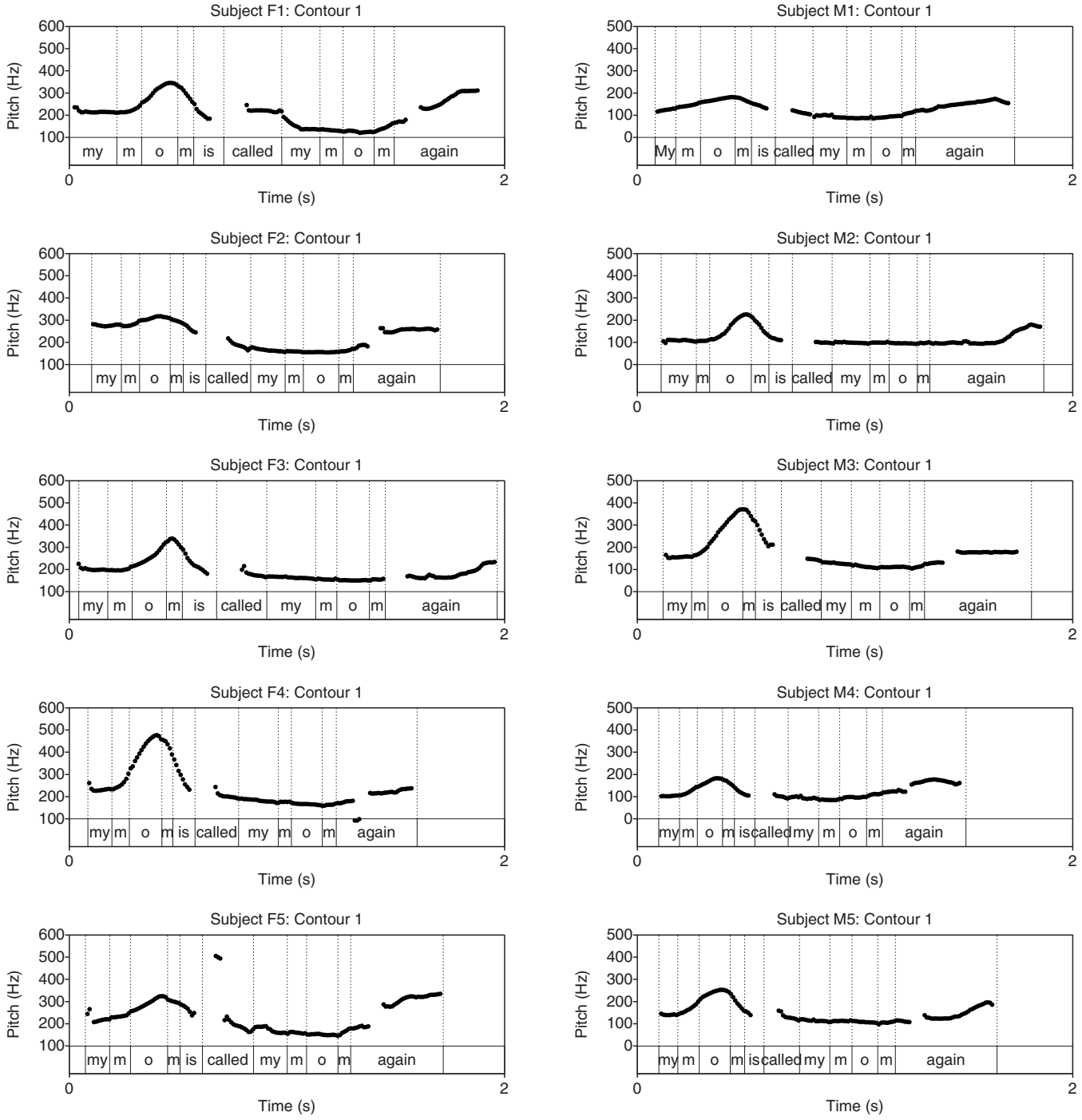


FIG. 3. Examples of the sentence "My mom is called my mom again" produced with intonation contour 1. The *F*0 contour for token 5 is shown for each subject. Female subjects are in the left column and male subjects are in the right column. (Note that these are "raw" *F*0 contours, which do not include the smoothing and manual corrections that were applied to the *F*0 contours used for the analysis.)

contours for the obstruent data were normalized in frequency to be similar to the *F*0 contours for the baseline /m/ data at midvowel. Normalized *F*0 at each point is

$$\widetilde{F0}(n) = F0(n) + N_{\text{nasal}} - N_{\text{obstruent}},$$

Where n=sample point, N_{nasal}=normalization factor for the baseline nasal, and $N_{\text{obstruent}}$=normalization factor for the obstruents.

The normalization factor N is obtained by averaging the *F*0 contour over a 50-ms window centered at either the peak *F*0 (most target syllables carrying a high *F*0) or 100 ms into the vowel (target syllables carrying low or neutral *F*0, for which the *F*0 contour tended to be flat, and target syllables carrying high pitch for which the *F*0 is flat or slopes down into the vowel). (The choice of 100 ms was based on our observations of where speakers tended to have their maximum *F*0.) Separate normalization factors were computed for contours averaged (1) across token, (2) across place, and (3) across subjects. Examples of this normalization are shown in Fig. 6.

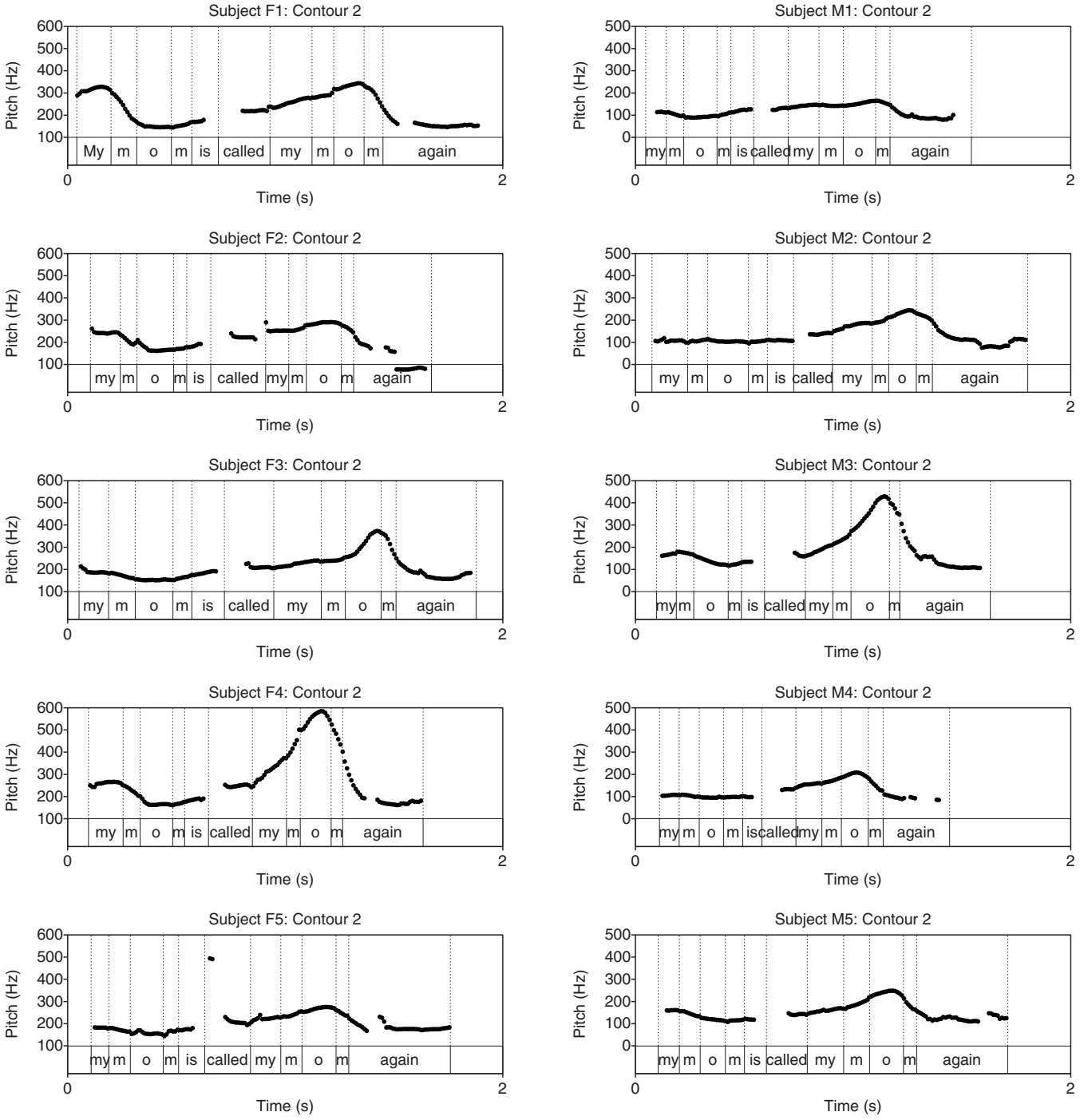


FIG. 4. Examples of the sentence "My mom is called my mom again" produced with intonation contour 2. The *F*0 contour for token 5 is shown for each subject. Female subjects are in the left column and male subjects are in the right column. (Note that these are raw *F*0 contours, which do not include the smoothing and manual corrections that were applied to the *F*0 contours used for the analysis.)

III. RESULTS

We first present results based on *F*0 contours averaged across subject (divided into male and female subjects). These results give us a general picture of the pitch-skip phenomenon. However, while all of the subjects exhibit the same general pattern, differences in the details of this pattern occur among them. Therefore, following the presentation of the general, across-subject results, we more closely examine the results for individual speakers.

A. Averages across subjects

Average *F*0 contours for the voiced and voiceless obstruents and the nasal consonant in the three pitch environments are summarized in Figs. 7–9.[3] Beginning with Fig. 7, the obstruents in high-pitch environment, we note that, as expected, *F*0 following the voiceless obstruents is raised relative to that of the voiced obstruents and the nasal consonant for about 100 ms into the vowel. However, *F*0 following the voiced obstruents tends to closely trace the baseline

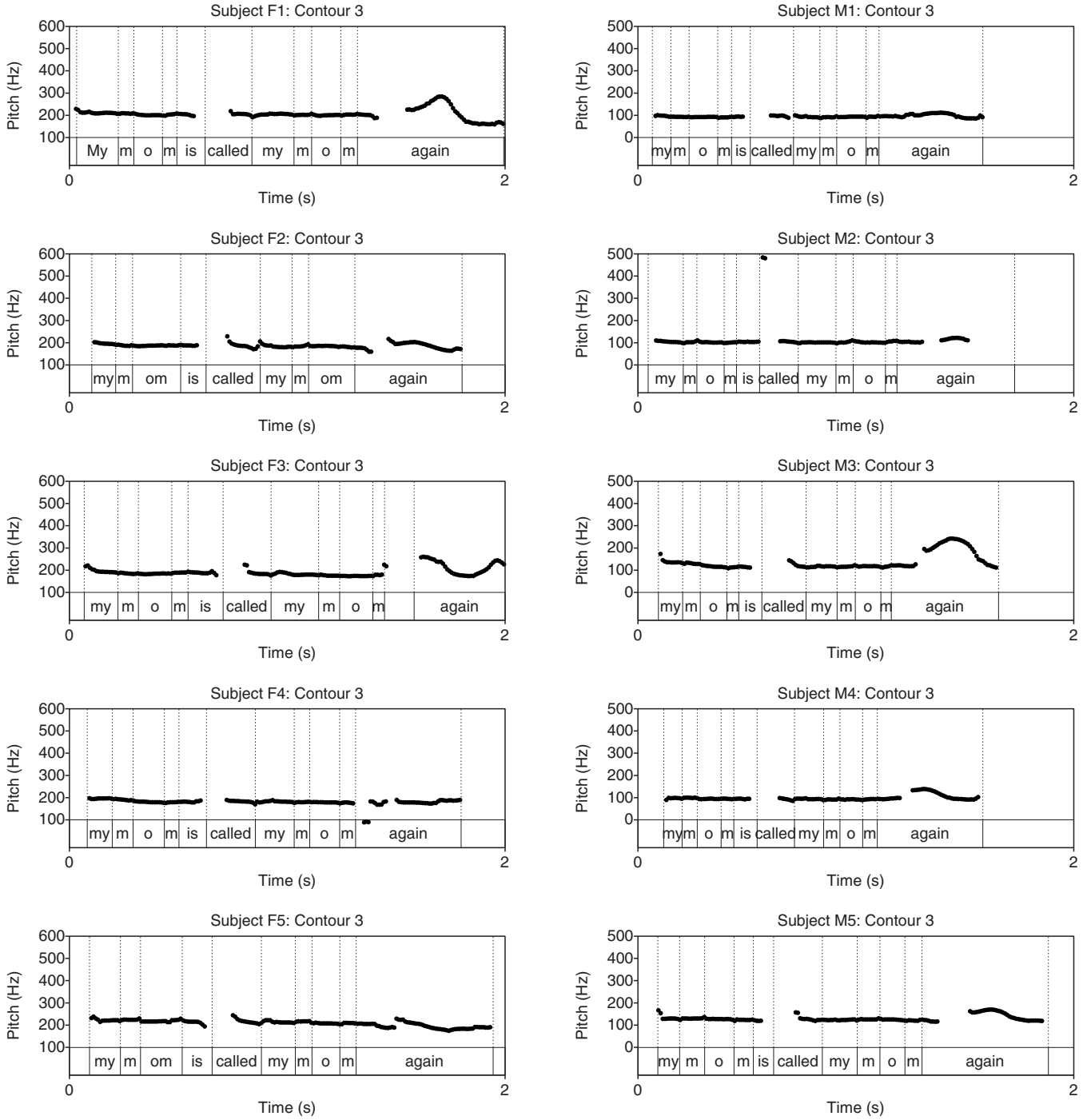


FIG. 5. Examples of the sentence "My mom is called my mom again" produced with intonation contour 3. The *F*0 contour for token 5 is shown for each subject. Female subjects are in the left column and male subjects are in the right column. (Note that these are raw *F*0 contours which do not include the smoothing and manual correction that were applied to the *F*0 contours used for the analysis.)

*F*0 following the nasal consonant, whereas one might have expected it to be lowered relative to the baseline (see Sec. I A 1). These results are consistent for both stop and fricative obstruents and for male and female speakers. Thus, assuming that the nasal consonant is truly neutral, it would seem that pitch skip is a phenomenon that applies only to voiceless obstruents and the vowels that follow them. We note also that there is little difference between the aspirated and unaspirated stop consonants.

We can also compare the results for the target syllables that occur early and late in the utterances. It appears that pitch skip occurs to a lesser degree for the target syllables late in the utterances. It is not clear if this effect is truly a function of syllable position because we note that the baseline *F*0 starts about 40 Hz higher for the late syllables than for the early syllables. For the female subjects, the maximum pitch in the target syllable is lower for the late syllable than for the early syllable. These two observations suggest that location in the pitch range or pitch excursion could also be a factor. We return to this point in the next section.

The *F*0 contours for the consonants in low-pitch environment (Fig. 8) seem to tell another story. There are few

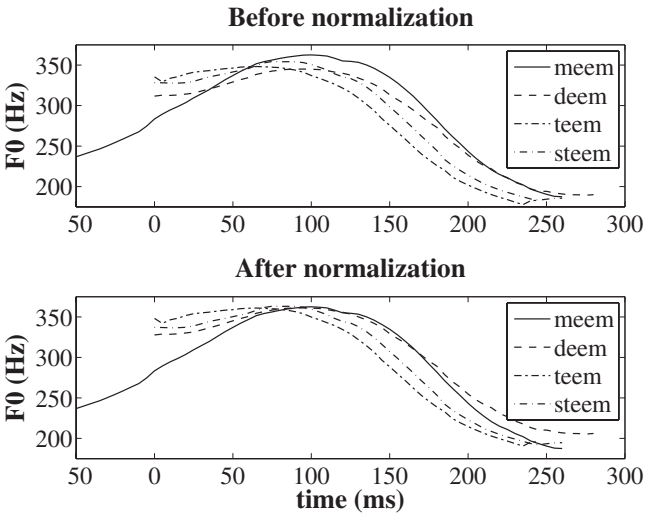


FIG. 6. An example of how averaged $F0$ contours for obstruents were normalized in frequency relative to averaged $F0$ contours for nasals (the baseline /m/ $F0$ contours). The example is for target syllables produced in a high-$F0$ environment by a female speaker. Top panel: averaged contours before normalization. Bottom panel: averaged contours following normalization. Note that $t=0$ in this and later graphs corresponds to the time t_0 of vowel onset for the CVm target syllables.

differences among the $F0$ contours for voiced and voiceless consonants, except in the first 10–15 ms, and those differences that occur are quite small. Again, this result is consistent for both male and female speakers and both stop and fricative consonants. The results for the consonants in the neutral pitch environment (Fig. 9) are nearly identical to those for the low-pitch environment. This similarity between the low and neutral $F0$ results is not surprising, given that the subjects produced similar $F0$ values for these target syllables. The contrast of the results for low and neutral $F0$ with those for the obstruents in high-pitch environment is quite striking. As Kohler (1982) suggested, there appears to be an interaction of segmental and phrase-level $F0$ effects; we discuss this result further in Sec. IV.

B. Comparisons among individual subjects

While the data for most of the subjects are qualitatively similar to the average $F0$ contours presented in Sec. III A, the details of the effects differ somewhat among the subjects. Although we do not think that these variations weaken the general result, it is worth acknowledging and describing them.

1. Degree of pitch skip in the three F0 environments

The general result is that pitch skip occurs in high-$F0$ environment for unvoiced obstruents only. However, some speakers deviate from this model in certain ways. For subject M1, the $F0$ contour following voiceless obstruents is only slightly higher than the baseline in all three $F0$ environments, as illustrated in Fig. 10 for fricatives early in the utterance. While it was generally true that $F0$ following voiced obstruents in a high-pitch environment closely followed the baseline $F0$ (following /m/), some subjects did exhibit an $F0$ that tended to be higher or lower than that following /m/ in certain environments. For example, subject F3's $F0$ following voiced obstruents early in an utterance tended to be lower than that following /m/, as illustrated in Fig. 11 for fricatives.

2. Aspirated stops v. unaspirated stops

Four of the subjects (F1, F2, F3, and M3) did show some difference between aspirated and unaspirated stops. For all four, $F0$ following aspirated stops tended to be somewhat higher than that following unaspirated stops when the target syllables occurred early in the utterance. Subjects F2 and M3 also exhibited this effect for the syllables occurring late in the utterances. Subject M3 displayed a particularly large effect, as seen in Fig. 12. Both the direction of the dichotomy and the variability among speakers are in line with previously reported data from Danish (Jeel, 1975; Reinholt Petersen, 1983) and American English (Ohde, 1984), described in Sec. I A 1.

3. Utterance position effect

In Sec. III A we mentioned that the degree of pitch skip in the high-$F0$ environment appeared to be stronger when the target syllables occurred early in the utterance than when they occurred later (see Fig. 7). Observation of the data for individual speakers suggests that most subjects do show some degree of change as a function of position (i.e., a smaller degree of pitch skip for late syllables), but the size of this change varies greatly across subjects. Figure 13 includes comparisons of $F0$ contours in early and late positions following fricatives for three subjects—M3 (no change with position), F1 (intermediate change with position, compared to other subjects), and M4 (extreme change with position, compared to other subjects). However, utterance position may not be the actual or only source of this apparent effect. We have observed for the baseline /mVm/ syllables that the maximum $F0$ in the vowel ($F0_{\max}$) tends to be lower in a subject's pitch range on the later syllables, while the $F0$ at vowel onset [$F0(t_0)$] tends to be *higher* in a subject's pitch range. That is, for syllables starting with /m/ in a high-$F0$ environment, the pitch excursion $\Delta F0_m = F0_{\max} - F0(t_0)$ is more compressed for the target syllables late in the utterances than early in the utterances. There is, in effect, less "room" for pitch skip to occur.

This effect is illustrated in Fig. 14. Figure 14(a) is a bar chart comparing the pitch excursions $\Delta F0_m^{\text{Early}}$ and $\Delta F0_m^{\text{Late}}$ (averaged across vowel) for the /mVm/ syllables in the early and late positions, respectively, for each subject. Only one subject (M3) has a larger pitch excursion for the late syllable than for the early syllable. Figure 14(b) is a bar chart showing the difference in pitch excursion $\Delta F0_m^{\text{Late}} - \Delta F0_m^{\text{Early}}$ for each subject. As expected from part (a), the difference is negative for all subjects but M3. The two subjects showing the smallest differences (F2 and M3) are also the subjects least likely (as judged by the author) to show a change in degree of pitch skip for early v. late utterance position. What is most striking about part (b) is the degree of variation among the subjects, from subject F4 whose pitch excursion is compressed by almost 120 Hz, to subject M3, whose pitch excursion expands slightly. We note also that these changes

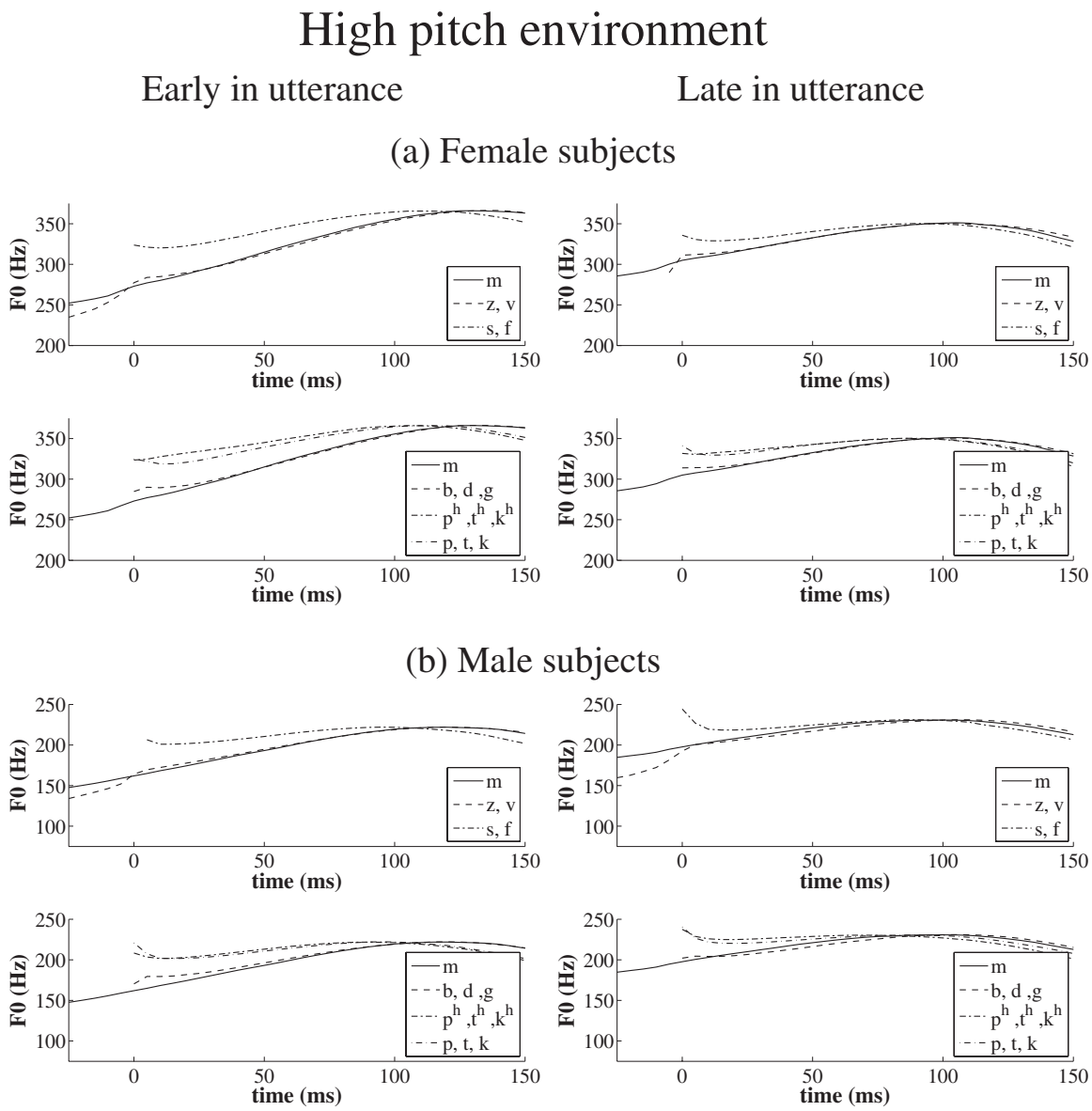


FIG. 7. *F*0 contours of target syllables in high-pitch environment, averaged across place of articulation and vowel for (a) female and (b) male subjects. The first column shows the contours for target syllables occurring early in the utterances, and the second column shows the contours for target syllables occurring late in the utterances. Each point in a contour includes 25–50 data points for /m/, 50–100 data points for the fricatives, and 75–150 for the stops.

in pitch excursion may be more perceptible for speakers with lower *F*0 ranges (100–150 Hz) than for speakers with higher-*F*0 ranges (200–400 Hz). Clearly there is a great deal more analysis that could be done to investigate the interaction of utterance position and pitch range in pitch skip.

IV. Summary and Discussion

The results can be broadly summarized as follows:

(1) When a (s)CVm syllable is in a high-pitch environment, *F*0 is greatly increased relative to a baseline *F*0 following voiceless obstruents, but *F*0 closely follows (or is only slightly higher than) the baseline following voiced obstruents.
(2) When a (s)CVm syllable is in a low-pitch environment, *F*0 is very slightly increased relative to a baseline following all obstruents.

These general results are consistent for both male and female subjects and both stop and fricative consonants. Details such as the degree of pitch skip (both in time and frequency) vary somewhat across subject, indicating that a model of this phenomenon should include speaker-specific parameters. In addition, an apparent effect of utterance position or pitch excursion was observed. Quantification of obstruent *F*0 then appears to be complicated, and must include considerations of speaker characteristics, prosodic stucture, and global as well as local *F*0 environments. Our original goal to quantify pitch shifts due to obstruent consonants turns out to have been somewhat naive. And yet, although we have not achieved that goal, our results have implications for models of speech production (Sec. IV A), which may be more relevant for improving speech synthesis than the original goal. Such an occurrence is common in research related to formant

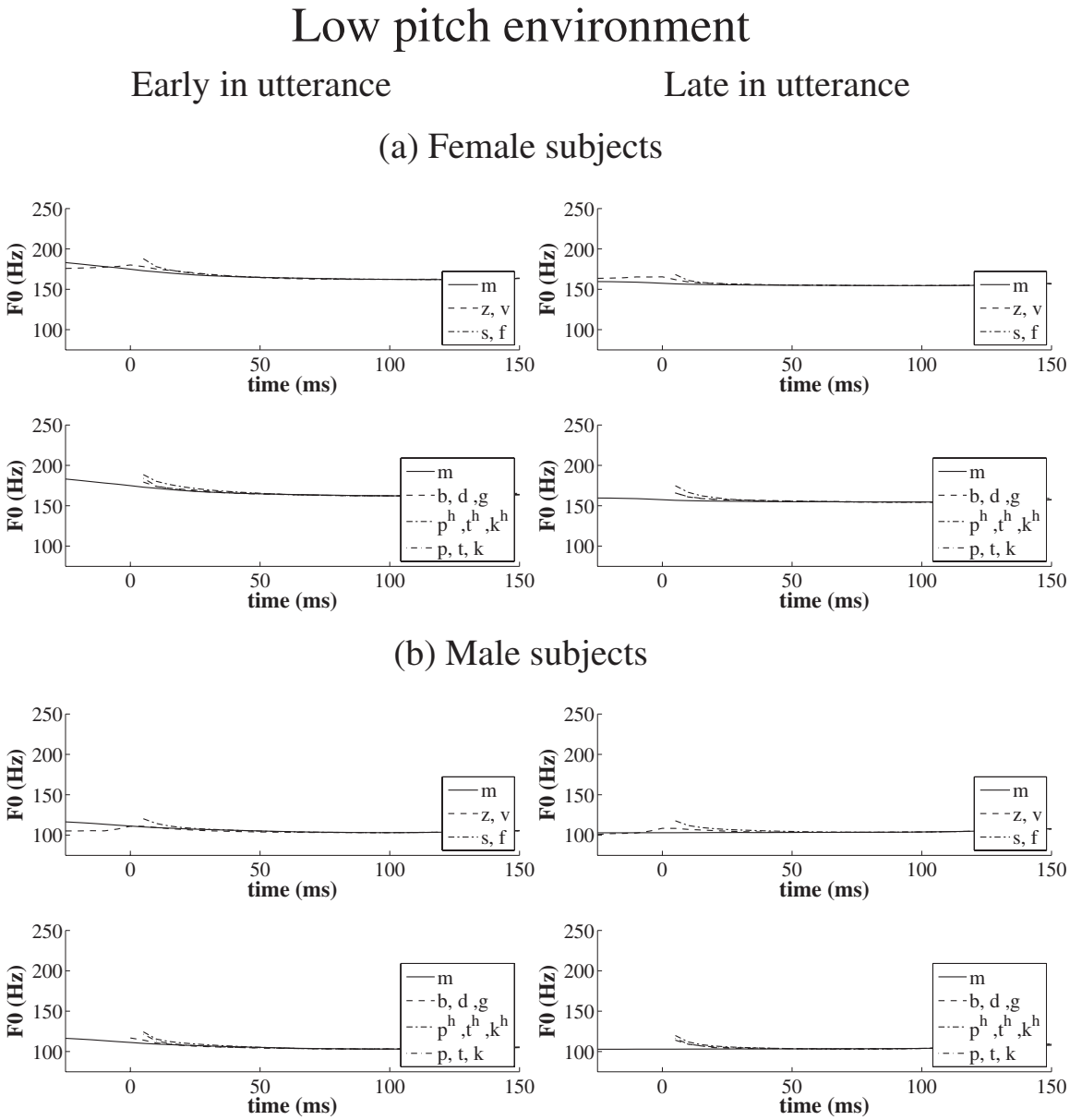


FIG. 8. *F*0 contours of target syllables in low-pitch environment, averaged across place of articulation and vowel for (a) female and (b) male subjects. The first column shows the contours for target syllables occurring early in the utterances, and the second column shows the contours for target syllables occurring late in the utterances. Each point in a contour includes 25–50 data points for /m/, 50–100 data points for the fricatives, and 75–150 for the stops.

or articulatory synthesis: insight is provided to issues of speech communication as synthesis is improved.

A. Interpretation of results

Because *F*0 following voiced stops is not lowered relative to that of /m/, it is probably not the result of a gesture intended to enhance the saliency of the [+voice] feature (Kingston and Diehl 1994). A more likely cause of the pitch skip observed in this data set is an increase in active vocal-fold stiffening during the voiceless obstruent consonants that carries over to the following vowel (Halle and Stevens, 1971; Löfqvist *et al.*, 1989), that is, an intrinsic effect that falls out from gestural overlap between a voiceless obstruent and a vowel that follows it. (We note, however, that an intrinsic effect can then be intentionally exaggerated by speakers to enhance the intended phonological contrast.) We propose that the observed difference between high- and low-pitch contexts occurs because of conflicts between the segmental and prosodic levels of speech production. The fact that we only observed a large effect in high-pitch environments is further support for the claim of Halle and Stevens (1971): high *F*0 in a vowel is in accord with vocal-fold stiffening in the preceding obstruent, so the gestural overlap in that case results in increased *F*0 at the vowel onset. However, when prosody demands low pitch in a vowel, a conflict with a stiffening gesture of a preceding voiceless obstruent arises; in such a case, the prosodic gesture trumps the segmental gesture, and the *F*0 contour is either not perturbed or perturbed to a degree that is probably not perceptible.

One might question the likelihood of a feature having a defining gesture that is so vulnerable to annihilation due to overlap with prosodic gestures. We suggest that this conun-

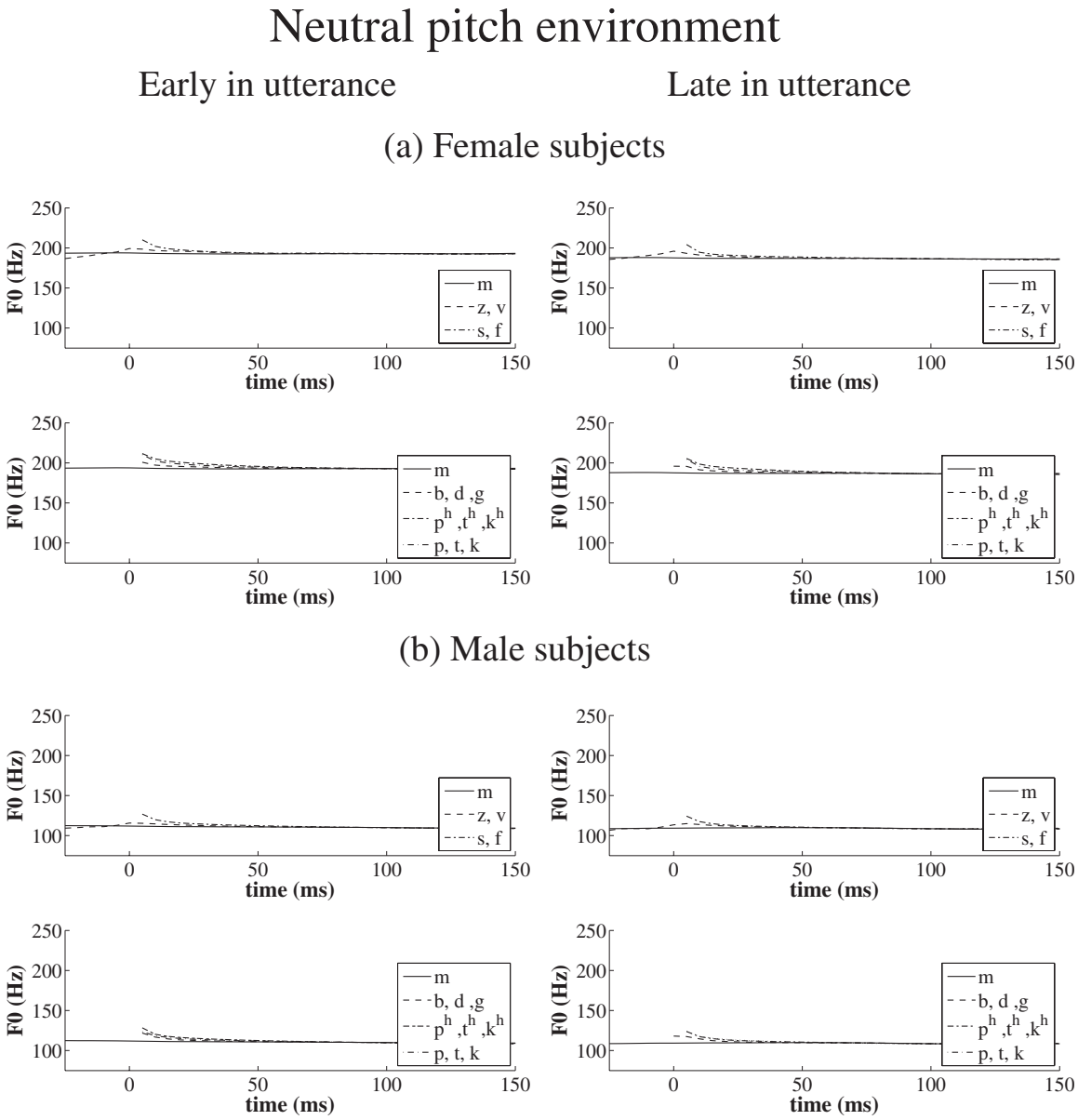


FIG. 9. *F*0 contours of target syllables in neutral pitch environment, averaged across place of articulation and vowel for (a) female and (b) male subjects. The first column shows the contours for target syllables occurring early in the utterances, and the second column shows the contours for target syllables occurring late in the utterances. Each point in a contour includes 25–50 data points for /m/, 50–100 data points for the fricatives, and 75–150 for the stops.

drum and our results can be explained within the framework of enhancement theory (Keyser and Stevens, 2006). According to this theory, a speech segment is defined as a bundle of binary features. A feature is defined as having a defining gesture and a corresponding defining acoustic characteristic. At the phonological planning stage, an utterance is comprised only of such bundles, but features are flagged if they are to occur in the context of other segmental features or prosodic elements that weaken their defining gestures and thus threaten the saliency of their defining acoustic characteristics. Features that are flagged as being vulnerable are reinforced with enhancing gestures. Enhancing gestures tend not to be subject to weakening or annihilation by overlap; they are specifically chosen to enhance the acoustic cues without fear of weakening. In such a scenario, our data can be explained as follows:

- In English, the voicing feature for obstruents is [stiff], and the defining gesture is stiffening of the vocal folds. This stiffening leads to inhibition of vocal-fold vibration during voiceless obstruents (Halle and Stevens, 1971). (Note that the same gesture during a vowel leads to an increase in *F*0.)
- The contrast between [+stiff] and [−stiff] is threatened when [+stiff] occurs in the context of a low *F*0 and when [−stiff] occurs in the context of a high *F*0. The contrast is thus made more salient through the use of enhancing gestures, specifically vocal fold spreading during voiceless obstruents (inhibiting vocal-fold vibration) and active expansion of the vocal tract during voiced obstruents (facilitating vocal-fold vibration) (Svirsky *et al.*, 1997). We note that spreading of the folds will be used as an enhancement

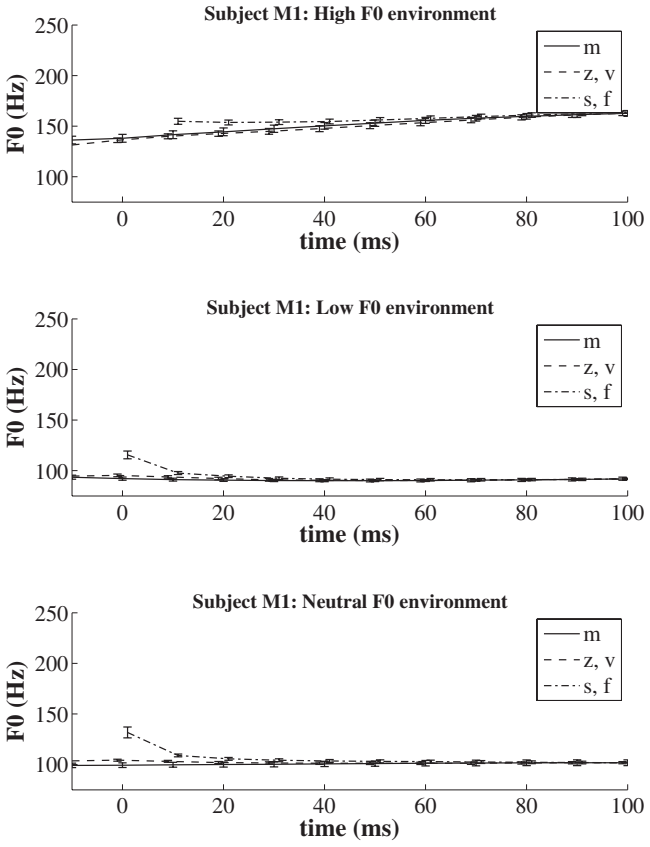


FIG. 10. *F*0 contours of target syllables occurring early in an utterance in high, low, and neutral *F*0 environment, for subject M1. Note that the voiceless obstruents exhibit a similar degree of pitch skip in all three *F*0 environments, contrary to the general results obtained by averaging across subject. Each point in a contour includes 5–10 data points for /m/ and 10–20 data points for the fricatives. Error bars indicate the standard error. (Contours for the obstruents are offset slightly along the time axis to improve clarity.)

for voiceless obstruents even in high-*F*0 environments. Likewise, vocal-tract expansion might occur even in low-*F*0 environments. While those claims may seem counterintuitive, the point of enhancement is to make *contrasts* more salient, and thus speakers may use enhancements for [+stiff] in an environment where [−stiff] is weakened, and vice versa.

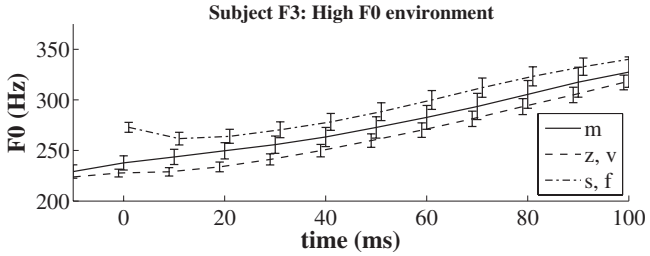


FIG. 11. *F*0 contours of target syllables occurring early in an utterance in high-*F*0 environment for subject F3. For this subject, *F*0 following the voiced obstruents in this environment tended to be low relative to the baseline contour following /m/. This result is contrary to the general result obtained by averaging across subjects. Each point in a contour includes 5–10 data points for /m/ and 10–20 data points for the fricatives. Error bars indicate the standard error. (Contours for the obstruents are offset slightly along the time axis to improve clarity.)

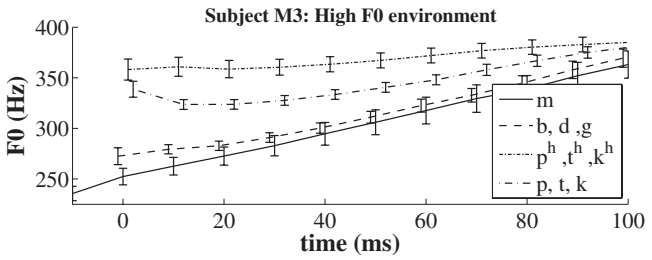


FIG. 12. *F*0 contours of target syllables occurring early in an utterance in high-*F*0 environment for subject M3. While most subjects did not show much difference between *F*0 following aspirated and unaspirated stop consonants, a few, such as this subject, exhibit a relatively strong difference. Each point in a contour includes 5–10 data points for /m/ and 15–30 data points for the stops. Error bars indicate the standard error. (Contours for the obstruents are offset slightly along the time axis to improve clarity.)

- Because of these enhancing gestures, gestures related to prosodic elements can override the defining segmental gestures in these cases, yet the saliency of the voicing contrast is preserved.

In contrast to features, which are believed to be universal in terms of both their defining gestures and acoustics, enhancing gestures are language specific (Keyser and Steven, 2006). Therefore, we emphasize that the proposed model of pitch skip is specific to English. Other languages may use the feature [stiff] contrastively; however, the acoustic manifestation could differ depending on how enhancing gestures are employed. For example, many tone languages include stop consonants, but the degree of observed pitch skip can vary greatly: Francis *et al.* (2006) reported that *F*0 perturbations are limited to the very early portion of the vowel in Cantonese (about 0–10 ms); Hombert (1977) reported perturbations that extend somewhat further into the vowel for Yoruba (about 50 ms for mid and low tones, and up to 100 ms for high tones); Kenstowicz and Park (2006) reported perturbations that extend at least until midvowel in Kyungsang Korean (specific times not provided). Francis *et al.* (2006) argued that Cantonese speakers intentionally limit the duration of pitch skip to protect the tone contrasts (see our interpretation below). On the other hand, the Kyungsang Korean speakers described by Kenstowicz and Park (2006) seem to have embraced the *F*0 perturbation as a means of shoring up the three-way unvoiced stop contrast, despite the competing demand of tonal contrasts on *F*0.

B. Implications for other studies

Our interpretation of the data within the theory of enhancement and overlap has implications for data reported in other studies. In Sec. I A 1 we described Korean data reported by Jun (1996). Although all three series of Korean stops are described as being unvoiced (underlyingly), an *F*0 dichotomy was observed at vowel onset following lenis stops, on the one hand, and tense or aspirated stops on the other. In our view, this finding suggests that the pitch-skip phenomenon is not due to voicing *per se* but to the manner in which voicing or its cessation is brought about. We would claim that a feature defining the lenis stops in Korean is [−stiff], while [+stiff] is a feature of the other two series of

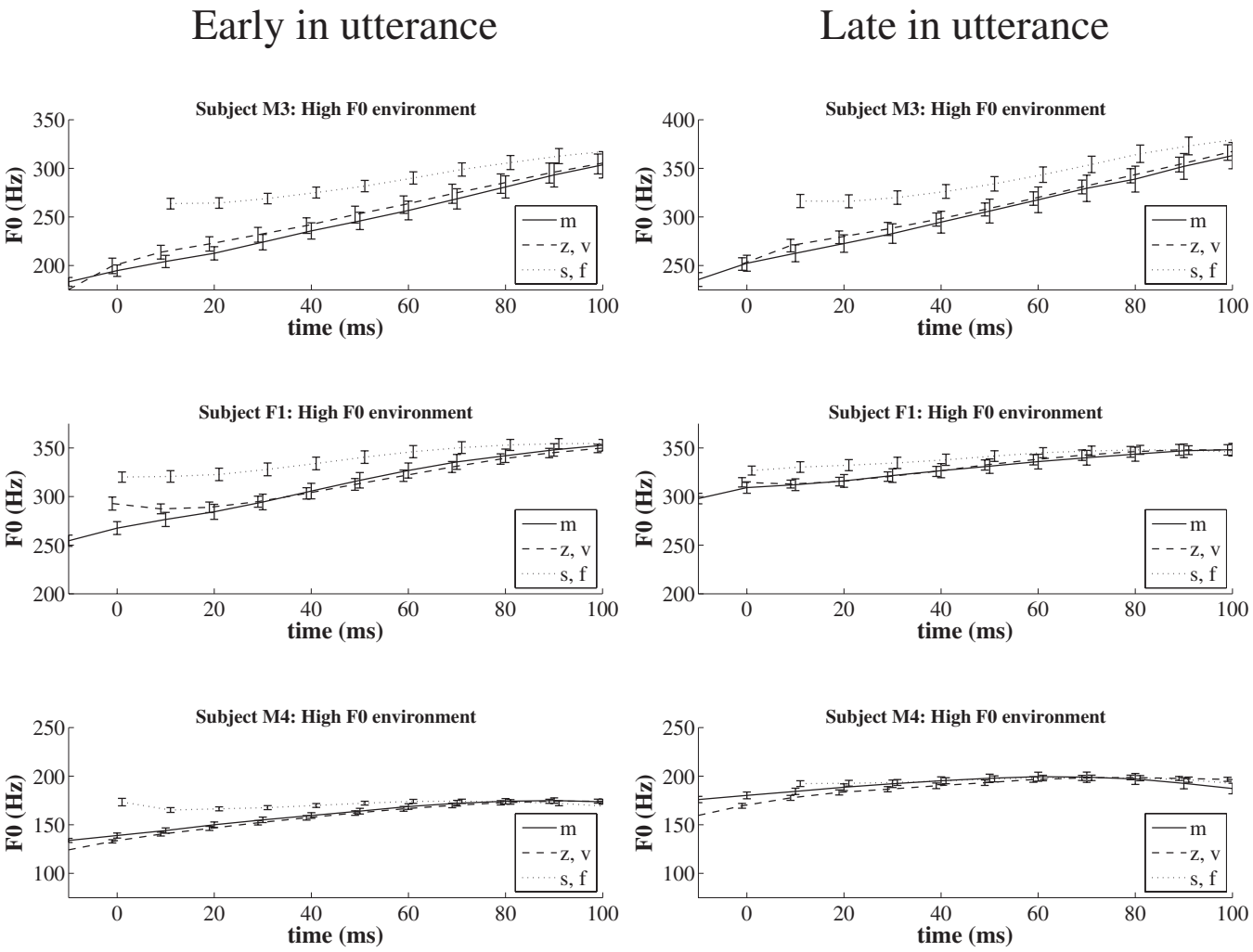


FIG. 13. *F*0 contours of target syllables occurring early and late in an utterance in high-*F*0 environment for subjects M3, F1, and M4. Subject M3 shows a similar degree of pitch skip across utterance position; subject F1 shows a reduced degree of pitch skip for syllables in late position, and subject M4 exhibits almost no pitch skip in late position. Each point in a contour includes 5–10 data points for /m/ and 10–20 data points for the fricatives. Error bars indicate the standard error.

stops. Further differentiation of the aspirated and tense stops would be due to the former having the feature [+spread], while the latter has the feature [−spread]. (The feature [spread] would not be defined for lenis stops.) Our interpretation of these Korean data is similar to the theory of Kim (1965) that proposes "tensity" as a feature independent of voicing and is similar in spirit to a proposal in Cho *et al.* (2002).

In their study of Cantonese, Francis *et al.* (2006) suggested that if pitch skip is intended to enhance a voicing contrast (Kingston and Diehl, 1994), it would be unlikely to be manifested in Cantonese because (1) there is no voice contrast and (2) it would interfere with lexical tone. While Francis *et al.* (2006) did observe a dichotomy between the aspirated and unaspirated stops in Cantonese, the effect was much smaller than in English, both in absolute difference and duration. This result was interpreted by them to indicate that while there was some intrinsic basis for pitch skip, the Cantonese speakers curtail this effect so as not to let it interfere with lexical tone. However, from our point of view, the feature that defines the contrast between aspirated and unaspirated stops in Cantonese would be [spread], not [stiff], and therefore, all else being the same, we would not expect to see a difference in the *F*0 of vowels following these obstruents. That Francis *et al.* (2006) did see a small effect may be due to other factors, such as aerodynamics, as suggested by Xu and Xu (2003) for Mandarin, or perhaps as a result of vocal-fold stiffening used to enhance voicelessness during the production of the unaspirated stops.

Similarly, our theory is in line with the observations of Kohler (1982) on medial stop consonants if one assumes that [stiff] is a feature for German stop consonants. When *F*0 is rising, the vowel *F*0 preceding the consonant is presumably low, and reduced vocal-fold stiffness during the vowel precludes the stiffening that might have occurred during the early phase of a voiceless consonant. Therefore, no effect of obstruent voicing is observed in the *F*0 of the preceding vowel. However, when *F*0 is falling, the relatively high *F*0 preceding the consonant does not conflict with stiffening of the folds early in the consonant and the pitch dichotomy is observed, while the drop in *F*0 for the following vowel masks the effects of stiffening later in the consonant.

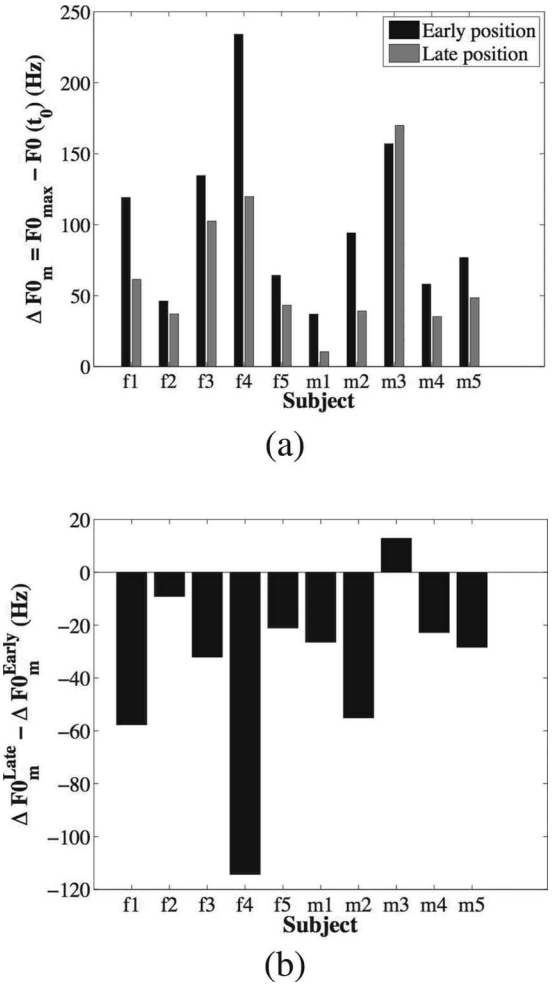


FIG. 14. Comparison of the pitch excursions $\Delta F0_m^{\text{early}}$ and $\Delta F0_m^{\text{late}}$ (averaged across vowel) for /mVm/ syllables in the early (dark bars) and late positions (gray bars). (b) Difference in pitch excursion $\Delta F0_m^{\text{late}} - \Delta F0_m^{\text{early}}$ for each subject. As expected from part (a), the difference is negative; i.e., pitch excursion is compressed in late position for all subjects but M3.

V. FUTURE DIRECTIONS

Certain predictions fall out from our model of pitch skip and are worth pursuing in future work. First, if implementation of a feature [stiff] is the source of pitch skip observed in most languages, the degree of pitch skip observed in vowels that either precede or follow unvoiced obstruents will depend on where *F*0 falls in a speaker's *F*0 range during those vowels. Therefore, a more complete test of this theory will observe obstruent effects on *F*0 contours that both rise and fall going into the obstruent. In addition, the effects should be observed on *F*0 contours that are neither rising nor falling, but rather constant at both low and high *F*0.

Likewise, our model predicts certain effects of *F*0 environment on the phonetic voicing of obstruents. For example, one might expect /b,d,g/ to show a lesser degree of vocal-fold vibration during the closure interval in a high-*F*0 environment than would be observed in a low-pitch environment. Preliminary data from two speakers support this prediction (Hanson, 2004).

As we have suggested, because pitch skip occurs consistently yet differently in the world's languages, studies of pitch skip can provide insight to models of speech production, speech perception, and sound change. Therefore, future studies of pitch skip will be most beneficial if they compare data across languages. Production and perception in bilingual speakers may be particularly insightful to studies of how different languages use enhancing gestures and acoustic cues to increase the saliency of universal feature contrasts.

ACKNOWLEDGMENTS

This work was carried out while the author was at the Speech Communication Group, MIT Research Laboratory of Electronics. The work was supported by NIH Grant Nos. DC04331 and DC00075. The programming assistance of Man Yin Yee is greatly appreciated. The PRAAT scripts for computing *F*0 contours and the C code for smoothing the contour were based on scripts and code provided by Dr. Yi Xu. The analysis and interpretation of the data have benefited from many conversations with Professor Kenneth N. Stevens. I thank Chris Shadle and three anonymous reviewers for their close reading of an earlier manuscript; their comments have greatly improved the paper.

[1]Note that lexical stress and utterance-level focus have been confounded in the Lea (1973) study, particularly for the second set of recordings. Therefore, it is not clear if segmental context interacts with stress or focus, or both. However, what is relevant for the discussion here is that segmental context appears to interact with suprasegmental elements that affect *F*0.

[2]By phrase-level prominence we mean what is often referred to as sentence stress or focus. Because these terms (especially the word *stress*) can mean different things to different readers, we use *prominence* to indicate that the speaker intended these words to be more prominent than the other words in a sentence, while not being specific about which acoustic cues were employed to signal that prominence.

[3]Error bars are omitted from these plots so as to avoid confusing variability among speakers for the variability within a given speaker.

Cho, T., Jun, S.-A., and Ladefoged, P. (**2002**). "Acoustic and aerodynamic correlates of Korean stops and fricatives," J. Phonetics **30**, 193–228.

Francis, A. L., Ciocca, V., Wong, V. K. M., and Chan, J. K. L. (**2006**). "Is fundamental frequency a cue to aspiration in initial stops?," J. Acoust. Soc. Am. **120**, 2884–2895.

Fujimura, O. (**1971**). "Remarks on stop consonants: Synthesis experiments and acoustic cues," in *Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen*, edited by L. L. Hammerich, R. Jakobson, and E. Zwirner (Akademisk Forlag, Copenhagen), pp. 221–232.

Haggard, M., Ambler, S., and Callow, M. (**1969**). "Pitch as a voicing cue," J. Acoust. Soc. Am. **47**, 613–617.

Halle, M., and Stevens, K. N. (**1971**). "A note on laryngeal features," Quarterly Progress Report No. 101 MIT Research Laboratory of Electronics, Cambridge, MA, pp. 198–213.

Han, M. S., and Weitzman, R. S. (**1970**). "Acoustic features of Korean /P,T,K/, /p,t,k/ and /p^h,t^h,k^h/," Phonetica **22**, 112–128.

Hanson, H. M. (**2004**). "The feature [stiff] interacts with intonation to affect vocal-fold vibration characteristics," J. Acoust. Soc. Am. **116**, 2546.

Hanson, H. M., and Stevens, K. N. (**2002**). "A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn," J. Acoust. Soc. Am. **112**, 1158–1182.

Hombert, J.-M. (**1977**). "Development of tones from vowel height?," J. Phonetics **5**, 9–16.

Hombert, J.-M. (**1978**). "Consonant types, vowel quality, and tone," in *Tone: A Linguistic Survey*, edited by V. A. Fromkin (Academic, New York), pp. 77–107.

Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (**1979**). "Phonetic explanations for the development of tones," Language **55**, 37–58.

House, A. S., and Fairbanks, G. (**1953**). "The influence of consonant environment upon the secondary acoustical characteristics of vowels," J. Acoust. Soc. Am. **25**, 105–113.

Jeel, V. (**1975**). "An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants," Technical

Report No. 9, Institute of Phonetics, University of Copenhagen, Copenhagen.
Jessen, M., and Roux, J. C. (**2002**). "Voice quality differences associated with stops and clicks in Xhosa," J. Phonetics **30**, 1–52.
Jun, S.-A. (**1996**). "Influence of microprosody on macroprosody: A case of phrase initial strengthening," Technical Report No. 92, University of California at Los Angeles, Los Angeles, CA.
Kenstowicz, M., and Park, C. (**2006**). "Laryngeal features and tone in Kyungsang Korean: A phonetic study," Studies in Phonetics, Phonology and Morphology **12**, 247–264.
Keyser, S. J., and Stevens, K. N. (**2006**). "Enhancement and overlap in the speech chain," Language **82**, 33–63.
Kim, C.-W. (**1965**). "On the autonomy of the tensity feature in stop classification (with special reference to Korean stops)," Word **21**, 339–359.
Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**, 419–454.
Kohler, K. J. (**1982**). "F_0 in the production of lenis and fortis plosives," Phonetica **39**, 199–218.
Kohler, K. J. (**1985**). "F_0 in the perception of lenis and fortis plosives," J. Acoust. Soc. Am. **78**, 21–32.
Lea, W. A. (**1973**). "Segmental and suprasegmental influences on fundamental frequency contours," in *Consonant Types and Tones*, Southern California Occasional Papers in Linguistics No. 1, edited by L. M. Hyman, (University of Southern California Press, Los Angeles), pp. 15–70.
Lehiste, I., and Peterson, G. E. (**1961**). "Some basic considerations in the analysis of intonation," J. Acoust. Soc. Am. **33**, 419–425.
Löfqvist, A., Baer, T., McGarr, N. S., and Story, R. S. (**1989**). "The cricothyroid muscle in voicing control," J. Acoust. Soc. Am. **85**, 1314–1321.
Löfqvist, A., Koenig, L. L., and McGowan, R. S. (**1995**). "Vocal tract aerodynamics in /aCa/ utterances: Measurements," Speech Commun. **16**, 49–66.
Massaro, D. W., and Cohen, M. M. (**1976**). "The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction," J. Acoust. Soc. Am. **60**, 704–717.
Matisoff, J. A. (**1973**). "Tonogenesis in Southeast Asia," in *Consonant Types and Tones*, edited by L. M. Hyman, Southern California Occasional Papers in Linguistics No. 1 (University of Southern California Press, Los Angeles), pp. 71–95.
Mohr, B. (**1971**). "Intrinsic variations in the speech signal," Phonetica **23**, 65–93.
Ohde, R. N. (**1984**). "Fundamental frequency as an acoustic correlate of stop consonant voicing," J. Acoust. Soc. Am. **75**, 224–230.
Reinholt Petersen, N. (**1983**). "The effect of consonant type on fundamental frequency and larynx height in Danish," Technical Report, Institute of Phonetics, University of Copenhagen, Copenhagen.
Shadle, C. H. (**1985**). "Intrinsic fundamental frequency of vowels in sentence context," J. Acoust. Soc. Am. **78**, 1562–1566.
Silverman, K. (**1984**). "*F*0 perturbations as a function of voicing of prevocalic and postvocalic stops and fricatives, and of syllable stress," in *Reproduced Sound: 1985 Autumn Conference, Windermere: Conference Handbook* (Institute of Acoustics, Windermere, Cumbria, Great Britain), Vol. **6**, pp. 445–452.
Silverman, K. (**1986**). "F_0 segmental cues depend on intonation: The case of the rise after voiced stops," Phonetica **43**, 76–91.
Stevens, K. N., and Bickley, C. A. (**1991**). "Constraints among parameters simplify control of Klatt formant synthesizer," J. Phonetics **19**, 161–174.
Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., and Wilhelms-Tricarico, R. (**1997**). "Tongue surface displacement during bilabial stops," J. Acoust. Soc. Am. **102**, 562–571.
Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1993**). "*F*0 gives voicing information even with unambiguous voice onset times," J. Acoust. Soc. Am. **93**, 2152–2159.
Xu, C. X., and Xu, Y. (**2003**). "Effects of consonant aspiration on Mandarin tones," J. Int. Phonetic Assoc. **33**, 165–181.
Xu, Y. (**1999**). "Effects of tone and focus on the formation and alignment of f_0 contours," J. Phonetics **27**, 55–105.

The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise

Kathryn Hopkins[a] and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 2 June 2008; revised 3 November 2008; accepted 6 November 2008)

Speech reception thresholds were measured with steady and amplitude-modulated noise maskers for signals processed to contain variable amounts of temporal fine structure (TFS) information. Subjects benefited more from TFS information for the modulated than for the steady masker. For both maskers, addition of TFS information up to 548 Hz improved performance, though the improvement was greater for the modulated masker. The addition of TFS information at higher frequencies improved performance further for the modulated masker only. These results are consistent with the idea that TFS information is important for listening in the dips of a fluctuating masker.
 [DOI: 10.1121/1.3037233]

PACS number(s): 43.71.Es, 43.71.An [RLF] Pages: 442–446

I. INTRODUCTION

For a given signal-to-noise ratio (SNR), normal-hearing subjects can identify speech better when it is presented in an amplitude-modulated noise background than in a steady noise background (Festen and Plomp, 1990). Similarly, speech reception thresholds (SRTs) for normal-hearing subjects are lower in modulated noise than steady noise (Duquesnoy, 1983; Peters *et al.*, 1998). This improvement in performance when listening in modulated noise has been termed "masking release." Masking release is thought to arise because subjects can listen to signal portions where the masker level is low and so the short-term SNR is high. This strategy has been labeled "listening in the dips."

For this strategy to be successful, temporal dips that contain information about the signal must be identified. One cue for identification could be a reduction in the modulation depth of the combined signal, though this cue is unlikely to be very salient, especially at low SNRs. Another possible cue is the change in temporal fine structure (TFS) in the dips of a masker when there is a signal present. This could allow identification of portions of the sound with the highest SNR. For this strategy to be successful, at least some of the TFS information about the target speech must be preserved, both in the physical signal and by the peripheral auditory system.

For subjects with cochlear hearing loss and cochlear implant users, masking release is often small or absent (Festen and Plomp, 1990; Peters *et al.*, 1998; Nelson *et al.*, 2003; Lorenzi *et al.*, 2006). Such subjects perform similarly when listening in steady and modulated noise. There are several possible reasons for this reduced ability to benefit from the dips in a masker.

(1) Hearing-impaired subjects often have broader auditory filters than subjects with normal hearing (Glasberg and Moore, 1986). Cochlear implants often have a small number of effective channels. This could reduce masking release in backgrounds that are modulated asynchronously across frequency, such as competing speech, because of increased spread of masking, which may mean that signals in the dips of a masker at one frequency are masked by masker components at nearby frequencies that are higher in level.
(2) Hearing-impaired subjects and cochlear implant users are often more susceptible to forward masking than normal-hearing subjects (Glasberg *et al.*, 1987). Peaks in a modulated masker would elevate the threshold in the following dips, making signal identification more difficult.
(3) Current cochlear implant processors discard much TFS information, and subjects with moderate cochlear hearing loss have a reduced ability to use TFS information (Lorenzi *et al.*, 2006; Hopkins and Moore, 2007; Hopkins *et al.*, 2008). TFS information may be important for identifying a signal in the dips of a fluctuating masker, as discussed above. This hypothesis is supported by data from Lorenzi *et al.* (2006). They measured the ability of hearing-impaired subjects to identify vowel-consonant-vowel (VCV) syllables processed to remove envelope information but to leave TFS information nearly intact. They found that the ability to identify these VCV syllables was correlated with masking release measured for intact VCV syllables in steady and amplitude-modulated noise.

Here we investigated the importance of TFS information for masking release, to test whether a reduced ability to use TFS information could explain the reduced masking release measured for hearing-impaired subjects and cochlear implant users. We used normal-hearing subjects and tone-vocoder processing, so that the effect of manipulating TFS information could be investigated without the possible confounding effects of changes in forward masking and frequency selectivity. The amount of TFS information that was preserved from the original signal was varied by vocoding information

[a]Author to whom correspondence should be addressed. Electronic mail: kh311@cam.ac.uk

above a "cut-off channel" (CO) but leaving lower frequency information intact (Hopkins *et al.*, 2008). The value of CO was varied, and performance was measured for both steady and amplitude-modulated noise.

II. METHODS

A. Subjects and materials

Ten normal-hearing subjects were tested. Subjects had thresholds of 15 dB hearing level (HL) or less at standard audiometric frequencies, and were paid for their time. The target speech for both training and testing was IEEE sentences (Rothauser *et al.*, 1969), spoken by a male native British English speaker. Sentences were presented in noise that had the same long-term spectrum as the target speech. The noise was either steady or amplitude modulated at a rate of 8 Hz. The modulation was sinusoidal on a decibel scale with a peak-to-valley ratio of 30 dB. The equation specifying the modulated noise was

$$F(t) = N(t) \times 10^{30[\sin(2\pi 8t)-1]/40}, \quad (1)$$

where $N(t)$ was the waveform of the unmodulated speech-shaped noise.

This combination of modulation rate and depth was chosen because pilot studies showed that these parameters led to a large amount of masking release with unprocessed stimuli, for the sentence material chosen for this study.

B. Processing and equipment

Processing was similar to that described by Hopkins *et al.* (2008). Speech and noise were summed at the required SNR and filtered into 32 channels using an array of finite-impulse-response filters, equally spaced on an ERB_N-number scale (Glasberg and Moore, 1990) between 100 and 10 000 Hz. The order of each filter was chosen so that its frequency response was approximately −6 dB at the point that the response intersected with the response of adjacent filters. Each filter was approximately 1-ERB_N wide, so that the filter bank roughly simulated the frequency selectivity of the normal auditory system. Channels up to and including a 'cut-off channel' (CO) were not processed further, whereas channels above CO were tone vocoded. The Hilbert transform was used to find the analytic signal for each channel signal and the envelope was calculated as the absolute value of each analytic signal. Each channel envelope was used to modulate a sine wave with frequency equal to the channel center frequency. Each modulated tone was subsequently filtered with the initial analysis filters. Consequently, envelope fluctuations with frequencies greater than half of the channel bandwidth were attenuated. The unprocessed channel signals and vocoded signals were time aligned and then combined. All signals were generated with a high-quality 16 bit PC soundcard (Lynx One) at a sampling rate of 22 050 Hz, passed through a Mackie 1202-VLZ mixing desk and presented to the subject monaurally via Sennheiser HD580 headphones. Subjects were seated in a double-walled sound-attenuating chamber. The target speech was presented at a constant level of 65 dB SPL, and the level of the noise was varied to give the required SNR.

C. Training

A period of training was conducted before testing to allow familiarization with the vocoder processing and the procedure. Eleven IEEE lists were presented with a fixed SNR for each list. The first list had CO=32 (corresponding to unprocessed signals) and steady noise, and each of the conditions was subsequently tested in a random order. For each list, the SNR was chosen to give performance of around 50%–80% correct. At the end of the training session, subjects completed two runs with the same procedure used in the testing session (described below) to allow familiarization with the adaptive testing procedure.

D. Testing

Five values of CO (0, 8, 16, 24, and 32) were used for each noise type (steady or modulated), making ten conditions in total. For each condition, the SNR needed to achieve 50% correct (the SRT) was measured using an adaptive procedure. Two IEEE sentence lists were presented consecutively to make a "run." The first sentence in each run was presented at an adverse SNR at which subjects were expected to identify no words correctly. The SNR was increased by 4 dB, and the sentence repeated until the subject correctly identified three or more of the five keywords. Subsequent sentences were presented once only. If the subject identified three or more keywords correctly, the following sentence was presented at a SNR that was k dB lower, and if two or fewer keywords were correctly identified, the following sentence was presented with a SNR that was k dB higher. For sentences before the first two turnpoints, k was equal to 4 dB; subsequently, k was equal to 2 dB.

E. Analysis

Within a run, for each SNR, the number of keywords presented and keywords correctly identified was recorded. These data were transformed to probit units (Finney, 1971), and linear regression analysis was used to estimate the slope and intercept parameters of the transformed data. The resulting probit function was then converted back to proportion correct units for each SNR and used as an estimate of the psychometric function for each subject and condition. The 50% correct points on these psychometric functions were identified to give SRTs that were used to compare performance across conditions.

III. RESULTS

Subjects performed better for both noise types as CO increased and more TFS information was available. The amount of benefit produced by the additional TFS information was found by calculating the improvement in performance (the reduction in SRT) for each value of CO relative to performance when CO=0, for each noise type. These results are plotted in Fig. 1. For both noise types, the largest benefit occurred when CO was increased from 0 to 8. However, for the modulated noise, performance improved further as more TFS information was added at higher frequencies. A within-subjects analysis of variance (ANOVA) on the SRTs

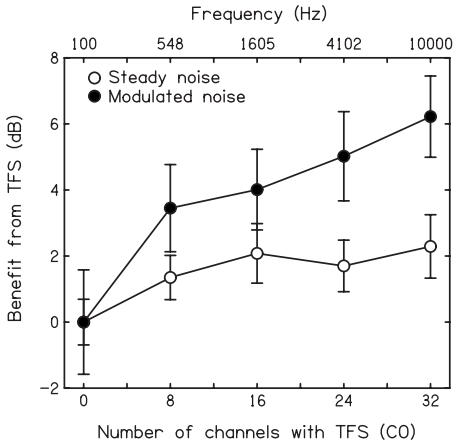


FIG. 1. Benefit of adding TFS information as measured by the SRT relative to that for CO=0, for steady and modulated noise. Mean SRTs for CO=0 for steady and modulated noise were −0.7 and −7.4 dB, respectively. The frequencies corresponding to each value of CO are shown along the top axis. Error bars show ± one standard error of the mean across subjects.

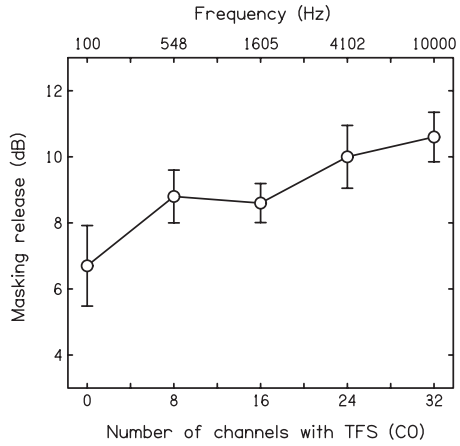


FIG. 2. Masking release plotted as a function of CO. The frequencies corresponding to each value of CO are shown along the top axis. Error bars show ± one standard error of the mean across subjects.

showed that the main effects of noise type and CO were highly significant ($p<0.001$) and that there was a significant interaction between the two factors [$F(4,36)=4.9$, $p=0.003$]. Fisher's least-significant-difference (LSD) test was performed to assess the effect of CO for each noise type. The results of this analysis are shown in Table I. Performance improved significantly for both noise types when CO was increased from 0 to 8. For the steady noise, no significant improvement in performance occurred as CO was increased further, but for the conditions where modulated noise was used there were significant improvements in performance for higher values of CO. For modulated noise, adding TFS information above 1605 Hz led to a significant improvement in the SRT. Thus, the benefit of TFS information is not restricted to the range covered by the voice fundamental frequency or its low harmonics. However, the improvement in SRT when the effective upper cut-off frequency for the TFS information was increased from 4102 to 10 000 Hz was small and nonsignificant, which is consistent with physiological data showing that the precision with which TFS information is represented in the auditory nerve weakens above about 5000 Hz.

TABLE I. Differences between mean SRTs measured with different pairs of values of CO for steady and modulated noise. Equivalent cut-off frequencies (in Hertz) are also shown. The LSD calculated using Fisher's procedure was 1.3. Differences equal to or above this value are shown in bold.

CO	CO Frequency	0 100	8 548	16 1605	24 4102	32 10 000
		Steady noise				
0	100	0				
8	548	**1.4**	0			
16	1 605	**2.1**	0.7	0		
24	4 102	**1.7**	0.3	−0.4	0	
32	10 000	**2.3**	0.9	0.2	0.6	0
		Modulated noise				
0	100	0				
8	548	**3.5**	0			
16	1 605	**5.0**	0.6	0		
24	4 102	**5.0**	**1.6**	1.0	0	
32	10 000	**6.2**	**6.2**	**2.2**	1.2	0

Figure 2 shows the masking release for each value of CO (SRT for steady noise minus the SRT for modulated noise). The amount of masking release increased as CO increased, and more of the original TFS was preserved. However, masking release still occurred when CO=0, when little or none of the original TFS information was preserved.

One concern with using the SRT as a measure of masking release is that psychometric functions measured using modulated noise are typically less steep than psychometric functions measured using steady noise (Qin and Oxenham, 2003; Stickney *et al.*, 2004), so the amount of masking release that is measured may depend on the percent correct that is tracked. Here, if a higher percentage correct was tracked, the masking release might have been smaller. Additionally, if the slopes of the psychometric functions changed for different values of CO, this might partially account for the different rates of change in SRT for steady and modulated noise as CO increased (as illustrated in Fig. 1).

To investigate these possibilities, a two-way within-subjects ANOVA was performed on the slopes of the fitted probit functions (as described earlier) for each subject and condition, with factors noise type and CO. The effect of noise type was significant [$F(1,9)=13.8$, $p=0.005$], but the effect of CO was not [$F(4,36)=0.15$, $p=0.96$], and there was no significant interaction between the two factors [$F(4,36)=0.24$, $p=0.91$]. The mean slopes, together with the associated standard deviations, are shown in Table II. The significant effect of noise type on the slopes of the psychometric functions is consistent with previous studies (Qin and Oxenham, 2003; Stickney *et al.*, 2004). The shallower slope of the psychometric functions for modulated noise than for steady noise means that the masking release measured in decibels by tracking the SNR needed to achieve a particular percent correct depends on the percent correct that is tracked; the higher the percent correct that is tracked, the less masking release is measured. The pattern of results that we report here (rather than the absolute values for masking release) should, however, be the same regardless of the percent correct that is

TABLE II. Mean and standard deviations of the slopes of the psychometric functions estimated for each noise type and value of CO.

	Steady noise		Modulated noise	
CO	Mean slope (probit units/dB)	Standard deviation	Mean slope (probit units/dB)	Standard deviation
0	0.39	0.23	0.20	0.11
8	0.41	0.13	0.23	0.09
16	0.40	0.19	0.24	0.11
24	0.40	0.17	0.25	0.11
32	0.38	0.20	0.26	0.12

tracked, as there was no significant effect of CO on the slopes of the psychometric functions, and no significant interaction between noise type and CO.

Psychometric functions are plotted in Fig. 3, based on the mean slopes and intercepts across subjects of the probit functions for each condition. From these psychometric functions, we derived a second measure of masking release by finding the difference in percent correct performance predicted by these psychometric functions for a SNR at which performance in the steady noise condition was predicted to be 10% correct for each value of CO. This gives an estimate of the masking release in percentage points, rather than decibels, so allowing comparison of our data with previous studies that used this measure. The masking release values estimated in this way are shown in the second column of Table III. The values increased progressively from 48% to 67% as CO was increased from 0 to 32. Similarly, the benefit of additional TFS information was estimated in percentage points, by measuring the improvement in performance in steady and modulated noise as CO increased for a SNR for which performance was estimated to be 20% correct when CO=0. These results are shown in the third and fourth columns of Table III. The maximum benefit (obtained with CO=32) was 31% for the steady noise and 47% for the modulated noise. The general pattern of results from these

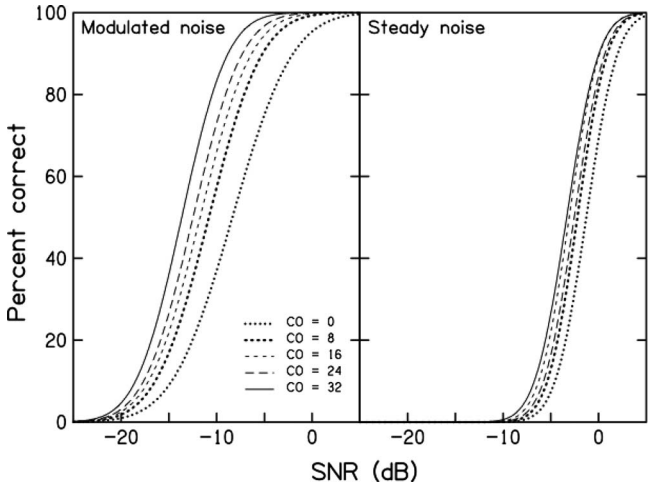


FIG. 3. Psychometric functions for each value of CO, for steady and modulated noise. The plotted functions are based on mean values of the slopes and intercepts of the probit functions fitted to the data for each subject and condition.

TABLE III. Masking release and benefit from additional TFS information expressed in percentage points. Masking release was calculated from the difference in psychometric functions in steady and modulated noise, at a SNR for which performance in steady noise was predicted to be 10% correct. The benefit from additional TFS as CO increased was measured by finding the difference in the psychometric functions at a SNR for which performance when CO=0 was predicted to be 20% correct.

		Benefit	
CO	Masking release	Steady noise	Modulated noise
0	48	0	0
8	59	14	19
16	60	27	28
24	65	19	35
32	67	31	47

analyses is consistent with that inferred from the SRTs; masking release increased as the number of channels containing TFS information increased, and the benefit of adding TFS information was greater for modulated than for steady noise.

IV. DISCUSSION

The greater benefit from the addition of TFS information for the modulated masker than for the steady masker, and the increased masking release as CO increased both suggest that TFS information is particularly important for listening in a modulated background. The same patterns of results were found when differences in performance were expressed as changes in SRT and as predicted differences in percent correct performance at a particular SNR. TFS information could be important for identification of signals in masker dips, thus directing attention to the portions of a signal with the most favorable SNR.

Significant (though reduced) benefit was also gained from TFS information at low frequencies when listening in steady noise, suggesting that TFS information is not important only for identification of signals in masker dips. TFS information could also allow better fundamental frequency identification. Current pitch perception models are based on the assumption that TFS information is important for low, resolved harmonics, and it has been suggested that TFS information may also be important for pitch perception based on harmonics with intermediate harmonic numbers that are unresolved (Moore *et al.*, 2006). Fundamental frequency may be an important grouping cue, allowing separation of the target and background into different auditory streams. Information about the fundamental frequency of the target speech would mainly be present at low frequencies, where there are resolved harmonics. The improvement in performance as TFS information was added up to 548 Hz, but not for higher frequencies (for the steady noise), is consistent with this idea.

It is possible that the vocoder processing used in the present study may have had an adverse effect on the temporal envelope information available to the central auditory system, as well as on the TFS. The widths of the analysis filters were chosen to be similar to those of the normal au-

ditory filters, and so auditory filters centered at the analysis channel center frequencies should have had similar temporal envelope cues at their outputs for the vocoded and unprocessed signals. However, only 32 contiguous filters were used, while the peripheral auditory system can be thought of as an array of many overlapping filters. Consequently, the temporal envelope information at the outputs of auditory filters centered between the analysis filters in the vocoder-processing scheme would have been somewhat degraded. This could partly account for the higher SRTs when CO was low.

Previous studies measured masking release for cochlear implant users and found little difference in performance when listening in steady or modulated noise (Nelson *et al.*, 2003). Similar results were found when using vocoder simulations of cochlear implant processing with normal-hearing subjects (Qin and Oxenham, 2003). This is in contrast with the present study, where 6.7 dB of masking release was measured when CO=0, and the signal was entirely vocoded. A possible explanation for this difference is the number of channels used in the vocoder processing. In this study, 32 1-ERB_N-wide channels were used to simulate the frequency selectivity of the normal auditory system. Cochlear implants have fewer effective channels, either because only a few electrodes are used or because of current spread between electrodes. Similarly, previous vocoder simulations often used fewer channels than the current study. The current data suggest that masking release can occur with little or no TFS information, provided that frequency selectivity is normal. This is consistent with the results of Gnansia *et al.* (2008), who measured the intelligibility of VCV signals in steady and modulated noise, processed to remove TFS information. As in the current study, many channels were used, so frequency selectivity was similar to that of the normal auditory system. Gnansia *et al.* (2008) found that performance was about 24% higher when identifying VCV syllables in modulated noise than in steady noise when the TFS information was removed. A larger improvement was seen when the stimuli were unprocessed, but the difference in masking release was smaller than measured in the present study. Similarly, Qin and Oxenham (2003) measured a small amount of masking release for sentences processed with a 24-channel vocoder (2.1 dB lower SRT for a modulated than for a steady noise masker) and Fu and Nogaki (2005) measured a small amount of masking release for sentences processed with a 16-channel vocoder, but no masking release when the number of vocoder-processing channels was reduced to four. The smaller masking release values reported in these previous studies probably reflect differences in the maskers used. The current study used a noise that was sinusoidally amplitude modulated on a decibel scale, whereas Gnansia *et al.* (2008) used noise that was sinusoidally amplitude modulated on a linear scale. Noise that is amplitude modulated on a decibel scale has more of its energy concentrated at the peaks of the waveform, with longer effective dips. This may give more opportunity for "dip listening" and consequently more masking release.

V. CONCLUSIONS

Adding TFS information to vocoder-processed speech improved performance more when listening in modulated noise than in steady noise, suggesting that TFS information contributes to masking release for normal-hearing subjects. A loss of ability to use TFS information in subjects with cochlear hearing loss and subjects with cochlear implants could partially account for the reduced masking release observed for such subjects.

ACKNOWLEDGMENTS

This work was supported by the MRC (UK). We thank Associate Editor Richard Freyman and two anonymous reviewers for helpful comments on earlier versions of this paper.

Duquesnoy, A. J. (**1983**). "Effect of a single interfering noise or speech source on the binaural sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74**, 739–743.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Finney, D. J. (**1971**). *Probit Analysis* (Cambridge University Press, Cambridge).

Fu, Q. J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

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., Moore, B. C. J., and Bacon, S. P. (**1987**). "Gap detection and masking in hearing-impaired and normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1546–1556.

Gnansia, D., Jourdes, V., and Lorenzi, C. (**2008**). "Effect of masker modulation depth on speech masking release," Hear. Res. **239**, 60–68.

Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "The effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech.," J. Acoust. Soc. Am. **123**, 1140–1153.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (**2006**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480–490.

Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.

Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "I.E.E.E. recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **AU-17**, 227–246.

Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Lexical and indexical cues in masking by competing speech

Karen S. Helfer and Richard L. Freyman
Department of Communication Disorders, University of Massachusetts Amherst, 358 North Pleasant Street, Amherst, Massachusetts 01003

(Received 21 March 2008; revised 30 October 2008; accepted 2 November 2008)

Three experiments were conducted using the TVM sentences, a new set of stimuli for competing speech research. These open-set sentences incorporate a cue name that allows the experimenter to direct the listener's attention to a target sentence. The first experiment compared the relative efficacy of directing the listener's attention to the cue name versus instructing the subject to listen for a particular talker's voice. Results demonstrated that listeners could use either cue about equally well to find the target sentence. Experiment 2 was designed to determine whether differences in intelligibility among talkers' voices that were noted when three utterances were presented together persisted when each talker's sentences were presented in steady-state noise. Results of experiment 2 showed only minor intelligibility differences between talkers' utterances presented in noise. The final experiment considered how providing accurate and inaccurate information about the target talker's voice influenced speech recognition performance. This voice cue was found to have minimal effect on listeners' ability to understand the target utterance or ignore a masking voice.
 [DOI: 10.1121/1.3035837]

PACS number(s): 43.71.Es, 43.71.Bp, 43.66.Dc [MSS] Pages: 447–456

I. INTRODUCTION

Over the past several years, there has been increased interest in studying how listeners are able to understand one talker in the presence of competing conversations. Many of the studies examining this ability have, to some measure, attempted to quantify the relative contributions of two types of masking involved in these listening situations: energetic masking and informational masking. Energetic masking is interference that is produced when a competing signal uses peripheral resources that are necessary to process the target. Informational masking is generally thought to be caused by confusion between the target and masking signals and/or uncertainty regarding the target. Speech maskers have the potential to produce both of these types of masking. This paper will describe a new set of sentence stimuli (the TVM sentences) designed for research on energetic and informational masking in competing speech paradigms and will discuss results of three studies using these sentences.

A. Rationale for the development of the TVM sentences

Many recent investigations of speech perception in a competing speech environment have used the coordinated response measure (CRM) corpus (Bolia *et al.*, 2000) (e.g., Arbogast *et al.*, 2005; Kidd *et al.*, 2005a; Kidd *et al.*, 2005b; Wightman and Kistler, 2005; Brungart *et al.*, 2006; Rakerd *et al.*, 2006; Brungart and Simpson, 2007). These sentences have a number of features that make them particularly useful for studying informational masking. First, each CRM sentence begins with, "Ready *cue name*…," where the cue name is one of eight possible names such as Ringo or Baron. This provides a means of orienting the listener to the target sentence. Second, analyses of error patterns in studies using the CRM sentences suggest that, at least when the target and masker are presented at approximately the same level, informational masking (rather than energetic masking) limits performance (e.g., Brungart *et al.*, 2001). Moreover, because of their closed-set nature and because of the independence of the colors and numbers, the CRM sentences can be used multiple times within and across test sessions with no risk of the listener learning specific stimuli.

While these sentences have proven to be extremely useful for studying a variety of characteristics of speech-on-speech masking, communication interactions outside of the laboratory often involve understanding messages that are not restricted to a small set of alternatives. Given the same set of target stimuli, closed-set tasks are considerably easier than open-set tasks (e.g., Sumby and Pollack, 1954) and, in many instances, do not accurately simulate the demands of finding the target word in lexical memory (e.g., Sommers *et al.*, 1997; Clopper *et al.*, 2006). Although open-set speech materials are available [e.g., the Harvard IEEE corpus (Rothauser *et al.*, 1969), the hearing in noise test (HINT) sentences (Nilsson *et al.*, 1994), and the BKB (Bamford–Kowal–Bench) sentences (Bench *et al.*, 1979)], none of these stimulus sets has the feature of a cue word, which is desirable when conducting studies in competing speech situations. This suggests the need for a large open-set corpus of stimuli that retains the cue word feature.

This paper will introduce the TVM sentences, a new open-set stimulus corpus designed specifically for research questions involving competing speech, and will describe three studies using these stimuli. The focus of these studies was on what cues listeners are able to use to identify and attend to a target message in the presence of competing messages. The primary experimental conditions presented a target TVM sentence from one talker in the presence of two masking TVM sentences with different cue names spoken by other talkers. As with the CRM corpus, if subjects are told the target cue (which, for TVM sentences, is the name Theo,

Victor, or Michael) before each trial, they can use that information to find and follow the target message. However, in order to do this, listeners must (1) hear the cue name and (2) somehow connect the key words in later parts of the sentence back to the cue name.

How do listeners make that connection? The most obvious possibility would seem to involve matching the voice reciting the cue name to the voice reciting the rest of the sentence, possibly assisted by additional matching of cue and key word loudness levels in cases where there are large target-to-masker level differences. Brungart *et al.* (2001) found that performance on the CRM stimuli was quite poor when target and masker were from the same talker, particularly at 0 dB target-to-masker ratio and below. Successful selective listening in competing speech may depend to a great degree on following a voice over time. Target voice information may therefore provide as much information as a semantic cue at the start of the target utterance.

B. Voice characteristics and speech masking

In order to successfully negotiate the challenge of listening to one utterance in the presence of other streams of speech, one must first determine the source of the target message and then attend to that signal while ignoring or deselecting the other messages. Faced with this task, an individual may use either semantic information (e.g., a certain word or phrase he/she is trying to find within a mixture of voices) or indexical information (the voice of the person to whom one wants to attend) to find the target utterance. Most of the research paradigms examining speech-on-speech masking use task instructions that direct the listener to a semantic cue, often a key word within the target signal.

Very little attention has been focused on the extent to which listeners can use indexical information in the talker's voice to identify and selectively attend to a target utterance. Experiment 3 in Brungart *et al.* (2001) examined whether knowledge of the target talker's voice (conveyed by blocking trials by the target talker) aided listeners' understanding of CRM sentences. On each trial subjects were instructed to listen for the sentence beginning with "Baron," which, in a given block, was always spoken by the same talker. They found that this indexical information provided no additional benefit besides cueing the listener to the sex of the target talker. That is, when the lexical cue name was available, subjects did not benefit from knowing the target talker's voice if the target and maskers were from same-sex talkers. This is consistent with the idea that the relative importance of any one cue used to distinguish a target utterance from a masking utterance likely depends on which other cues are available (e.g., Kidd *et al.*, 2005a). It is possible that listeners might have benefited from the voice or indexical information if the lexical cue name was not also presented.

It is well established that the amount of masking (both energetic and informational) produced by speech is related to the similarity of the speech target to the speech masker. In most situations, same-sex maskers produce greater masking (especially informational masking, which is related to confusion between the target and the masker) than do maskers produced by individuals differing in sex from the target talker (e.g., Festen and Plomp, 1990; Brungart, 2001; Brungart *et al.*, 2001; Darwin *et al.*, 2003). Moreover, intelligibility of a target utterance in the presence of a speech masker can be enhanced by introducing fundamental frequency differences between the masker and the target (e.g., Brokx and Nooteboom, 1982; Darwin *et al.*, 2003). Few studies have sought to quantify and account for the differences in masking effectiveness when the target and masker are spoken by people of the same sex. Some data suggest that certain voices appear to be more resistant to same-sex speech masking than others (e.g., Brungart, 2001). A recent study from our laboratory found substantial differences in the amount of informational masking produced by various combinations of two-talker female maskers (Freyman *et al.*, 2007). It is reasonable to assume that at least some of the variability in the amount of informational masking produced by specific voices is related to the listener's ability to differentiate the target voice from the masking voice(s).

Research has clearly demonstrated that voice information and phonetic information from an utterance are intertwined and are processed together. For example, a number of studies have documented a reduction in speech understanding ability when there is uncertainty regarding the talker from trial to trial as compared to performance measured when the same talker is used throughout the experiment (Mullennix *et al.*, 1989; Nygaard *et al.*, 1995). Performance on speeded classification tasks (where one must attend to a talker's voice while ignoring the lexical content of the message) shows that listeners cannot ignore irrelevant changes in one of these dimensions while attending to the other (e.g., Mullennix and Pisoni, 1990). Another piece of evidence of the connectivity of voice and phonetic information comes from investigations of the influence of talker familiarity on speech recognition. These studies show that listeners find it easier to perceive words in noise that are spoken by familiar (versus unfamiliar) talkers (e.g., Nygaard *et al.*, 1994). Hence, voice information appears to be an important component of speech recognition and, as such, may play a significant role in how well listeners can understand speech in adverse listening conditions.

One purpose of the studies described in this paper is to examine how listeners use lexical and indexical information in a competing speech situation. It has been decades since Broadbent (1952) reported that listeners can use information about the talker's voice to help resolve a single-talker-interference task. Since then, surprisingly little attention has been devoted to examining the degree to which listeners can use voice information in a competing speech task as a cue to identifying the target talker. The fact that trial-by-trial target talker uncertainty has a detrimental effect on speech perception in the presence of speech maskers (e.g., Brungart *et al.*, 2001) suggests that information about the target talker's voice is somehow used by listeners in competing speech situations.

This paper will first describe the development of the TVM sentence corpus. We then will summarize the results of three studies using the TVM sentence stimuli. The first study examined how the type of cue to which the subject was

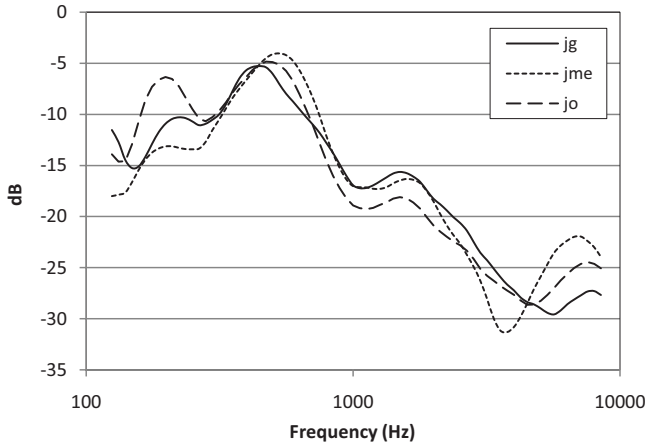


FIG. 1. Long-term third-octave smoothed spectra derived from a sample of 20 sentences from each talker.

asked to attend (key word or target talker voice) influenced speech recognition. Since differences in intelligibility between talkers' utterances were noted in the first study, a second experiment was conducted to determine the source of this talker variability. In this study, each sentence was presented in steady-state noise, and data analysis focused on identifying differences in intelligibility among the three talkers' recordings. The third study was designed to further examine the role of exposure to a voice on the ability to attend to or ignore a target or masking talker.

II. DEVELOPMENT OF THE TVM SENTENCE CORPUS

Each of the stimuli in the TVM corpus has the format, "*Name* discussed the ____ and the ____ today," where *Name* is the cue name *Theo*, *Victor*, or *Michael* and blanks correspond to one- or two-syllable nouns used for scoring. The cue names were chosen to be distinctive in both the auditory and the visual (i.e., lipreading) domains. The scoring words were common nouns, most of which were taken from the Thorndike–Lorge lists (Thorndike and Lorge, 1952). A total of 1080 unique sentences were created (360 beginning with each of the three cue names), with no scoring word repeated across the corpus. Examples of TVM stimuli are, "Theo discussed the swamp and the whisper today" and "Victor discussed the plant and the book today."

The sentences were recorded onto digital videotape from three male talkers who had no discernible regional dialect. Each talker was recorded saying each of the 1080 sentences. Recordings were generated in a sound-treated audiometric chamber. A remote microphone (Shure MX 183) was clipped to the talker's shirt approximately 6 in. below the mouth. The output of the microphone was routed to a preamplifier (PreSonus TubePre) and then sent to a digital video camera (Panasonic PV-DV953). Talkers were instructed to speak in a conversational manner but to attempt to put equal emphasis on the cue name and the two scoring words within each sentence. The studies described in the present paper used only the audio portions of these stimuli. Long-term third-octave smoothed spectra derived from the same 20 sentences from each talker are displayed in Fig. 1.

Stimuli were transferred to a PC for editing, storage, and presentation. Each sentence was saved in a separate file and scaled to produce utterances equal in rms amplitude. Three college-aged adults independently verified that the cue name and scoring words in each sentence were intelligible when heard in quiet. Sentences deemed unacceptable were discarded and then re-recorded.

III. EXPERIMENT 1: TASK INSTRUCTIONS AND SPEECH-ON-SPEECH MASKING

This study examined performance on the TVM corpus in a speech-on-speech masking task where listener instructions were manipulated. Specifically, on some trials subjects were instructed to repeat the sentence beginning with one of the cue names; on other trials subjects were given a preview of the target voice and told to repeat the sentence spoken by that talker. The primary purpose of this experiment was to compare listeners' use of these two types of cues (semantic and indexical) in both spatially coincident and spatially separated listening conditions.

A. Procedures

On a given trial, three sentences were presented simultaneously, each beginning with a different cue name and spoken by a different talker. One of these sentences was designated as the target sentence, while the other two were maskers. All three sentences began simultaneously and ended at approximately the same time (depending on the specific sentence length).

Testing took place in an IAC sound-treated room that has been used in previous competing speech experiments (Helfer and Freyman, 2005; Freyman *et al.*, 2007). The reverberation time in this chamber ranges from 0.12 s in the high frequencies to 0.24 s in the low frequencies (Nerbonne *et al.*, 1983). Our previous studies in this room have demonstrated similar amounts of spatial release from masking to that obtained in an anechoic chamber (e.g., Freyman *et al.*, 1999).

On half of the trials, the target and masking sentences were presented from a front loudspeaker located at 0° azimuth and at a distance of 1.3 m from the subject's head and at a height of 1.2 m from the floor (ear height for the average adult when seated); this spatial condition is hereafter referred to as F-F (for front-front). On the other trials (F-RF for front–right front), the target sentence was presented from the front loudspeaker, while the masking sentences were presented from the front and from a loudspeaker located 60° to the right of the subject at the same distance and height as the front, with a 4-ms time lead favoring the right. Due to the precedence effect, this spatial configuration produces the perception of the masker being located toward the right, well separated from the target. Comparing data obtained in the F-F and F-RF configurations allows us to identify the relative contributions of energetic and informational masking, as will be described in Sec. III B.

Target and masking sentences were mixed by summing the respective digital waveforms and then were presented from the computer's sound board. The stimuli were attenu-

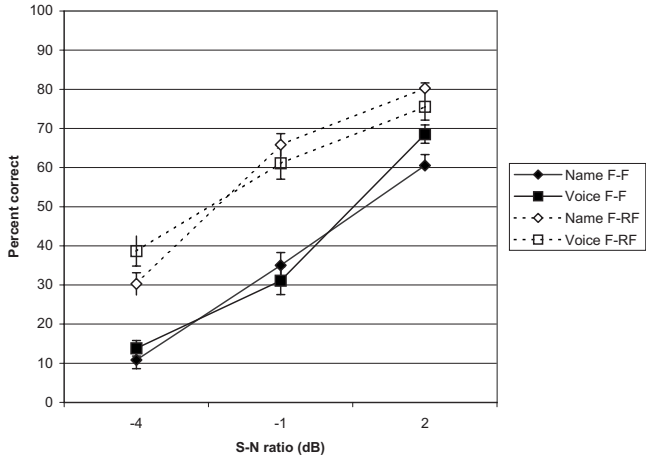


FIG. 2. Percent-correct recognition of key words in the TVM sentences. Filled symbols represent performance in the F-F (spatially coincident) condition, and open symbols display performance in the F-RF (spatially separated) condition. Performance for the name instruction mode is denoted by the diamond-shaped symbols; identification accuracy for the voice instruction mode is shown with the square symbols. Error bars represent one standard error.

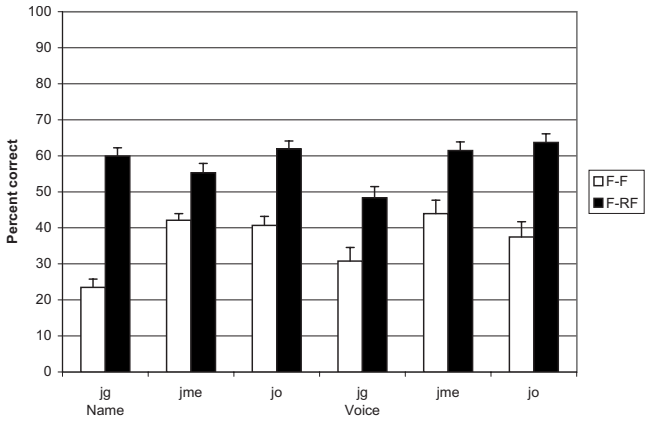


FIG. 3. Percent-correct recognition of key words in the TVM sentences with data aggregated by target talker and instruction mode. Open bars represent performance in the F-F (spatially coincident) condition, and filled bars display performance in the F-RF (spatially separated) condition. Data are averaged across S-N ratios. Error bars represent one standard error.

ated (TDT PA4), amplified (TDT HBUF5), and then power amplified (TOA P75D) before being sent to the loudspeaker(s) (Realistic Minimus 7). The target sentence was presented at a level of 56 dBA (re speech peaks).

Twelve conditions were run in this experiment: all possible combinations of two spatial configurations (F-F and F-RF), three signal-to-noise (S-N) ratios (−4, −1, and +2 dB), and two instruction modes. In the present study, the nominal S-N ratio was based on the level of the combination of the two maskers (i.e., a 0 dB S-N ratio was produced by presenting the target and the mixture of the two maskers at an equal level). No adjustment in the expressed S-N ratio was made for the F-RF condition even though the masker was presented from an additional loudspeaker (and, as a result, was 3 dB higher than in the F-F condition). In the *name cue* instruction mode, subjects were told to repeat the sentence beginning with a specified cue name (e.g., Theo, Victor, or Michael). This cue was presented to the subject (via text) on a screen of a laptop computer immediately before sentence presentation for a period of approximately 3 s. For the *voice cue* instruction mode, subjects heard a preview of the target talker's voice (saying, "this is the target sentence") prior to being presented with the three simultaneous sentences. They were instructed to repeat the sentence spoken by the target voice.

All variables were randomized on a trial-by-trial basis. Nine young normal-hearing subjects (age: 21–33 years; mean: 23.50 years) each participated in one session lasting approximately 2 h. Subjects heard 30 trials per condition. Data described below are based on 540 responses per condition (30 trials × 2 scoring words per trial × 9 listeners).

B. Results

Accuracy of identification of scoring words in the two instruction modes is displayed in Fig. 2. Several patterns can be observed in the data. First, there was little difference in performance between the two instruction conditions. It appears that listeners were, in general, equally adept at using semantic and lexical cues to resolve this listening task. Second, when three sentences were presented together, the TVM stimuli produced a substantial amount of informational masking. We assume that the F-RF condition, in which the masker is clearly heard in a location different from that of the target, produces little or no informational masking. Hence, our premise is that better performance in the F-RF condition versus the F-F condition indicates the presence of informational masking as differences in the amount of energetic masking (which is slightly greater in the F-RF condition since the masker comes from two loudspeakers rather than one) cannot explain this result (e.g., Freyman *et al.*, 1999). Repeated-measure analysis of variance (ANOVA) on these data [transformed into rationalized arcsine units (rau) (Studebaker, 1985)] confirmed these trends. Significant main effects were found for both S-N ratio [$F(2,7)=251.50$, $p<0.001$] and spatial condition [$F(1,8)=115.90$, $p<0.001$], but the main effect of instruction mode was not significant. The interaction of S-N ratio x spatial condition was significant [$F(2,7)=12.00$, $p=0.005$], as were the instruction mode x S-N ratio interaction [$F(2,7)=6.82$, $p=0.023$] and the three-way interaction [$F(2,7)=4.96$, $p=0.046$]. Hence, there was a small effect of instruction mode that depended on S-N ratio.

Examination of the data suggests that the voice cue was slightly more effective than the name cue at −4 dB S-N ratio, more so for the F-RF condition than for the F-F condition. The most parsimonious explanation for this finding is that at −4 dB S-N ratio, listeners might fail to correctly perceive the cue name (which was only briefly presented at the beginning of the target sentence), while they had a longer opportunity to capture the target talker's voice, which was available throughout the trial. For the F-RF spatial condition, the name cue was slightly more effective than the voice cue for the more advantageous S-N ratios. The reverse was true at +2 dB S-N ratio for the F-F condition.

Data were also analyzed to examine differences in performance among the talkers. Results of this analysis can be seen in Fig. 3. ANOVA on these data (averaged across S-N

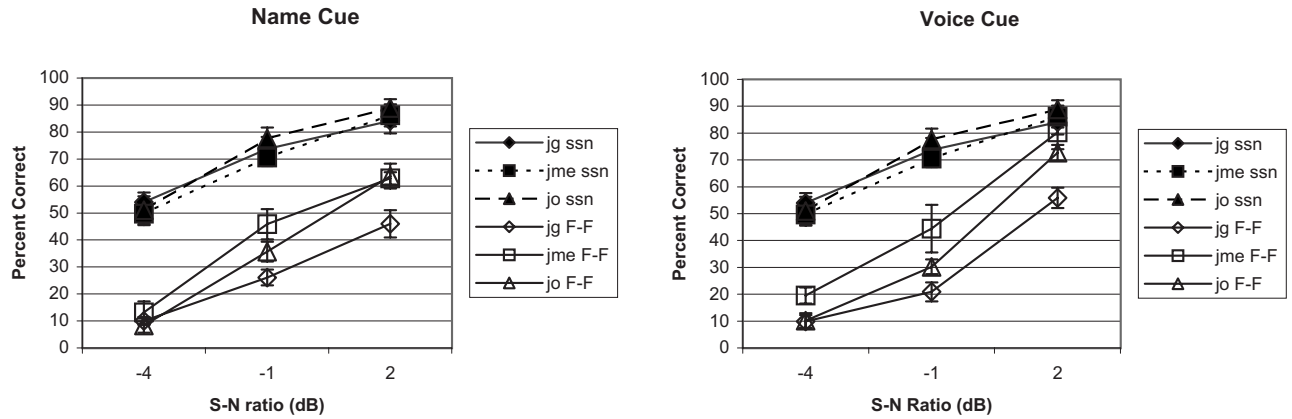


FIG. 4. Comparison of percent-correct identification of key words in the TVM sentences presented in SSN and in the presence of speech maskers. Filled symbols are data from experiment 2 (SSN), and open symbols are from experiment 1 (competing speech maskers) for the name instruction mode (left panel) and the voice instruction mode (right panel). Data are aggregated by target talker (jg, jme, and jo). Error bars represent one standard error.

ratios after being converted to rau) revealed significant main effects for talker $[F(2,7)=18.33,\ p=0.002]$ and for spatial condition $[F(1,8)=138.23,\ p<0.001]$ as well as a significant interaction between these two factors $[F(2,7)=6.74,\ p=0.023]$. There were no significant main or interaction effects involving instruction mode except for the three-way interaction, which just reached significance $[F(2,7)=4.81,\ p=0.049]$. When the utterances of these three talkers were presented simultaneously, talker JG was more difficult to understand than the other two talkers, and talker differences were greater in the spatially colocated F-F condition than when presented with spatial cues. This result suggests that voices that are equally intelligible when presented with what is assumed to be a purely energetic masker (i.e., the F-RF condition) may be differentially affected by informational masking (which is presumed to contribute in the F-F condition). *Post hoc* t-tests indicated that intelligibility of all three talkers was significantly different from one another during F-F presentation. For the F-RF configuration, talker JO was significantly easier to understand than the other two talkers; intelligibility differences between JME and JG were not significant in this spatial configuration.

IV. EXPERIMENT 2: TVM SENTENCE INTELLIGIBILITY IN SPEECH-SHAPED NOISE

The purpose of this experiment was to determine whether the talker differences noted in experiment 1 (with speech maskers) were also obtained when noise was used as competition. Because talker differences were greater in the first experiment in the spatially coincident F-F condition than when stimuli were spatially separated, we had reason to believe that informational masking played a key role in this pattern of results; in other words, stimuli from the different talkers varied in intelligibility in experiment 1 because their voices were more or less confusable. We were interested in determining the extent to which talker differences were seen when the masker was purely energetic.

A. Procedure

Each sentence was played in the presence of speech-shaped noise (SSN) that was derived from the TVM sentences. To produce the SSN, 15 of the scaled sentences from each talker were concatenated. A custom software program was used to extract the spectral envelope from this waveform and to shape a white noise with this derived envelope. Ten young college-aged listeners (age: 20–31 years; mean: 23.2 years) with normal hearing (verified via pure-tone screening) participated in this experiment. None of these subjects participated in experiment 1.

Testing was completed in an IAC sound-treated chamber. Across all subjects, each sentence from each of the three talkers was presented at three S-N ratios (−4, −1, and +2 dB). The experimental setup was identical to that used in experiment 1, with the exception that the speech and SSN stimuli were always delivered from the front loudspeaker. The speech stimuli were played at a level of 56 dBA (re speech peaks), and the noise level was adjusted to produce the desired S-N ratios.

Each subject listened to 324 sentences (36 sentences for each of the nine talker/S-N ratio combinations). The target talker and S-N ratio were varied on a trial-by-trial basis. Subjects were instructed to verbally repeat the target sentence, and an experimenter, seated in a control room, scored the sentences online. Testing took place in one session lasting approximately 1.25 h.

B. Results

Percent-correct performance for experiment 2 is displayed in Fig. 4; for comparison purposes, scores from experiment 1 are also plotted in this figure. The left panel of Fig. 4 shows performance in the presence of SSN compared to data from experiment 1 in the name cue conditions, while the right panel contrasts performance in SSN to that using the voice cue. Data collected in experiment 2 (converted to rau) were analyzed via repeated-measure ANOVA with talker and S-N ratio as the independent variables. The main effect of S-N ratio $[F(2,8)=33.36,\ p<0.001]$ was significant, but neither the main effect of talker nor the S-N ratio by talker interaction reached statistical significance. Hence, differences in intelligibility among the three talkers' recordings were very small (and nonsignificant) when the sentences were presented in steady-state noise. Sentence perception in

the steady-state SSN (filled symbols) was better than performance in the presence of a two-talker masker (open symbols). This result is consistent with the speech masker producing informational masking.

Results of these two experiments demonstrate that talker differences that are not apparent in the presence of steady-state noise can be seen when three sentences are presented simultaneously, especially when target and masker are spatially coincident. Moreover, at least for the present recordings, listeners can use either indexical or semantic cues in speech-on-speech masking, although the extent to which these cues are effective appears to depend on the S-N ratio as well as on the particular talker. Our next experiment was designed to further examine subjects' use of indexical information in speech-on-speech masking.

V. EXPERIMENT 3: THE EFFECT OF A VOICE CUE ON TVM SENTENCE PERCEPTION

This experiment was an exploration of the effect of listeners' exposure to voice information. Experiment 1 showed that knowledge of the talker's voice helped the listeners locate and attend to the target talker. However, it is unclear whether the processing of voice cues in competing speech is automatic or mandatory. Data from previous research suggest that listeners have difficulty ignoring indexical information during speech recognition. For example, Mullennix and Howe (1999) presented a same- or different-voice prime either before or after a target word (presented in quiet at a low intensity level). They found that a different-voice prime degraded performance, but only when the listener was required to attend to the prime. Our experience with presenting a semantic prime prior to stimulus presentation (Freyman *et al.*, 2004) suggests that it causes the target utterance to "pop out" from within a mixture of other voices, but that study did not include a prime that provided only indexical or voice information. The present experiment was designed to determine the extent to which presentation of a voice cue prior to a trial automatically draws the listener's attention to that voice. Specifically, we were interested in determining whether giving a voice cue or prime from a *masking* talker draws the subject's attention to that voice and makes it difficult to ignore.

A. Procedures

The same equipment, spatial conditions, and S-N ratios used in experiment 1 were also used in this experiment. On each trial the listener was given the name cue (Theo, Victor, or Michael) corresponding to that in the target sentence. As in experiment 1, this cue was displayed on the screen of a computer immediately prior to stimulus presentation. On 1/3 of the trials (hereafter called name trials), subjects were given only this name cue. On 1/3 of the trials, listeners were also given the voice cue corresponding to the target talker's voice; that is, they heard a preview of the target talker saying "this is the target sentence" immediately prior to stimulus presentation (this will be referred to as congruent trials). On the remaining trials, subjects were given the name cue but also heard a voice cue from one of the two masking talkers

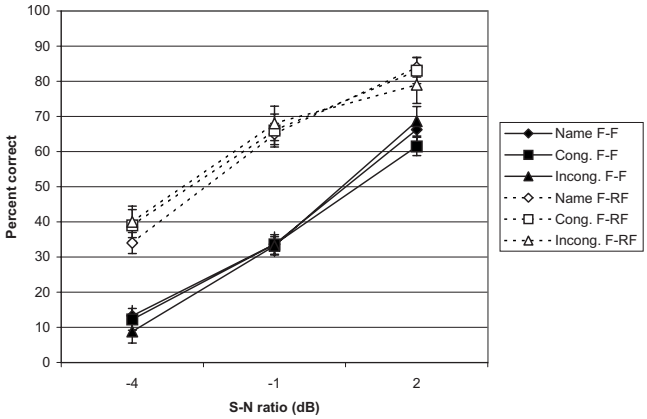


FIG. 5. Percent-correct identification of key words in the TVM sentences in the name, congruent (cong.), and incongruent (incong.) conditions. Filled symbols represent performance in the F-F (spatially coincident) condition, and open symbols display performance in the F-RF (spatially separated) condition. Error bars represent one standard error.

(these will be called incongruent trials). Subjects were told that on any given trial, the voice preview may or may not be the target voice. Each of ten young college-aged subjects (age: 19–25 years; mean: 20.88 years) listened to an average of 30 trials in each of the 12 conditions (all combinations of three S-N ratios, two spatial conditions, and three trial types). None of these individuals had participated in experiment 1 or experiment 2. The precise number of trials per condition varied slightly from subject to subject, and conditions were randomized on a trial-by-trial basis.

B. Results

One subject was not able to complete this experiment, and data from another showed clear outliers in the F-F condition. Hence, the analyses discussed below are based on data from eight subjects. Performance in percent correct for experiment 3 is shown in Fig. 5. It is clear that the presentation of an incongruent voice cue does not adversely affect performance. Averaged across S-N ratios and spatial conditions, performance for the three types of conditions was almost identical, between 49% and 50% correct. Repeated-measure ANOVA (on the data transformed into rau) confirmed this trend: significant main effects of spatial condition $[F(1,7)=102.70,\ p<0.001]$ and S-N ratio $[F(2,6)=505.42,\ p<0.001]$ and their interaction $[F(2,6)=8.36,\ p=0.018]$ were found, with insignificant main effects and interactions involving condition type. It appears that the presence of a congruent voice cue in addition to a name cue does not provide benefit, and the presentation of an incongruent voice cue does not hinder performance. It should be kept in mind that subjects could choose to ignore the voice prime since they were told that it might or might not contain useful information. Therefore, results of this study could be interpreted as suggesting that voice information is not automatically encoded while listening to speech. Although subjects were not asked about their subjective impressions, it is likely that providing a voice cue from a masking talker did not cause that voice to pop out, at least not to the extent that it interfered with the perception of the target

sentence.

VI. TRENDS ACROSS EXPERIMENTS

A. Error patterns

We examined error patterns in the data from experiments 1 and 3 in terms of the types of responses listeners made when their perception was incorrect. Errors were classified into two categories. Nonconfusion errors were those in which subjects' responses were either errors of omission (e.g., the subject said "I don't know" or something similar) or (usually) close acoustic approximations to the target (e.g., "face" for "pace"). Confusion errors were those in which the incorrect response was a word from one of the maskers, which presumably reflects some confusion between the target and the maskers.

For each subject, the proportions of each error type (out of the total number of errors) were calculated for each condition, aggregated across S-N ratios and talkers. By far, the most frequent type of error in the data from both experiments 1 and 3 were nonconfusion errors. This type of error accounted for 78%–95% of all errors across listening conditions. It should be noted that the overwhelming prevalence of nonconfusion errors found here is quite different from that found with the CRM corpus, where most of the errors involve responses from a masker (Brungart, 2001; Brungart and Simpson, 2002; Brungart and Simpson, 2007; Kidd *et al.*, 2005a; Wightman and Kistler, 2005). The difference in error patterns most likely reflects the fact that the CRM is a closed-set measure, while listeners are not given alternatives when using the TVM sentences.

The proportion of nonconfusion errors was substantially higher in the spatially separated F-RF condition than in the F-F condition. In other words, subjects' confusions between the target and masker were more common in the spatially coincident F-F condition than when the target and masker were spatially separated, as shown in Fig. 6, consistent with results of studies using the CRM stimuli (e.g., Arbogast *et al.*, 2002; Arbogast *et al.*, 2005; Brungart and Simpson, 2002; Brungart and Simpson, 2007).

It should also be noted that error patterns varied substantially among listeners, which could reflect individual differences in response style and/or differences in perception. Particularly notable was the observation that some listeners made a relatively large number of confusion errors, while other subjects rarely responded in this way. For example, within the data from one target talker in one listening condition (F-F for name trials with talker JO in experiment 3), the percentage of confusion errors across listeners ranged from 6% to 38%. This could indicate that some listeners had a more difficult time than others in distinguishing the target from the masker. Conversely, it could also reflect differences in response bias, wherein some individuals were more likely than others to respond with a word that they heard within the mixture even if they were unsure of whether or not it came from the target, while other subjects chose not to respond when unsure.

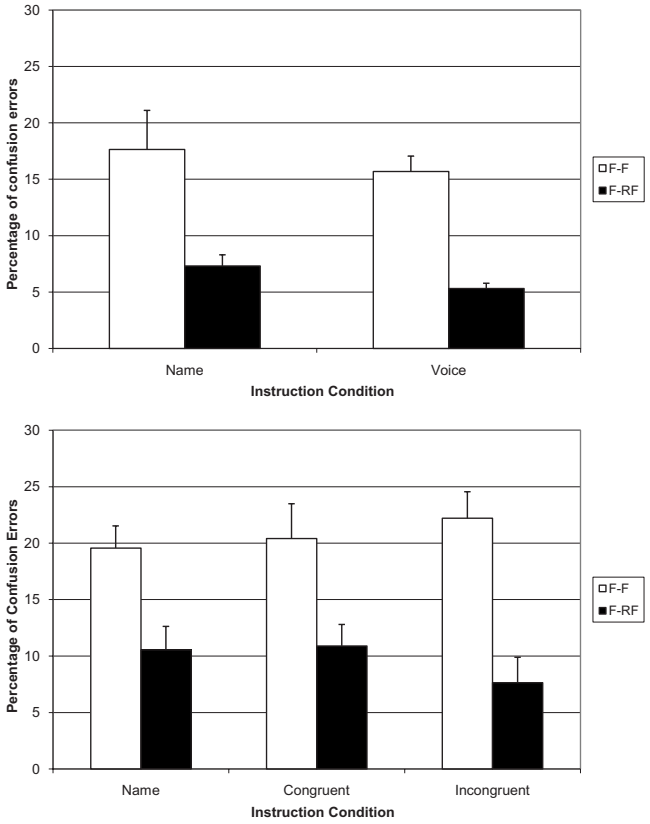


FIG. 6. Proportions of errors that involved responses from maskers in experiment 1 (top panel) and experiment 3 (bottom panel), aggregated by instruction condition in each study. Open bars represent performance in the F-F (spatially coincident) condition, and filled bars display performance in the F-RF (spatially separated) condition. Error bars represent the standard error.

B. Learning effects

Perceptual differences between target and masking voices may be used by listeners to resolve a competing speech situation. Listeners can learn to differentiate and identify voices with fairly high accuracy (e.g., Nygaard *et al.*, 1994), and the learning of voices can occur incidentally (that is, without conscious intention). We were interested in examining the extent to which incidental learning of the voices of the three talkers used in these experiments affected performance. We expected to see a small learning effect over the course of the experiment as subjects gained experience with the task and determined the effectiveness of certain strategies. Our premise was that listeners would become familiar with the three talkers' voices over the course of the experiment and that this exposure would contribute to a larger learning effect in certain conditions than in others. Specifically, we theorized that repeated exposure to voice information would provide greater benefit in conditions in which confusion between the target talker and masking talkers limited performance (that is, in the F-F condition) versus that in conditions in which informational masking is minimized (in the F-RF condition for speech maskers or when the masker was noise).

Performance on the first and last 25 trials (that is, the first and last 50 scored items, with two scored words per trial) presented to each listener for each condition (F-F and

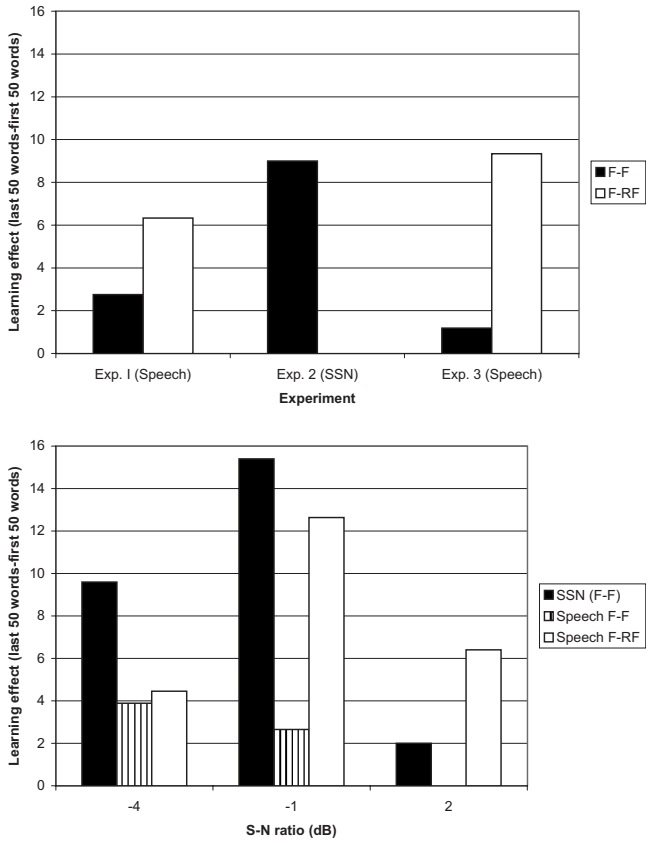


FIG. 7. The difference in recognition ability between the first 25 sentences (i.e., first 50 key words) and last 25 sentences (last 50 key words). The top panel displays data aggregated by experiment, averaged across S-N ratios. The bottom panel shows learning effects aggregated by condition (F-F/F-RF and S-N ratio) with data in the presence of speech maskers averaged across experiments 1 and 3.

F-RF for each of the S-N ratios) was compared for the three experiments. These data are shown in Fig. 7. Learning effects were greatest for conditions that were unlikely to cause informational masking (in the presence of the SSN masker and in the F-RF presentation mode with speech maskers) and smallest in the F-F condition with speech maskers.

These trends were confirmed with repeated-measure ANOVA (with the data transformed to rau). For experiment 2, the difference in recognition of the first versus last 50 target words was significant $[F(1,9)=35.06,\ p<0.001]$, as was the interaction between order and S-N ratio $[F(2,8)=9.72,\ p=0.007]$. Analysis of the data from experiment 1 also showed significant effects of order $[F(1,8)=12.46,\ p=0.008]$ and significant interactions between spatial condition and order $[F(1,8)=6.14,\ p=0.038]$ and between the three variables of spatial condition, order, and S-N ratio $[F(2,7)=12.48,\ p=0.005]$. ANOVA on data from experiment 3 revealed no significant main or interaction effects involving order.

Results of this analysis did not support the idea that learning of the talkers' voices would provide greater benefit in conditions where confusion between talkers is likely to play a role in performance. It is possible that learning effects could have been related simply to task difficulty rather than to spatial condition or masker. However, there was no systematic influence of S-N ratio on the order effect (see the bottom panel of Fig. 7), suggesting that a simple explanation of task difficulty cannot explain the data. This finding is somewhat contrary to a recent report of Van Engen and Bradlow (2007), who found learning effects in their task of speech-on-speech masking using native and non-native maskers. They did not analyze their data in terms of learning effects in maskers that did and did not produce informational masking, so a direct comparison between our results and theirs cannot be made.

VII. DISCUSSION

The studies described in this paper investigated the types of cues listeners use in a competing speech situation. Experiment 1 demonstrated that listeners can use both semantic information and indexical (or voice) information to identify and attend to a target message in a multitalker environment. The differences that were noted in listeners' ability to use these types of cues were small and depended on both the specific talker and the S-N ratio. There was some indication that the voice cue was slightly more effective than the name cue at the lowest S-N ratio and that the reverse was true at higher S-N ratios (although this trend was not entirely consistent).

The relative equivalence in performance when using the two types of cues (name versus voice) was perhaps an unexpected finding, as the ability to remember an unfamiliar voice and use that information to identify the target utterance is inherently a more difficult task than simply finding the sentence beginning with a specific name. Hence, it might be expected that performance using the voice cue would be poorer than that obtained using the name cue. It is possible that this result was not found because, at least at the most adverse S-N ratio, the cue name itself was masked to such an extent—during the brief period it was available—that listeners could not always determine the target utterance. The observation that the difference in instruction cues was larger in the F-RF than in the F-F condition supports this premise as energetic masking was greater in the F-RF trials (because the masker energy was 3 dB higher). In essence, a voice cue prior to sentence presentation may give the listener the same information as specifying a semantic cue at the beginning of a sentence, assuming that listeners are using the voice they hear reciting the cue name to find the stream corresponding to the target key words.

The recordings produced by the three talkers varied in intelligibility when presented in a multitalker situation but not when played in steady-state noise. Moreover, talker differences were larger when the voices were presented in the spatially coincident F-F condition than when spatially separated. Taken together, these two findings suggest that informational masking may contribute more than energetic masking to the differences in intelligibility found among the talkers' utterances.

It is possible that talker differences occurred primarily because the talkers' productions varied in intelligibility when presented together (that is, one of the voices "stood out" from the others or, conversely, was more confusable with one of the other voices). This result would be consistent with that

reported by Brungart (2001), whose data suggest that some voices are more resistant to (or produce more) informational masking than others. It is also feasible that talker differences were caused by the relative masking effectiveness of the specific two-talker masker combinations. This explanation is consistent with data collected previously in our laboratory, where we found that different pairs of two-talker maskers produced greater differences in informational masking than in energetic masking (Freyman *et al.*, 2007). Because the same two-talker masker was always used for a given target voice in the present study, we are unable to tease apart these factors.

Although it was clear that listeners could use voice information to help identify and attend to the target, such information did not appear to be automatically encoded. Results from experiment 3 demonstrated that subjects could readily ignore a voice prime presented just prior to stimulus presentation. The presentation of a prime that was consistent with the target voice did not help the listener when the name cue was also available. This result agrees with data from Brungart *et al.* (2001), who also found that knowing the target talker's voice provided little benefit in same-sex masking conditions when the cue name was presented. Moreover, the presence of an incongruent prime did not hinder the subjects' ability to find and understand the target message. It is likely that because subjects were told that the voice cue might or might not be useful on any given trial, they chose to ignore it. If voice information is encoded automatically, it could be argued that subjects would not be able to ignore a voice cue, even if they knew it was irrelevant. Future research needs to be done to determine whether encoding of indexical information is mandatory in a competing speech situation.

One unexpected finding in the present studies had to do with learning effects. Listeners' performance did improve slightly over the course of the experiment, but more so in conditions that produced only or predominantly energetic masking (in the presence of steady-state noise in experiment 2 and in the F-RF modes in experiments 1 and 3). Learning effects were minimal in the F-F condition with speech maskers, which produced both informational and energetic masking. This result is in opposition to what we had anticipated: that learning or experience would play a greater role in reducing informational masking than energetic masking.

Although studies of informational masking using tonal stimuli (Neff and Callaghan, 1988; Neff and Dethlefs, 1995; Oxenham *et al.*, 2003) or speech detection (Balakrishnan and Freyman, 2008; Freyman *et al.*, 2008) have noted little or no evidence of learning effects, to our knowledge only one study (Van Engen and Bradlow, 2007) has examined the extent to which this finding persists in a competing speech recognition task. Van Engen and Bradlow (2007) did indeed note that experience with the task and/or the target talker's voice improved subjects' performance over the course of their study. Intuitively, repeated exposure to the talkers' voices should make it easier to differentiate one voice from another. Assuming that informational masking is caused (in part) by uncertainty regarding the target utterance versus the masking signals, enhancing the ability to differentiate between voices within the mixture should lead to reduced informational masking. We currently can offer no explanation for the pattern of learning effects found in the present study.

In summary, results of the studies described in this paper suggest several potentially important factors related to competing speech perception. Listeners can use either talker voice (lexical) or cue name (semantic) information to resolve this listening task; voices that do not differ in susceptibility to energetic masking may still vary in intelligibility when presented under conditions with informational masking; and incidental learning of the talkers' voices across an experimental session may play a greater role in reducing energetic masking than informational masking.

Finally, these studies suggest that the TVM sentences may prove to be a viable corpus for use in competing speech research. These stimuli have the advantage of incorporating a cue name but are open-set in nature, allowing the listener to use higher-level linguistic skills to aid in understanding. Perhaps because of their open-set nature, the TVM sentences are more difficult than the CRM sentences at equivalent S-N ratios and so may be appropriate for future research where closed-set stimuli such as the CRM are likely to be too easy, such as during auditory-visual presentation. Future studies will examine the extent to which individual sentences can be repeated across (and between) experimental sessions without the risk of remembering specific stimuli, as well as measure statistical properties of the corpus.

ACKNOWLEDGMENTS

We thank Sugata Bhattacharjee, Beth Ann Jacques, J. Ackland Jones, Amanda Lepine, and April Teehan for their contributions to this project. Portions of these data were presented at the Acoustical Society of America Conference, Providence, RI, June 2006. This work was supported by NIDCD Grant No. 01625.

Arbogast, T. L., Mason, C. R., and Kidd, G. K., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T. L., Mason, C. R., and Kidd, G. K., Jr. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Balakrishnan, U., and Freyman, R. L. (**2008**). "Speech detection in spatial and non-spatial speech maskers," J. Acoust. Soc. Am. **123**, 2680–2691.

Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol. **13**, 108–112.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communication research," J. Acoust. Soc. Am. **107**, 1065–1066.

Broadbent, D. E. (**1952**). "Listening to one of two synchronous messages," J. Exp. Psychol. **44**, 51–55.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., Iyer, N., and Simpson, B. D. (**2006**). "Monaural speech segregation using synthetic speech signals," J. Acoust. Soc. Am. **119**, 2327–2333.

Brungart, D. S., and Simpson, B. D. (**2002**). "Within-ear and across-ear interference in a cocktail-party listening task," J. Acoust. Soc. Am. **112**, 2985–2995.

Brungart, D. S., and Simpson, B. D. (**2007**). "Effect of target-masker simi-

larity on across-ear interference in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **122**, 1724–1734.
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.
Clopper, C. G., Pisoni, D. B., and Tierney, A. T. (**2006**). "Effects of open-set and closed-set task demands on spoken word recognition," J. Am. Acad. Audiol. **17**, 331–349.
Darwin, C., Brungart, D., 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.
Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2008**). "Spatial release from masking with noise-vocoded speech," J. Acoust. Soc. Am. **124**, 1627–1637.
Freyman, R. L., Helfer, K. S., and Balakrishnan, U. (**2007**). "Variability and uncertainty in masking by competing speech," J. Acoust. Soc. Am. **121**, 1040–1046.
Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation on the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.
Helfer, K. S., and Freyman, R. L. (**2005**). "The role of visual speech cues in reducing energetic and informational masking," J. Acoust. Soc. Am. **117**, 842–849.
Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005a**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.
Kidd, G., Jr., Mason, C. R., Brughera, A., and Hartmann, W. M. (**2005b**). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," Acta. Acust. Acust. **91**, 526–536.
Mullennix, J. W., and Howe, J. N. (**1999**). "Selective attention in perceptual adjustments to voice," Percept. Mot. Skills **89**, 447–457.
Mullennix, J. W., and Pisoni, D. B. (**1990**). "Stimulus variability and processing dependencies in speech perception," Percept. Psychophys. **47**, 379–390.
Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (**1989**). "Some effects of talker variability on spoken word recognition," J. Acoust. Soc. Am. **85**, 365–378.
Neff, D. L., and Callaghan, B. P. (**1988**). "Effective properties of multicomponent simultaneous maskers under conditions of uncertainty," J. Acoust. Soc. Am. **83**, 1833–1838.
Neff, D. L., and Dethlefs, T. M. (**1995**). "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," J. Acoust. Soc. Am. **98**, 125–134.
Nerbonne, G. P., Ivey, E. S., and Tolhurst, G. C. (**1983**). "Hearing protector evaluation in an audiometric testing room," Sound Vib. **17**, 20–22.
Nilsson, M., Soli, S. D., and Sullivan, J. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.
Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1994**). "Speech perception as a talker-contingent process," Psychol. Sci. **5**, 42–46.
Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1995**). "Effects of stimulus variability on perception and representation of spoken words in memory," Percept. Psychophys. **57**, 989–1001.
Oxenham, A. J., Fligor, B. J., Mason, C. R., and Kidd, G. (**2003**). "Informational masking and musical training," J. Acoust. Soc. Am. **114**, 1543–1549.
Rakerd, B., Aaronson, N. L., and Hartmann, W. M. (**2006**). "Release from speech-on-speech masking by adding a delayed masker at a different location," J. Acoust. Soc. Am. **119**, 1597–1605.
Rothauser, E. H., Chapman, W. D., Guttman, N., Norby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "I.E.E.E. recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.
Sommers, M. S., Kirk, K. I., and Pisoni, D. B. (**1997**). "Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners I: The effects of response format," Ear Hear. **18**, 89–99.
Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.
Thorndike, K. I., and Lorge, I. (**1952**). *The Teacher's Word Book of 30,000 Words* (Columbia University Press, New York).
Van Engen, K. J., and Bradlow, A. R. (**2007**). "Sentence recognition in native- and foreign-language multi-talker background noise," J. Acoust. Soc. Am. **121**, 519–526.
Wightman, F. L., and Kistler, D. J. (**2005**). "Informational masking of speech in children: Effects of ipsilateral and contralateral distracters," J. Acoust. Soc. Am. **118**, 3164–3176.

Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference

Andrew J. Oxenham[a] and Andrea M. Simonson
Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

(Received 18 October 2007; revised 24 September 2008; accepted 13 October 2008)

Speech intelligibility was measured for sentences presented in spectrally matched steady noise, single-talker interference, or speech-modulated noise. The stimuli were unfiltered or were low-pass (LP) (1200 Hz cutoff) or high-pass (HP) (1500 Hz cutoff) filtered. The cutoff frequencies were selected to produce equal performance in both LP and HP conditions in steady noise and to limit access to the temporal fine structure of resolved harmonics in the HP conditions. Masking release, or the improvement in performance between the steady noise and single-talker interference, was substantial with no filtering. Under LP and HP filtering, masking release was roughly equal but was much less than in unfiltered conditions. When the average *F*0 of the interferer was shifted lower than that of the target, similar increases in masking release were observed under LP and HP filtering. Similar LP and HP results were also obtained for the speech-modulated-noise masker. The findings are not consistent with the idea that pitch conveyed by the temporal fine structure of low-order harmonics plays a crucial role in masking release. Instead, any reduction in speech redundancy, or manipulation that increases the target-to-masker ratio necessary for intelligibility to beyond around 0 dB, may result in reduced masking release. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3021299]

PACS number(s): 43.71.Gv, 43.66.Hg [RLF] Pages: 457–468

I. INTRODUCTION

A common complaint of hearing-impaired listeners and cochlear-implant users relates to their difficulty in understanding speech in complex acoustic backgrounds. One illustration of this problem comes from studies that compare speech intelligibility in steady noise with intelligibility in spectrally and/or temporally fluctuating maskers. Speech intelligibility, as measured in normal-hearing listeners, improves dramatically when temporal fluctuations are introduced into a noise masker, or when the noise masker is replaced by a single-talker interferer with the same long-term power spectrum (e.g., Miller and Licklider, 1950; Festen and Plomp, 1990). This "masking release" has been ascribed to normal-hearing listeners' ability to "listen in the valleys," making use of favorable target-to-masker ratios (TMRs) in brief (or narrowband) low-energy temporal (or spectral) epochs in the masker. In contrast, hearing-impaired listeners typically show much less release from masking when the masker is changed from steady noise to a fluctuating noise or single-talker interferer (e.g., Festen and Plomp, 1990; Peters *et al.*, 1998), and cochlear-implant users have been found to show either no masking release, or even an increase in masking, under similar conditions (Nelson *et al.*, 2003; Stickney *et al.*, 2004). For hearing-impaired listeners, the lack of masking release has traditionally been explained in terms of a putative loss of temporal and spectral resolution, limiting their ability to make use of brief or narrowband dips in masker energy. Similarly, for cochlear-implant users, at least some of the deficit may be due to severely reduced spectral resolution.

More recently, a number of studies have suggested that deficits in fundamental frequency (*F*0) coding and pitch perception may also be a possible cause for the lack of masking release in listeners with hearing impairment (Summers and Leek, 1998) or cochlear implants (Nelson and Jin, 2004; Stickney *et al.*, 2004). In contrast to explanations based on a loss of spectral or temporal resolution, a reduction in masking release due to reduced pitch salience does not necessarily imply that the target was masked by the interferer in the traditional sense. Instead, it suggests that the cues necessary to distinguish the target from the interferer were not sufficiently salient for successful perceptual segregation to occur. The failure to distinguish and segregate a target from an interferer could result in the masker fluctuations being confused for those of the target, thus actually producing more (rather than less) masking. This phenomenon has been referred to as a form of "informational masking," rather than the more traditional "energetic masking" that is believed to dominate for a steady-noise masker (e.g., Brungart, 2001). In cases where the interferer is speech, temporal fine-structure cues may help to distinguish between the different *F*0s, or *F*0 trajectories, of the target and the interferer (Brokx and Nooteboom, 1982; Bird and Darwin, 1998; Assmann, 1999; Stickney *et al.*, 2007). In cases where the interferer is a modulated noise, the fact that parts of the speech are voiced may be sufficient to produce some segregation between the speech and the noise (Qin and Oxenham, 2003; Lorenzi *et al.*, 2006).

In normal hearing, pitch information appears to be conveyed primarily by the temporal fine structure of the lower-

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

order harmonics, which are thought to be resolved in the peripheral auditory system. The ability of listeners to discriminate small differences in $F0$ is considerably better for low-order harmonics than high-order (unresolved) harmonics (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003), and when $F0$ information from low-order harmonics conflicts with that from high-order harmonics, the low-order harmonics usually dominate the percept, both for isolated harmonic complex tones (Plomp, 1967; Micheyl and Oxenham, 2007) and for speech (Bird and Darwin, 1998). In fact, studies using harmonic tone complexes consisting only of high-order unresolved harmonics suggest that listeners are not capable of extracting the pitches of two complexes presented simultaneously in the same spectral region (Carlyon, 1996; Micheyl *et al.*, 2006).

Hearing-impaired listeners typically have less access to resolved harmonics, due in part to abnormally broad auditory filters and, in many cases, poorer coding of single pure tones (e.g., Bernstein and Oxenham, 2006). It has also been claimed that hearing-impaired listeners may have a reduced ability to process temporal fine-structure information (Lorenzi *et al.*, 2006), although it is difficult to separate this from the effects of poorer frequency selectivity. Whatever the underlying mechanisms, Lorenzi *et al.* (2006) showed that poorer temporal fine-structure processing seems to correlate negatively with masking release when comparing speech reception in steady noise with that in fluctuating noise.

The coding schemes used in cochlear implants do not typically convey temporal fine-structure information in the waveforms of each electrode, and the spectral representation is too coarse to provide temporal fine structure in terms of place of excitation of individual harmonics. Thus, if the accuracy of pitch coding plays a major role in masking release, one reason why both hearing-impaired listeners and cochlear-implant users show severely reduced amounts of masking release may be the reduced pitch salience experienced by them, which is induced by reduced or no access to the temporal fine structure associated with low-order resolved harmonics.

Given the proposed link between pitch salience and temporal fine structure on the one hand, and masking release on the other, an obvious prediction is that normal-hearing listeners should show reduced masking release if pitch salience and temporal fine-structure cues are disrupted in some way. Recent studies using acoustic simulations of cochlear-implant processing in normal-hearing listeners seem to support this prediction (e.g., Nelson *et al.*, 2003; Qin and Oxenham, 2003, 2006; Gnansia *et al.*, 2008; Hopkins *et al.*, 2008). For instance, Qin and Oxenham (2003) measured speech intelligibility in various types of interferers with the stimuli processed through noise-excited envelope vocoders that reduce spectral resolution (depending on the number of frequency channels) and replace the original temporal fine structure with a noise carrier. They found no release from masking (and even a slight increase in masking) when a steady noise was replaced by a single talker (male or female) interferer, even with the relatively good spectral resolution afforded by 24-channel simulations. Less masking release was also observed for a noise masker that was modulated with the temporal envelope of speech. Nevertheless, vocoder simulations distort the signal in other ways than simply reducing access to the original temporal fine structure; they also reduce spectral resolution somewhat (even in the 24-channel case) and, for noise-excited vocoders, introduce spurious envelope fluctuations that may also affect performance (Whitmal *et al.*, 2007). More recent studies (Gnansia *et al.*, 2008; Hopkins *et al.*, 2008) have used even more channels (32 in the case of Gnansia *et al.*, 2008) and tonal carriers, and have still found reduced masking release. However, even here, it is difficult to rule out the role of potential temporal envelope distortions due, for instance, to beating between neighboring carriers.

The present study provides a more direct test of the hypothesis that the pitch information provided by the temporal fine structure of low-order resolved harmonics is important for masking release. Sentences were presented in steady or fluctuating backgrounds in unfiltered, low-pass-(LP), or high-pass-(HP) filtered conditions. The rationale was that HP filtering the stimuli should eliminate the low-order harmonics necessary for accurate pitch perception, while LP filtering the stimuli should retain these low-order harmonics. Thus, if low-order resolved harmonics are necessary for masking release, then masking release should be observed in the LP-filtered condition, while little or no masking release should be observed in the HP-filtered condition. Experiment 1 tests this hypothesis by comparing speech intelligibility in steady noise with that in the presence of a single-talker interferer in broadband (BB), LP-filtered, and HP-filtered conditions. Experiment 2 increases the pitch differences between the target and interfering speech by artificially shifting the pitch contour of the interferer to lower $F0$s. Experiment 3 tests whether temporal fine structure is important for situations in which the masker is a temporally fluctuating noise, as suggested by some previous studies (Qin and Oxenham, 2003; Lorenzi *et al.*, 2006).

II. EXPERIMENT 1: INTELLIGIBILITY IN STEADY NOISE AND SINGLE-TALKER MASKERS

A. Methods

1. Stimuli

The target sentences were Hearing in Noise Test (HINT) sentence lists (Nilsson *et al.*, 1994), spoken by a male talker. The average $F0$ of these HINT sentences is approximately 110 Hz, with an average within-sentence standard deviation of 24 Hz. Speech-shaped noise maskers consisted of BB Gaussian noise that was filtered to match the long-term power spectrum of the HINT sentences. Competing-talker maskers were created using recordings of a subset of IEEE sentences (IEEE, 1969), also produced by a male. These sentences were also spectrally shaped to match the long-term power spectrum of the HINT sentences. The sentences were concatenated and then broken into 4-s blocks to be used as maskers, without regard to the beginnings of sentences. In other words, a 4-s block could start in the middle of a sentence. The gaps that occurred between concatenated sentences occurred anywhere within each 4-s block, and had an

average duration of 135 ms (with gaps defined as contiguous periods during which the level remained at least 35 dB below the long-term rms level of the speech). The average *F*0 of the competing-talker masker was approximately 106 Hz, with an average within-sentence standard deviation of 28 Hz.

The target sentences and maskers were added together at the appropriate TMR for a given condition before further filtering. The filtered stimuli were created by either LP or HP filtering the combined target and masker using a fourth-order Butterworth filter. The LP condition involved stimuli that were LP filtered with a cutoff frequency of 1200 Hz; the HP condition used a cutoff frequency of 1500 Hz. These particular cutoff frequencies were chosen in an attempt to equate overall performance in the speech-shaped steady-noise conditions, based on the results from a pilot study. A complementary off-frequency noise was subsequently added to the HP- and LP-filtered stimuli. The noise had the same long-term spectrum (before filtering) as the HINT target speech, and its unfiltered level was set to be 12 dB below the level of the unfiltered target speech. In the LP condition, the complementary noise was HP filtered at 1200 Hz; in the HP condition, the complementary noise was LP filtered at 1500 Hz (also using fourth-order Butterworth filters). The purpose of the complementary noise was to limit the possibility for "off-frequency listening" to information in the slopes of the filters, and to mask any low-frequency distortion products that might be generated by the speech in the HP-filtered conditions.

In all trials, the speech or noise masker and the complementary noise (when present) had a total duration of 4 s, gated on and off with 5-ms ramps. The target sentences ranged in duration from 1.17 to 2.46 s, and began 750 ms after the beginning of the masker.

2. Subjects

Twenty-four subjects (18 female) participated in this experiment. None had previously listened to HINT sentences. The subjects all reported normal hearing, and had audiometric thresholds of 20 dB HL or less at octave frequencies between 250 and 8000 Hz. Their ages ranged from 19 to 55, with a mean age of 25.3. Subjects were paid for their participation.

3. Procedure

Three groups of eight subjects each were tested using one of the three filter conditions (unfiltered, HP filtered, and LP filtered). Subjects were told that they would hear distorted sentences in a noisy background and they were instructed to type what they heard via a computer keyboard. It was explained that some of the utterances would be hard to understand, and that they should make their best guess at as many words as possible.

Approximately 1 h of practice was provided before testing. Practice target sentences were taken from 12 lists of sentences from the IEEE/TIMIT corpus (IEEE, 1969). These sentences were produced by a different talker from that used to create the maskers. Subjects heard half the sentences in a background of speech-shaped noise, and the other half in a competing-talker background, at several TMRs in BB, LP, and HP conditions. After each practice sentence, listeners were prompted to type what they had heard via a computer keyboard. Feedback was then provided by displaying the sentence text on the computer monitor. Listeners had the option to repeat the sentence as often as they wished before moving on to the next sentence.

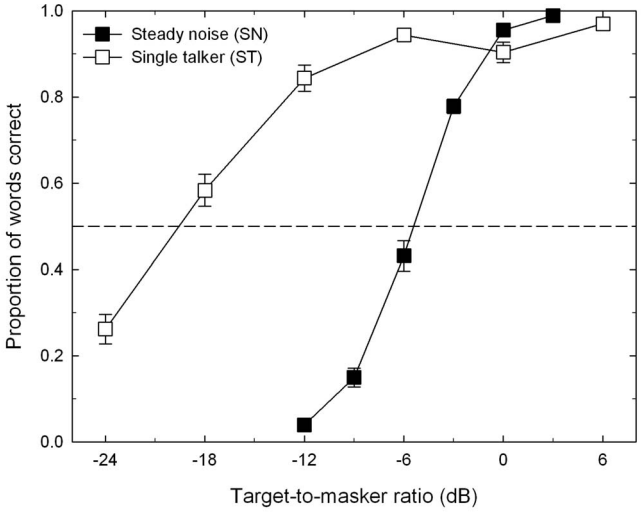


FIG. 1. Proportion of words correctly reported from HINT sentence lists as a function of TMR. Open symbols represent performance in the presence of a single-talker interferer and filled symbols represent performance in the presence of a steady-noise interferer. The square symbols in this and all other figures represent unfiltered conditions. Error bars represent ±1 standard error of the mean across subjects.

In the actual experiment, the stimuli were presented at six different TMRs for each masker type (speech-shaped noise, competing talker), resulting in 12 conditions. Two lists, each comprising ten HINT target sentences, were presented for each condition, resulting in a total of 24 lists of HINT sentences in the experiment. No target sentence was presented more than once to any subject. Each condition was presented once before any was repeated. The order of presentation was counterbalanced across subjects for the masker type, and always progressed from highest (easiest) to lowest (hardest) TMR. No feedback was given during the actual experiment, which took approximately 1 h and immediately followed the practice session.

For both the practice and the actual experiment, the stimuli were presented diotically at an overall level of 70 dB SPL (after processing) in each ear over Sennheiser HD580 headphones. The stimuli were played out via a Lynx22 (LynxStudio) sound card at 16 bit resolution with a sampling rate of 22.05 kHz. After the session, the listeners' responses were scored offline. All words except for "a" and "the" were counted, and the percentage of correct words in each condition was calculated. Obvious misspellings were counted as correct.

B. Results and discussion

The mean results from the unfiltered conditions are shown in Fig. 1, with proportion of words correctly reported plotted as a function of TMR. Error bars represent ±1 standard error across the eight subjects. The raw percent-correct

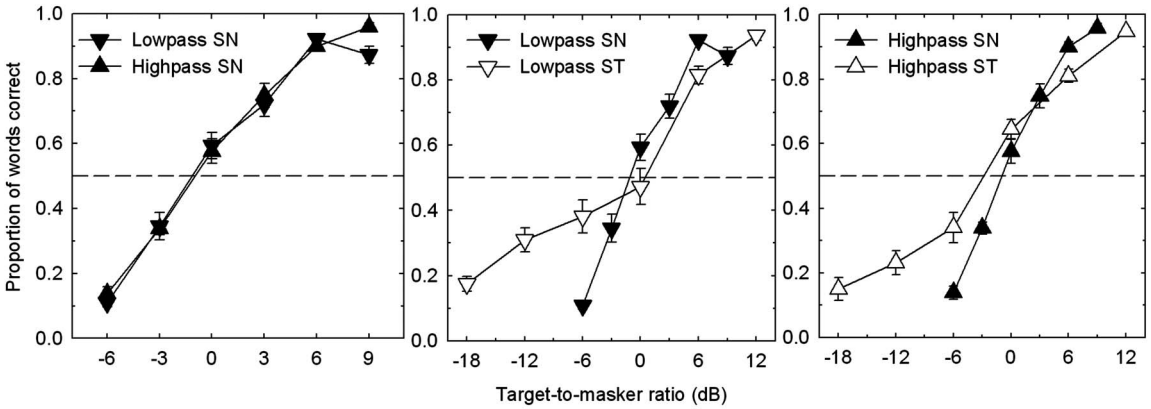


FIG. 2. Proportion of words correctly reported from HINT sentence lists as a function of target-to-masker ratio. Upward- and downward-pointing triangles represent HP and LP conditions, respectively. The left panel shows results in steady noise (filled symbols); the center and right panels compare those results with performance in a single-talker interferer (open symbols) in LP and HP conditions, respectively. Error bars represent ±1 standard error of the mean across subjects.

scores are shown; however, for the purposes of statistical analyses, the data were subjected to an arcsine transform. Filled symbols denote performance in steady-noise and open symbols denote performance in the single-talker background. As has been shown in many previous studies (e.g., Duquesnoy, 1983; Festen and Plomp, 1990; Peters *et al.*, 1998; Summers and Molis, 2004), a single-talker interferer produces considerably less masking than a steady noise with the same overall level and long-term spectrum. The amount of masking release can be quantified in various ways. One way is to consider the increase in proportion of words correctly reported at a given TMR (Takahashi and Bacon, 1992; Jin and Nelson, 2006). Quantified in this way, masking release was about 80 percentage points at a TMR of −12 dB, 50 percentage points at a TMR of −6 dB, and no masking release (or a "negative release" of 5 percentage points) at 0 dB TMR, perhaps because of ceiling effects. Another way to quantify masking release is to determine the change in TMR necessary to obtain 50% of words correct; this TMR is known as the speech reception threshold (SRT) (e.g., Qin and Oxenham, 2003).[1] The SRTs were estimated from our data by fitting a logistic function to the individual data from all TMRs, and determining the TMR at which the fitted functions crossed the 50% line. The logistic functions were fitted using maximum likelihood with an assumed binomial distribution. The SRTs, averaged across the SRTs from logistic fits to the data from individual subjects, were −5.6 and −19.5 dB for the steady-state noise and single-talker interferer, respectively, resulting in an average masking release of around 14 dB.

The left panel of Fig. 2 shows the results from the LP and HP-filtered conditions in a steady-noise background. The similarity of the two curves indicates that we were successful in selecting LP and HP filter cutoffs that produced roughly equal performance in steady noise for the two conditions. A mixed-model analysis of variance (ANOVA) on the arcsine transformed data confirmed a significant effect of TMR ($F_{5,70}$=241.0, $p<0.0001$) but no effect of filter type ($F_{1,14}$ =1.39, p=0.26).[2] The interaction between filter type and TMR was significant ($F_{5,70}$=3.84, p=0.01), presumably due to the downturn in performance at the highest TMR in the LP condition. The middle panel of Fig. 2 replots the LP results from the left panel and compares them to the LP results using a competing-talker interferer. Some masking release is present, at least at negative TMRs, but it is considerably less than that found in the unfiltered conditions. The right panel shows the same comparison of steady noise and single-talker interferer, but for the HP-filtered conditions. Again, the amount of masking release is greatly reduced, relative to the unfiltered conditions, but is similar to that found in the LP filtered conditions. The similarity of the HP and LP-filtered conditions was confirmed by a mixed-model ANOVA comparing the HP- and LP-filtered single-talker conditions, with TMR as a within-subject variable and filter condition as a between-subject variable. As expected, there was a significant effect of TMR ($F_{5,70}$=263.8, $p<0.0001$) but no main effect of filter type ($F_{1,14}<1$, n.s.).

Overall, the amount of masking release was small in both filtered conditions. In the LP condition, a within-subject ANOVA, using only TMRs that were measured in both the single-talker and steady-noise maskers, revealed no main effect of masker type ($F_{1,7}<1$, n.s.), but a significant interaction between masker type and TMR ($F_{2,14}$=39.14, $p<0.0001$). In the HP condition, the main effect of masker type failed to reach significance ($F_{1,7}$=3.97, p=0.086), but the interaction between TMR and masker type was significant ($F_{2,14}$=13.51, p=0.0006). In both HP and LP conditions, it appears that masking release is present only at negative TMRs, and is effectively absent, or even reversed, at TMRs of 0 and +6 dB.

The SRTs for the LP and HP conditions were obtained by averaging the SRTs from logistic fits to the individual data. In the LP condition, the SRT was −0.5 dB in the steady noise and −3.7 dB in the single-talker masker, resulting in an overall masking release of only 3.2 dB. The SRT for the LP single-talker masker, as estimated using a logistic function does not match the "visual" SRT well (estimated by where the interpolated data points in Fig. 2 cross the 50% line), mainly because the logistic function cannot capture the apparent dip in performance at a TMR of 0 dB. This dip was not found in the HP-filtered conditions, where the estimated SRTs for the steady noise (−0.7 dB) and the single talker

(−3.9 dB) agree better with visual inspection of the data in Fig. 2, to again produce an estimated masking release of 3.2 dB.

In summary, masking release was similar in both the LP and HP conditions and was reduced relative to the unfiltered conditions. The results suggest that masking release in normal-hearing listeners can be severely reduced simply by limiting the available frequency spectrum, without any further degradation. The results are not consistent with our initial prediction that the LP-filtered condition should result in greater masking release than that observed in the HP-filtered conditions, because of the presence of low-order resolved harmonics in the LP, but not in the HP, condition.

There are a number of possible reasons for the failure of the results to match predictions. One possible explanation is that the HP filter cutoff was not set high enough, and that some resolved harmonics were present in that condition. To test for this, we carried out an *F*0 discrimination experiment using complex tones with the same long-term spectral envelope as that of the target speech. As described in the Appendix, we found very pronounced differences between the LP- and HP-filtered conditions, with large (poor) *F*0 difference limens (DLs) and strong phase dependencies in the HP conditions, as expected from unresolved harmonics, and small (good) *F*0 DLs and no phase dependence in the LP conditions, as expected from resolved harmonics. Thus, it seems that the filters used in this experiment were successful at restricting the stimuli in the HP condition to only unresolved harmonics.

Another possible explanation is that the average *F*0s of the target (110 Hz) and interferer (106 Hz) were very similar in this experiment. It is possible that the pitch differences were not sufficient to provide a strong segregation cue, and that other cues that help us to identify different talkers, such as vocal tract length (Darwin *et al.*, 2003), were sufficient to provide whatever small masking release was observed in both LP- and HP-filtered conditions. In fact, the similarity in average *F*0 between the target and masker may explain the apparent dip in performance at 0 dB TMR in the LP condition: it may be that confusion between the target and masker was greatest when there were no overall level differences between the two (e.g., Brungart, 2001) and that other cues, such as vocal tract length, were conveyed better in the HP than in the LP condition. This idea is tested in experiment 2 by artificially lowering the *F*0 of the interferer, while retaining all other speaker characteristics, such as vocal tract size, in order to increase the pitch difference between the target and interferer.

III. EXPERIMENT 2: INTELLIGIBILITY IN A SINGLE-TALKER MASKER WITH SHIFTED *F*0

A. Methods

1. Subjects

Nine subjects (four female) participated in this experiment. None had previously listened to HINT sentences. The subjects all reported normal hearing and had audiometric thresholds of 20 dB HL or less at octave frequencies between 250 and 8000 Hz. Their ages ranged from 19 to 31, with a mean age of 21.8. Subjects were either paid or given course credit for their participation.

2. Stimuli and procedure

The stimuli were very similar to those used for the single-talker condition in experiment 1, with the exception that the *F*0 of the masker was lowered by four semitones using PRAAT software (Boersma and Weenink, 2008). The processed-*F*0 interferer had an average *F*0 of 85 Hz, with an average within-sentence standard deviation of 22 Hz, and was thus quite far removed from the average *F*0 of the target (110 Hz). As before, the interfering speech was spectrally shaped to match the long-term spectrum of the target speech. Subjects heard all the sentences in a single-talker background, at four different TMRs, and each subject was tested using all three filter conditions (unfiltered, HP filtered, and LP filtered). The order of presentation was counterbalanced across subjects for the filter type and always progressed from highest (easiest) to lowest (hardest) TMR. As in experiment 1, the LP-filter cutoff was 1200 Hz, the HP-filter cutoff was 1500 Hz, and the complementary off-frequency noise was presented in each of the filtered conditions. Because the interferer *F*0 was lower than in experiment 1, the harmonics of the interferer should have been even less resolved than those of the interferer in experiment 1. As in experiment 1, subjects were provided with approximately 1 h of training with feedback (on a different sentence corpus) before testing.

B. Results and discussion

The results are shown in Fig. 3, with the data from the steady-noise masking conditions from experiment 1 replotted as small symbols for comparison. The left panel shows results from the unfiltered condition, and the right panel shows the results from the filtered conditions. In the unfiltered condition, the shift in the interferer *F*0 resulted in about the same amount of masking release as was found in the original unfiltered conditions, with a mean estimated SRT of −18.5 dB (compared with −19.5 dB in the unprocessed condition). In contrast, for both the LP- and HP-filtered stimuli, more masking release was observed with the shifted-*F*0 maskers than with the original single-talker maskers. A comparison of the shifted and original single-talker conditions indicated a main effect of *F*0 shift, with improved performance using the shifted-*F*0 masker [main effect of shift: $F_{1,15}=4.70$, $p=0.047$ (LP) and $F_{1,15}=6.08$, $p=0.03$ (HP)]. As found in experiment 1, LP and HP filtering resulted in very similar performance overall. A within-subject ANOVA comparing LP- and HP-filtered conditions with the shifted-*F*0 interferer confirmed no significant difference between the two ($F_{1,8}<1$, n.s.), with no significant interaction between filter condition and TMR ($F_{3,24}=1.809$; $p=0.173$). Analysis in terms of SRT yielded similar conclusions. The SRTs, estimated from the logistic fits to the individual data, were −7.1 and −6.9 dB (resulting in a release from masking of about 6.6 and 6.2 dB) for the LP and HP conditions, respectively. Interestingly, in this case there was no apparent difference in performance between the LP and HP conditions

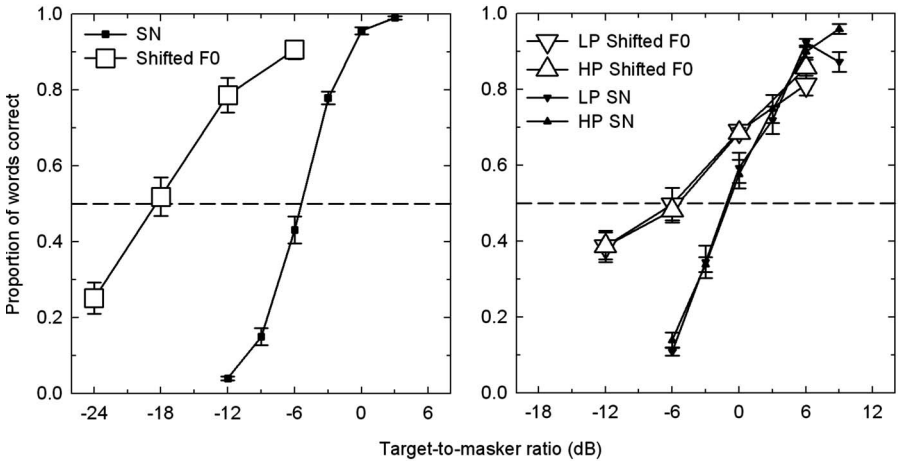


FIG. 3. Proportion of words correctly reported from HINT sentence lists as a function of TMR in conditions where the *F*0 of the single-talker interferer was shifted down by four semitones. Results in steady noise from Figs. 1 and 2 are replotted as small symbols for comparison. The left panel shows results from unfiltered conditions, and the right panel shows results from the LP- and HP-filtered conditions. Error bars represent ±1 standard error of the mean across subjects.

at a TMR of 0 dB, as was observed in experiment 1. It may be that the confusion in experiment 1 between the target and masker was greater in the LP than in the HP condition, and that a greater difference in average *F*0 between the target and masker reduced the confusion, and hence the difference between the LP and HP conditions at a 0 dB TMR.

In summary, shifting the *F*0 of the interfering talker increased the amount of masking release in both LP- and HP-filtered conditions. However, the overall amount of masking release was still very similar in both conditions. As in experiment 1, the results are not in line with our prediction that the more salient and accurate pitch conveyed by the lower-order resolved harmonics should lead to greater masking release in the LP- than in the HP-filtered conditions.

There are again a number of potential reasons why our expectations were not met. For instance, it may be a coincidence that the LP and HP conditions in experiments 1 and 2 both produce very similar amounts of masking release. It should be remembered that the low and high spectral regions can convey very different forms of speech information, even if the overall intelligibility is the same (e.g., Northern and Downs, 2002). In fact, both the LP and HP conditions lack a considerable portion of the spectrum between 300 and 3500 Hz, which is considered to contain the most important speech information (French and Steinberg, 1947; Assmann and Summerfield, 2004). Also, even at an acoustic level, the strength of masker fluctuations may vary somewhat in different frequency regions; for instance, Festen and Plomp (1990) found that there was typically more speech modulation energy in high-frequency regions than in low-frequency regions. Thus, it is possible that the greater amount of pitch information in the low-frequency region may be offset to some degree by the greater masker modulation depth (and hence greater scope for masking release) in the high-frequency region. Nevertheless, it seems clear that the temporal fine-structure information from low-order harmonics in voiced speech does not play a dominant role in determining masking release, at least when comparing steady-noise maskers to single-talker interferers. The aim of the final experiment was to generalize our findings to another condition for which masking release is often measured, involving masking by amplitude-modulated noise.

IV. EXPERIMENT 3: MASKING RELEASE IN SPEECH-MODULATED NOISE

A. Rationale

Hearing impairment, cochlear implants, and simulations of cochlear implants using channel vocoders all lead to reduced masking release, relative to normal hearing, when comparing modulated noise with steady-state noise maskers (e.g., Festen and Plomp, 1990; Bacon *et al.*, 1998; Peters *et al.*, 1998; Qin and Oxenham, 2003; Nelson and Jin, 2004; Jin and Nelson, 2006; Rhebergen *et al.*, 2006). Various types of modulated noise have been used in previous studies, including sinusoidally amplitude (or intensity) modulated noise, square-wave gated noise, and noise that is modulated by the broadband temporal envelope of speech from a single talker. As with masking release using single-talker interferers, it has been proposed that the reduced ability of hearing-impaired listeners and cochlear-implant users to take advantage of temporal dips in a modulated noise may be related to reduced access to temporal fine-structure cues (e.g., Qin and Oxenham, 2003; Lorenzi *et al.*, 2006). The rationale is that severely reduced (or absent) temporal fine-structure cues may render it more difficult to distinguish the fluctuations in the (partially periodic) target speech from those of the noise masker.

In this experiment, we measured speech intelligibility in the presence of speech-modulated noise. Because the temporal envelope of BB speech was used, the modulation energy was the same in both low- and high-frequency regions. This enabled us to test whether any differences emerge between masking release in the LP and HP conditions when the modulation energy of the masker is the same in both regions.

B. Methods

1. Subjects

Nine subjects (eight female) participated in this experiment. None had previously listened to HINT sentences. The subjects all reported normal hearing, and had audiometric thresholds of 20 dB HL[3] or less at octave frequencies between 250 and 8000 Hz. Their ages ranged from 19 to 31, with a mean age of 20.7. Subjects were either paid or given course credit for their participation.

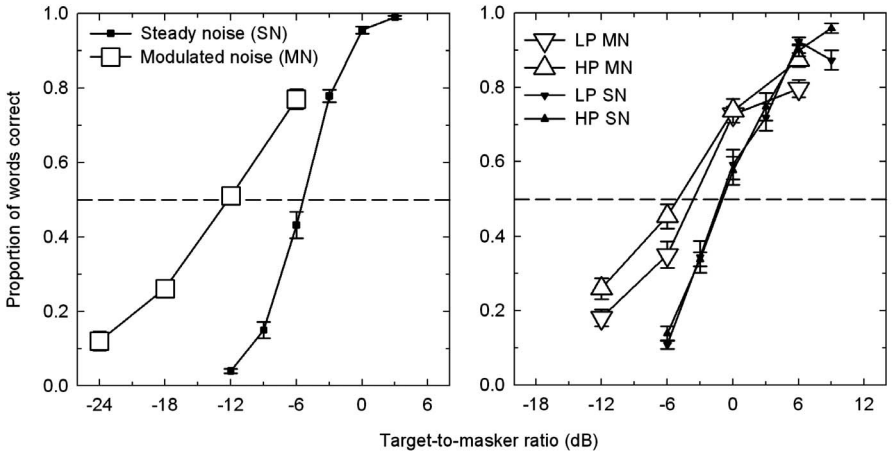


FIG. 4. Proportion of words correctly reported from HINT sentence lists as a function of TMR. Open symbols represent results in the presence of a speech-modulated noise masker. Results in steady noise from Figs. 1 and 2 are replotted as small symbols for comparison. The left panel shows results from unfiltered conditions, whereas the right panel show results from the LP- and HP-filtered conditions. Error bars represent ±1 standard error of the mean across subjects.

2. Stimuli and procedure

Modulated-noise maskers were generated based on the envelopes of the maskers from experiment 1. Envelopes were obtained by half-wave rectifying and LP filtering the BB speech stimulus using a sixth-order Butterworth filter with a cutoff frequency of 50 Hz. The resulting envelopes were used to modulate a noise that was spectrally shaped to have the same long-term power spectrum as the target HINT sentences.

Subjects heard the target HINT sentences in a modulated-noise background at four different TMRs. As in the previous experiment, each subject was tested under all three filter conditions (unfiltered, HP filtered, and LP filtered) using a LP-filter cutoff of 1200 Hz and a HP-filter cutoff of 1500 Hz, with complementary off-frequency noise in the filtered conditions. The testing session was preceded by a 1 h training session with feedback, as described in experiment 1.

C. Results and discussion

The results are shown in Fig. 4, with the steady-noise data from experiment 1 replotted as small symbols for comparison. The left panel shows results from the unfiltered condition and the right panel shows results from the filtered conditions. The unfiltered results showed a significant amount of masking release, when compared with the speech-shaped noise condition from experiment 1, at the TMRs that both conditions had in common ($F_{1,15}=243.53$; $p<0.0001$), although performance was lower overall than that observed with the single-talker interferer in experiment 1 ($F_{1,15}=95.93$; $p<0.0001$). The finding of less masking release for speech-modulated noise than for a single-talker interferer is consistent with earlier studies (e.g., Peters *et al.*, 1998).

Both the LP and HP conditions showed a significant amount of masking release when compared with the comparable steady-noise conditions from experiment 1 (LP: $F_{1,15}=5.93$, $p=0.028$; HP: $F_{1,15}=16.46$, $p=0.001$). A within-subject comparison of the low-pass and high-pass conditions revealed that, contrary to our initial hypothesis, performance in the HP condition actually exceeded that in the LP condition overall ($F_{1,8}=33.17$, $p<0.001$), although the difference did not exceed ten percentage points at any TMR. There was no interaction between TMR and filter condition ($F_{3,24}=1.10$, $p=0.37$). The mean estimated SRTs were −3.4 and −5.5 dB in the low-pass and high-pass conditions, respectively.

The fact that the LP condition did not result in more masking release than the HP condition when the masker modulation energy was equated across the two regions does not provide support for the idea that the higher masker modulation energy in the HP conditions of experiments 1 and 2 counteracted any benefit of temporal fine structure in the LP conditions. However, the results do not provide strong evidence *against* the hypothesis either. This is because there are many differences between the single-talker interferer of experiments 1 and 2 and the modulated noise of experiment 3, which make a direct comparison problematic. On one hand, the acoustic differences between the modulated noise and the target are much greater than those between the interfering talker and the target. This might imply that the pitch cues encoded in the temporal fine structure may not be as necessary when segregating speech from noise, as compared to segregating speech from speech. In other words, the more similar acoustics, as well as the intelligibility (Rhebergen *et al.*, 2005), of a speech masker may produce more informational masking than a modulated-noise masker, which may in turn be alleviated more by pitch cues. On the other hand, a speech interferer has spectral, as well as temporal, dips, which may enhance the potential amount of masking release. It is not clear which of these (or other) competing factors governs the similarities and differences between the different masker types. Nevertheless, it is clear that our initial findings of no specific benefit of pitch-related temporal fine-structure cues in the LP region generalize from single-talker interferers to temporally modulated noise maskers.

V. GENERAL DISCUSSION

A. Summary of results

Three experiments were carried out to compare word intelligibility in sentences using steady or fluctuating maskers under different spectral conditions (unfiltered, LP filtered, and HP filtered). The cutoff frequencies (LP 1200 Hz; HP 1500 Hz) were selected to produce roughly equal speech intelligibility for the two filtered conditions in the presence of steady noise and to minimize the availability

of peripherally resolved harmonics in the HP condition (see Appendix). Experiment 1 found a substantial release from masking in the unfiltered conditions when the steady noise was replaced by a single-talker interferer. The amount of masking release was greatly reduced in both the LP and HP conditions, and no significant differences between the two filtered conditions were found. Experiment 2 used the same single-talker masker as was used in experiment 1, but artificially lowered the *F*0 of the masker to increase the average *F*0 difference between the target and the masker. Increasing the average *F*0 difference between the target and masker did not affect performance in the unfiltered conditions, where the masking release remained substantial. The increase in *F*0 difference did increase the release from masking found in the two filtered conditions, but it did so by the same amount for both the LP and HP conditions. Experiment 3 measured intelligibility in the presence of a noise masker that was temporally modulated with a broadband speech stimulus. The same speech modulator was used in the unfiltered and in the LP and HP conditions. As in the previous two experiments, more masking release was observed (relative to performance in unmodulated noise) in the unfiltered conditions than in the LP and HP conditions, which again both produced similar amounts of masking release; in fact, the amount of masking release was somewhat greater for the HP than for the LP condition.

B. Possible interpretations and implications

The results are not consistent with the idea that the pitch information carried in the temporal fine structure of low-order resolved harmonics plays a crucial role in masking release. The results do not therefore seem to support earlier conjectures that loss of temporal fine-structure and pitch information may underlie the reduced masking release experienced by hearing-impaired listeners (e.g., Lorenzi *et al.*, 2006) and cochlear-implant users (e.g., Qin and Oxenham, 2003; Stickney *et al.*, 2004).

One possible interpretation is that any form of band limiting of the stimuli may result not only in poorer performance overall but also in a reduction in the advantage that listeners can take of spectrotemporal fluctuations in a masking stimulus. It may be that a large proportion of the masking release observed in BB conditions stems from the inherent redundancy in the speech signal, and that a reduction in that redundancy, by band-limiting the stimuli, reduces the scope for masking release. Loss of redundancy may also explain in part why hearing-impaired listeners, as well as simulated and real cochlear-implant users, typically exhibit less (or no) masking release. It may be, for instance, that impaired frequency resolution is sufficient to reduce redundancy to a point where little or no masking release is observed.

Another, not mutually exclusive, possibility is that masking release can only occur at relatively low TMRs. It is noteworthy that in all our conditions (filtered and unfiltered), masking release was only observed at TMRs of 0 dB or less. In the unfiltered conditions, the interpretation is made difficult by the fact that performance was near ceiling at 0 dB TMR. However, even in these conditions there is some hint of a "crossover" around 0 dB TMR, as is more clearly observed in the filtered conditions, where the single-talker interferer actually becomes a more effective masker than the steady noise at high TMRs. As suggested by Kwon and Turner (2001), masker fluctuations can in principle hinder as well as help performance, by introducing spurious fluctuations into the target. In cases where the TMR is positive, where there is little scope for confusion between the target and the masker, it may be that the detrimental effects of spurious masker fluctuations outweigh any benefit of dips in the masker energy. A potentially related finding was recently reported by Freyman *et al.* (2008) in the domain of spatially induced (as opposed to *F*0-induced) masking release. Using noise-excited envelope-vocoded stimuli, they found that perceived spatial separation between target speech and interfering speech only improved performance when TMRs were negative. For tasks that involved highly positive TMRs, no benefit of perceived spatial separation was found.

Interestingly, it appears that pitch does play some role in masking release: when the *F*0 of the masker was shifted away from that of the target, masking release increased in both the LP and HP conditions. However, the improvement due to the pitch difference was the same in both conditions. This suggests that although the pitch of unresolved high-order harmonics is weaker than that for low-order harmonics, it is still sufficient for source segregation in dynamic situations, such as speech, where the moment-by-moment *F*0 differences between the sources are typically much larger than the *F*0 DL.

The conclusion that the pitch from unresolved harmonics may be sufficient for speech segregation, at least in certain circumstances, is consistent with basic psychoacoustic studies of stream segregation using only unresolved harmonics (Vliegen *et al.*, 1999; Vliegen and Oxenham, 1999). However, it seems at odds with the conclusions of Deeks and Carlyon (2004) and Qin and Oxenham (2005). Both these studies used different forms of envelope vocoders to study the role of pitch perception in speech segregation. Deeks and Carlyon (2004) used six-band envelope vocoders, excited by pulse trains with different *F*0s (or rates). They found no benefit of presenting the masker and target speech on different *F*0s in separate channels. However, their stimuli were much more impoverished than those used here: only three broad frequency bands were available for the target. Although this may be a realistic scenario for cochlear-implant users, it is not possible to conclude that it was simply the lack of resolved harmonics that led to poor performance and no benefit of *F*0 differences. Qin and Oxenham (2005) found that the poorer *F*0 discrimination produced by noise-excited envelope vocoding resulted in an inability to make use of *F*0 differences between two simultaneously presented synthetic vowel stimuli. There are at least two reasons why they did not observe a benefit of *F*0 difference while we did. First, modulated-noise stimuli from an envelope vocoder produce an even poorer pitch percept than unprocessed unresolved harmonics, meaning that the pitch information in the Qin and Oxenham (2005) study was more impoverished than that in the present study. Second, it may be that pitch information is more beneficial in sequential situations than in simultaneous

ones. For synchronously gated pairs of vowels, *F*0 information from two sources must be extracted simultaneously; for more natural speech situations, much can be gained from "glimpsing" in time, as well as frequency, such that pitch differences across time can be used to form perceptual streams corresponding to the two sources.

The studies of masking release by Scott *et al.* (2001) and Elangovan and Stuart (2005) came to somewhat different conclusions from the ones reached here. Comparing masking by a steady noise with masking by a randomly interrupted noise masker, Scott *et al.* (2001) found that LP filtering reduced masking release, whereas Elangovan and Stuart (2005) found that HP filtering did not. Their results were discussed in terms of potentially poorer temporal resolution at low frequencies. There are several differences between our study and theirs. First, single-word recognition was tested, as opposed to sentences in the present study. Second, their noise was gated on and off abruptly, whereas we used either a single talker or a noise modulated by the broadband envelope of a single talker. It may be that the limits of temporal resolution (if any) are more apparent in abruptly gated maskers because the limits of temporal resolution (as opposed to the temporal slopes of the stimuli) are more likely to govern performance for abruptly gated maskers than for maskers with more shallow envelope slopes. Finally, it appears from Fig. 2 in Elangovan and Stuart (2005) that the spectra of their maskers and targets were not as carefully matched as in the present study, leaving open the possibility that some of the differential effects of filtering were due to spectrally local differences in long-term TMR. Thus, although the conclusions differ from ours, their data do not necessarily contradict our findings.

The results of Freyman *et al.* (2001) are somewhat related to, and are also generally consistent with, our findings. They found that the release from masking obtained by introducing a perceived spatial separation between a two-talker masker and a target speaker was no greater (in fact, somewhat less) when the stimuli were HP filtered at 1500 Hz than when they were unfiltered. This suggests that the masker and target were no more confusable after HP filtering, despite the reduction in the availability of low-frequency information. However, the use of female speech (with a higher *F*0) and the absence of LP masking noise make it difficult to rule out the possibility that some low-frequency resolved harmonics were available in the stimuli of Freyman *et al.* (2001).

Whatever the reasons underlying the loss of masking release in the present experiments, it seems clear that simply bandlimiting the stimuli can be enough to severely reduce the masking release experienced by normal-hearing subjects. This implies that the proportion of the speech spectrum that is audible may itself be a significant predictor of how much masking release a given hearing-impaired listeners will demonstrate (e.g., Bacon *et al.*, 1998), as well as the many other potential factors, such as frequency and temporal resolution, that have been investigated in the past (e.g., George *et al.*, 2006, 2007).

VI. CONCLUSIONS

In unfiltered conditions, substantial release from masking was found when a steady-noise masker was replaced by a single-talker interferer or a speech-modulated-noise masker with the same long-term spectrum. The amount of masking release was greatly reduced in both LP- and HP-filtered conditions. In contrast to predictions based on the hypothesized role of pitch and temporal fine structure in masking, masking release in LP conditions was not significantly greater than in the HP conditions in any of the experimental conditions. This suggests that the temporal fine structure of resolved harmonics does not play a crucial or unique role in masking release. It may instead be that the amount of perceptual redundancy in the target determines masking release, or that any stimulus manipulation (or hearing condition) that increases the TMR necessary for intelligibility in steady noise to levels higher than 0 dB results in a loss of masking release.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (Grant No. R01 DC 05216). We thank Christophe Micheyl, Neal Viemeister, and Koen Rhebergen for many useful discussions, and Christian Lorenzi, Richard Freyman, Christophe Micheyl, and two anonymous reviewers for helpful comments on an earlier version of this paper.

APPENDIX: *F*0 DIFFERENCE LIMENS AND HARMONIC RESOLVABILITY WITH SPEECH-SHAPED HARMONIC TONE COMPLEXES

1. Methods

a. Stimuli

In this experiment *F*0 DLs were measured for harmonic tone complexes that were spectrally shaped to match the long-term power spectrum of the speech materials used in the main experiments. Equal-amplitude harmonic tone complexes, extending up to 10 kHz, were passed through a filter to provide them with the same spectral envelope as sentences from the HINT (Nilsson *et al.*, 1994). Two values of *F*0 were tested, 110 and 135 Hz, which represent the mean *F*0 and one standard deviation above the mean *F*0 of the target speech, respectively. After spectral shaping, the complexes were either presented without any additional filtering (BB condition), or were LP or HP filtered, using a fourth-order Butterworth filter. When used, the LP filter cutoff was 1200 Hz and the HP filter cutoff was 1500 Hz, as in the main speech experiments. The components of the complex tones were added either in sine phase or in random phase. In the case of random phase, the phase of each component was chosen independently and randomly with uniform distribution from 0 to 2π for each stimulus presentation.

In the LP and HP conditions, a complementary noise was added to reduce the possibility of off-frequency listening and the detection of distortion products. This noise had the same speech-shaped spectral envelope as the tone complex but was 12 dB lower than the nominal level of the complex (before LP or HP filtering). In the LP condition, the noise

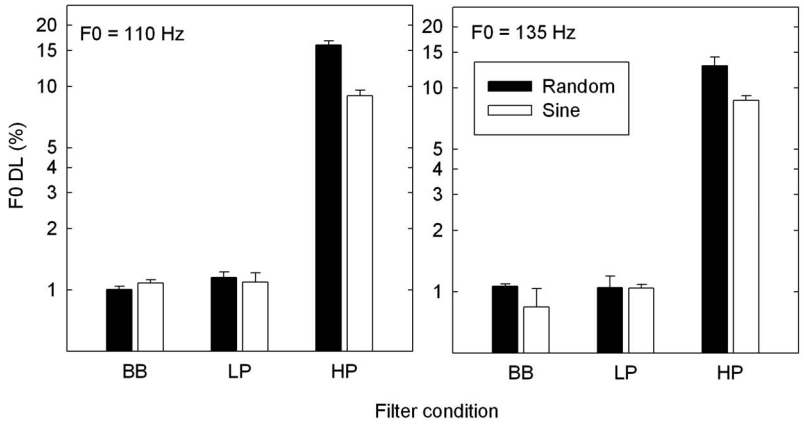


FIG. 5. $F0$ DLs for tone complexes spectrally shaped to have the same spectral envelope as the target HINT sentences. Filter conditions include unfiltered (BB), LP filtered, and HP filtered. Solid bars represent conditions with components in random starting phase, and open bars represent conditions in sine starting phase. Error bars represent one standard error of the mean across subjects.

was HP filtered at 1200 Hz; in the HP condition, the noise was LP filtered at 1500 (fourth-order Butterworth filter in all cases). Thus the long-term spectral characteristics of the stimuli in this experiment were the same as the target and complementary noise in the main speech experiments. The tone complexes were 400 ms in duration, including 50-ms raised-cosine (Hanning) onset and offset ramps. The complementary noise was gated synchronously with the HP- or LP-filtered tone complex.

b. Procedure

Thresholds were measured using a two-interval two-alternative forced-choice procedure. In each trial, listeners were presented with two presentations of the tone complex, separated by a silent interval of 300 ms. The baseline $F0$ ($F0_{base}$) was roved over two semitones with uniform distribution across each trial (but not across intervals). The $F0$s in the two intervals were $F0_{base}(1+\Delta/100)^{1/2}$ and $F0_{base}(1+\Delta/100)^{-1/2}$, where Δ is the difference in $F0$ between the two intervals, in percent. The listeners' task in each trial was to select the interval with the higher $F0$. Thresholds were estimated using a 3-down 1-up adaptive tracking procedure that estimates the 79.4% correct point of the psychometric function (Levitt, 1971). The initial value of Δ was 15.85%. This value was decreased after three consecutive correct responses and increased after each incorrect response. The transition from decreasing to increasing (or from increasing to decreasing) values defines a reversal. The initial step size was an increase or a decrease in Δ by a factor of 2.51. After two reversals, the step size decreased to a factor of 1.58, and after two more reversals, the step size was reduced to its final value of a factor of 1.26. A run terminated after six reversals at the final step size, and threshold was defined as the geometric mean value of Δ at those final six reversals.

The average overall level of the complex tone was set to 75 dB SPL but was roved over a uniform 6 dB range across intervals and trials. The level and $F0$ roving were introduced to avoid the possibility that listeners learned to respond correctly based on cues other than $F0$ (such as a small increase in loudness as a harmonic moved to the peak of the spectral envelope) and to simulate some aspects of the dynamic nature of speech, where the mean level may not be wholly representative. As in the speech experiments, the level of the complementary background noise was not roved.

The stimuli were presented diotically to listeners seated in a double-walled sound-attenuating booth. Stimuli were generated digitally and converted to voltage using a Lynx-Studio Lynx22 sound card with 24 bit resolution at a sampling rate of 22.05 kHz. The stimuli were either fed directly to HD580 headphones (Sennheiser) or were first passed through a headphone amplifier (TDT HB7). In both cases, levels were verified using a voltmeter and a sound pressure level meter coupled to the headphones using an artificial ear (Brüel and Kjær).

c. Subjects

Four normal-hearing listeners (all female; ages between 20 and 41) participated as subjects in this experiment. One was the second author; the others were paid for their services and had also participated in one of the speech experiments.

2. Results

The mean results, geometrically averaged across listeners, are shown in Fig. 5. In line with earlier studies using equal-amplitude harmonic tone complexes, $F0$ DLs in the BB condition were all low ($\sim$1%) and showed no systematic effect of component phase relationship. In the LP conditions, $F0$ DLs were similarly low, and also showed no phase dependencies. In contrast, $F0$ DLs in the HP conditions were higher by nearly an order of magnitude, and showed a systematic phase effect, being higher (worse) in the random-phase conditions than in the sine-phase conditions. These results confirm that pitch coding accuracy for harmonic tone complexes with the same long-term spectral shape as the speech stimuli used in the main experiments is considerably worse in the HP-filtered conditions than in the LP-filtered or BB conditions. This is true for the $F0$ corresponding to the mean $F0$ of the sentences and for an $F0$ one standard deviation above the mean. The poor thresholds and phase dependence in the HP-filtered conditions suggest that only unresolved harmonics were accessible to the subjects in the main speech experiments.

[1]Some studies define the SRT as the TMR at which 50% of whole sentences (as opposed to words) are reported with no errors (e.g., Bacon *et al.*, 1998; Peters *et al.*, 1998).

[2]Results from the ANOVAs are quoted using the Greenhouse–Geisser correction for sphericity where appropriate (with original degrees of freedom cited).

[3]One subject had a threshold of 25 HL at 8000 Hz in one ear and a threshold of 25 HL at 4000 Hz in the other ear. However, the subject's data did not deviate significantly from those of others in that group.

Assmann, P. F. (**1999**). "Fundamental frequency and the intelligibility of competing voices," in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, pp. 179–182.

Assmann, P. F., and Summerfield, A. Q. (**2004**). "The perception of speech under adverse conditions," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. Fay (Springer, New York), pp. 231–308.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," J. Speech Lang. Hear. Res. **41**, 549–563.

Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," J. Acoust. Soc. Am. **113**, 3323–3334.

Bernstein, J. G., and Oxenham, A. J. (**2006**). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," J. Acoust. Soc. Am. **120**, 3929–3945.

Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.

Boersma, P., and Weenink, D. (**2008**). "Praat: Doing phonetics by computer," http://www.praat.org/ (Last viewed April 2008).

Brokx, J. P., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Carlyon, R. P. (**1996**). "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," J. Acoust. Soc. Am. **99**, 517–524.

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.

Deeks, J. M., and Carlyon, R. P. (**2004**). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation," J. Acoust. Soc. Am. **115**, 1736–1746.

Duquesnoy, A. J. (**1983**). "Effect of a single interfering noise or speech source on the binaural sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74**, 739–743.

Elangovan, S., and Stuart, A. (**2005**). "Interactive effects of high-pass filtering and masking noise on word recognition," Ann. Otol. Rhinol. Laryngol. **114**, 867–878.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2008**). "Spatial release from masking with noise-vocoded speech," J. Acoust. Soc. Am. **124**, 1627–1637.

George, E. L., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295–2311.

George, E. L., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., and Houtgast, T. (**2007**). "Auditory and nonauditory factors affecting speech reception in noise by older listeners," J. Acoust. Soc. Am. **121**, 2362–2375.

Gnansia, D., Jourdes, V., and Lorenzi, C. (**2008**). "Effect of masker modulation depth on speech masking release," Hear. Res. **239**, 60–68.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Jin, S. H., and Nelson, P. B. (**2006**). "Speech perception in gated noise: The effects of temporal resolution," J. Acoust. Soc. Am. **119**, 3097–3108.

Kwon, B. J., and Turner, C. W. (**2001**). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," J. Acoust. Soc. Am. **110**, 1130–1140.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Micheyl, C., Bernstein, J. G., and Oxenham, A. J. (**2006**). "Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise," J. Acoust. Soc. Am. **120**, 1493–1505.

Micheyl, C., and Oxenham, A. J. (**2007**). "Across-frequency pitch discrimination interference between complex tones containing resolved harmonics," J. Acoust. Soc. Am. **121**, 1621–1631.

Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.

Nelson, P. B., and Jin, S. H. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.

Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.

Nilsson, M., Soli, S., and Sullivan, J. A. (**1994**). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Northern, J., and Downs, M. (**2002**). *Hearing in Children* (Lippincott, Williams & Wilkins, New York).

Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.

Plomp, R. (**1967**). "Pitch of complex tones," J. Acoust. Soc. Am. **41**, 1526–1533.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.

Qin, M. K., and Oxenham, A. J. (**2006**). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," J. Acoust. Soc. Am. **119**, 2417–2426.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2005**). "Release from informational masking by time reversal of native and non-native interfering speech," J. Acoust. Soc. Am. **118**, 1274–1277.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2006**). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," J. Acoust. Soc. Am. **120**, 3988–3997.

Scott, T., Green, W. B., and Stuart, A. (**2001**). "Interactive effects of low-pass filtering and masking noise on word recognition," J. Am. Acad. Audiol. **12**, 437–444.

Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.

Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F. G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.

Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Summers, V., and Leek, M. R. (**1998**). "F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss," J. Speech Lang. Hear. Res. **41**, 1294–1306.

Summers, V., and Molis, M. R. (**2004**). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47**, 245–256.

Takahashi, G. A., and Bacon, S. P. (**1992**). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," J. Speech Hear. Res. **35**, 1410–1421.
Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (**1999**). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," J. Acoust. Soc. Am. **106**, 938–945.
Vliegen, J., and Oxenham, A. J. (**1999**). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am. **105**, 339–346.
Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Helfer, K. S. (**2007**). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2388.

English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy

Kota Hattori[a] and Paul Iverson
Division of Psychology and Language Sciences, University College London, 2 Wakefield Street, London WC1N 1PF, United Kingdom

(Received 1 November 2007; revised 7 October 2008; accepted 8 October 2008)

Native speakers of Japanese often have difficulty identifying English /r/ and /l/, and it has been thought that second-language (L2) learning difficulties like this are caused by how L2 phonemes are assimilated into ones native phonological system. This study took an individual difference approach to examining this relationship by testing the category assimilation of Japanese speakers with a wide range of English /r/-/l/ identification abilities. All Japanese subjects were assessed in terms of (1) their accuracy in identifying English /r/ and /l/, (2) their assimilation of /r/ and /l/ into their Japanese flap category, (3) their production of /r/ and /l/, and (4) their best-exemplar locations for /r/, /l/, and Japanese flap in a five-dimensional set of synthetic stimuli (F1, F2, F3, closure duration, and transition duration). The results demonstrated that Japanese speakers assimilate /l/ into their flap category more strongly than they assimilate /r/. However, there was little evidence that category assimilation was predictive of English /r/-/l/ perception and production. Japanese speakers had three distinct best exemplars for /r/, /l/, and flap, and only their representation of F3 in /r/ and /l/ was predictive of identification ability.
 [DOI: 10.1121/1.3021295]

PACS number(s): 43.71.Hw, 43.71.Es [AJ] Pages: 469–479

I. INTRODUCTION

Adult second-language (L2) learners can have difficulties learning L2 phonemes. For example, native speakers (L1) of Japanese are often poor at identifying and producing the English consonants /ɹ/ and /l/ (e.g., Goto, 1971; Miyawaki *et al.*, 1975), although they can somewhat overcome these problems by receiving phonetic training with talker and stimulus variability (e.g., Hazan *et al.*, 2005; Iverson *et al.*, 2005; Lively *et al.*, 1993; Logan *et al.*, 1991) or by having long periods of exposure to English-speaking environments (e.g., Yamada, 1995). However, even highly experienced Japanese speakers can have difficulty identifying English /ɹ/ and /l/ at a native speaker's level (e.g., Yamada and Tohkura, 1992).

Some current L2 speech perception models attribute such learning difficulties to the similarity between L1 and L2 phonetic or phonemic categories. For example, Best's perceptual assimilation model (PAM) (Best, 1995; Best *et al.*, 2001) states that L2 phonemes are perceived based on their articulatory similarity to the listener's closest L1 phonemes; pairs of L2 phonemes are particularly hard to distinguish when they both sound like equally good examples of the same L1 category. Likewise, Flege's speech learning model (SLM) (Flege, 1995) states that the learner's L1 and L2 categories interact in a common phonological system; learning L2 categories is particularly difficult when there is an existing L1 category that is similar, and learning is easier when the new L2 categories fall into relatively unoccupied regions of the listener's phonological space.

Previous research has suggested that these kinds of similarity relationships may be the cause of Japanese learners' problems with English /ɹ/-/l/. Japanese has an apicoalveolar tap /ɾ/ (Vance, 1987) that has been described as being related to English /ɹ/ and /l/. Best and Strange (1992) hypothesized that both English /ɹ/ and /l/ are perceived as poor exemplars of this Japanese phoneme (or as the Japanese /w/); according to PAM, English /ɹ/ and /l/ are hard to perceptually distinguish because they are the same with regard to the Japanese phonological system. Subsequent assimilation rating experiments have generally supported this view (Guion *et al.*, 2000; Iverson *et al.*, 2003); when Japanese listeners hear English /ɹ/ and /l/ stimuli and are asked how close these are to Japanese phonemes, they rate that both are poorly related to Japanese /ɾ/ or /ɯr/ (i.e., the Japanese phoneme preceded by a vowel). Aoyama *et al.* (2004) noted that there are trends for Japanese speakers to rate English /l/ as being somewhat closer to Japanese /ɾ/, even though both are poor exemplars, and also found that Japanese children were better at learning English /ɹ/ than /l/. This fits with SLM's claim that learning difficulties are caused by an interaction between L1 and L2 categories; English /ɹ/ is thought to be somewhat easier to learn because the formation of this category is less affected by the listener's existing Japanese /ɾ/ category.

The present study further investigated the role of assimilation by testing Japanese adults who have a wide range of English /ɹ/-/l/ identification abilities and by determining whether these individual differences are related to the degree that they assimilate these phonemes into their L1 /ɾ/ category. If there is a causal relationship between L1 assimilation and L2 category learning, it seems likely that individuals who assimilate these phonemes less strongly into their L1 /ɾ/ category would be better at identifying and producing Eng-

[a]Author to whom correspondence should be addressed.

lish /ɹ/ and /l/. That is, inexperienced Japanese learners of English have been claimed to assimilate all of these phonemes into the same category (e.g., Best and Strange, 1992). Individuals who have learned to accurately identify English /ɹ/ and /l/ may thus have learned to dissimilate them from Japanese /ɾ/ (e.g., Flege, 1995; Flege *et al.*, 2003), allowing these individuals to create three independent categories within the same phonological space. Alternatively, it is possible that individuals who are more accurate with English /ɹ/ and /l/ have changed their assimilation patterns asymmetrically. For example, auditory training results (Iverson *et al.*, 2005) suggest that Japanese listeners who are learning English /ɹ/-/l/ may actually increase their assimilation of /l/ into the Japanese /ɾ/ category, but not change their assimilation of /ɹ/. Successful learners may thus merge English /l/ and Japanese /ɾ/ into a single compromise category rather than dissimilate (see Flege, 2003; MacKay *et al.*, 2001), and learn a separate category for English /ɹ/.

Previous work has quantified assimilation by having individuals identify and rate which L1 categories are most similar (e.g., Guion *et al.*, 2000; Iverson *et al.*, 2003), but these types of subjective ratings may not be appropriate for individual difference studies. For example, if a Japanese listener rates that /l/ is more similar to Japanese /ɾ/ than does another Japanese listener, it is not clear whether the difference between their rating scores would reflect a real difference in the degree of assimilation or a superficial difference in their use of the scale (e.g., a bias to use the middle of the scale rather than the extremes). Such biases are essentially factored out in traditional studies that average across individuals. Alternatively, acoustic analyses of productions (e.g., Lotto *et al.*, 2004) could be used to measure category assimilation (e.g., one could measure whether an individual's production of English /l/ is acoustically distinct from their Japanese /ɾ/). However, such a measure has at least three disadvantages. One is that English /ɹ/, /l/, and Japanese /ɾ/ cannot always be compared on common acoustic dimensions (e.g., Japanese /ɾ/ productions can sometimes have a burst that is not typical in English /ɹ/ and /l/). Another disadvantage is that between-talker differences (e.g., vocal tract length and speaking rate) introduce variability that is hard to control. Finally, speech production does not always correlate strongly with speech perception (e.g., Bradlow *et al.*, 1997), so it is not clear whether acoustic analyses of speech production would be relevant for perception.

We instead took two new approaches to measuring category assimilation. Our first approach was to use a bilingual identification task including English /ɹ/, English /l/, and Japanese /ɾ/ as response alternatives on the same trial. This allowed us to determine, for example, the extent to which individuals actually confused English /l/ with Japanese /ɾ/, rather than measuring whether they are subjectively related. If English /l/ and Japanese /ɾ/ were frequently confused, this would indicate that the stimuli mapped onto the same underlying phonetic categories (i.e., strong category assimilation). This method, however, may be less effective for measuring weaker degrees of assimilation, where stimuli are subjectively judged to be a poor exemplar of an L1 category without being similar enough to actually sound like that category.

Our second approach was to have individuals find best exemplars for /ɹa/, /la/, and /ɾa/ in a large five-dimensional acoustic space that varied F1, F2, and F3 frequencies during the initial closure, and the durations of the closure and transition, using efficient search methods previously developed for vowels (Evans and Iverson, 2004, 2007; Iverson and Evans, 2003, 2007; Iverson *et al.*, 2006). The approach is similar to acoustically measuring productions, but it has the advantages of mapping all ratings onto the voice of a single speaker and a set of common dimensions, as well as being a perceptual task that may thus be more relevant to identification accuracy. If listeners have similar best exemplars for English /l/ and Japanese /ɾ/, for example, this would again suggest that they have a single category representation for each phoneme.

The same set of Japanese adults participated in a series of three experiments. Experiment 1 measured their baseline identification ability for English /ɹ/-/l/, and used a three-alternative bilingual identification task to measure assimilation. Experiment 2 assessed the spoken accent of these subjects. Experiment 3 mapped best exemplars of the L1 and L2 phonemes in a five-dimensional space. The aim was to determine whether the degree to which Japanese adults assimilate English /ɹ/ and /l/ into their Japanese /ɾ/ category is related to their ability to identify and produce these English phonemes.

II. EXPERIMENT 1: PHONEME IDENTIFICATION

The baseline English /ɹ/-/l/ identification abilities of subjects were assessed by eliciting forced-choice /ɹ/-/l/ judgments for initial-position minimal-pair words (e.g., *rock* or *lock*) that were recorded from multiple native speakers of British English.

The degree of assimilation between the English and Japanese phonemes was measured by having listeners make a three-way bilingual forced-choice judgment between English /ɹ/, English /l/, and Japanese /ɾ/. In the interests of removing any cues for whether the stimuli were from English or Japanese that were not related to the consonant, CV syllables were used with vowels that are common to both languages (i.e., /i/, /e/, /ɑ/, /o/, and /u/), and highly skilled Japanese-English bilingual speakers recorded all phonemes, rather than having separate groups of Japanese and English speakers. Listeners heard all stimuli mixed in the same block, and judged which of the three consonants they heard.

A. Method

1. Subjects

Thirty-nine adult native speakers of Japanese were tested; 3 were omitted from the data due to computer problems, leaving a total of 36 participants. Their ages ranged from 19 to 48 years (median=25 years), they started learning English between 6 and 13 years (median=13 years), and had received English instruction for 7–25 years (median =9 years). All participants were brought up in monolingual environments in Japan, and all participants were tested in London. They had lived in English-speaking countries be-

tween 1 month and 13 years (median=3 months). None of the participants reported having hearing problems.

2. Stimuli and apparatus

Stimulus recordings were made in an anechoic chamber with 44 100 16 bit samples/s. The stimuli were recorded with Radio Spares (RS) 249-946 microphone, Edirol USB Audio Capture UA-25, and Dell Optiplex GX 260. The stimuli were played to subjects using Sennheiser HD 280 headphones while subjects sat in a sound-treated room. Responses were recorded using PRAAT (Boersma and Weenik, 2008).

The English stimuli were initial-position /ɹ/-/l/ minimal-pair words (e.g., rock and lock) used previously by Iverson *et al.* (2005). Four British English speakers (two male and two female) recorded a total of 120 minimal-pair words (words were not repeated between speakers).

The bilingual stimuli were 15 CV syllables created by combining three consonants (i.e., /ɹ/, /l/, and /ɾ/) and five vowels (i.e., /i/, /e/, /ɑ/, /o/, and /u/). Five English-Japanese bilingual speakers made the recordings. Two were native speakers of British English (one male and one female) who had majored in Japanese at UK universities and had studied abroad in Japan. The others (two male and one female) were native speakers of Japanese who had lived in the UK for more than 15 years; two of them moved to the UK before age 5, and the other speaker immigrated to the UK in his 30 s. All speakers recorded two repetitions of each syllable. All stimuli were screened by native British and Japanese phonetically trained listeners, to ensure that all consonants sounded like normal native versions. One speaker (native Japanese) was unable to correctly produce the Japanese /ɾ/ followed by /i/; these recordings were omitted, leaving a total of 148 stimuli (15 syllables × 5 speakers × 2 repetitions −2 errors).

B. Procedure

1. English /ɹ/-/l/ identification

On each trial, the participants heard one stimulus word and clicked on a button to indicate whether it began with /ɹ/ or /l/. They heard each stimulus once and did not receive feedback. The experiment began with a practice block of 10 trials, and was followed by 120 experimental trials that were blocked by talker but were otherwise in a random order.

2. Bilingual /ɹ/-/l/-/ɾ/ identification

On each trial, the participants heard a syllable and clicked on a button to indicate whether it began with English /ɹ/, English /l/, or Japanese /ɾ/. These options were spelled out on the screen as "R," "L," and "JR;" the categories were described to subjects to ensure that they understood which categories were indicated. They heard each stimulus once and did not receive feedback. The experiment began with a practice block of 30 trials, and was followed by 148 trials that were blocked by talker but were otherwise in a random order.

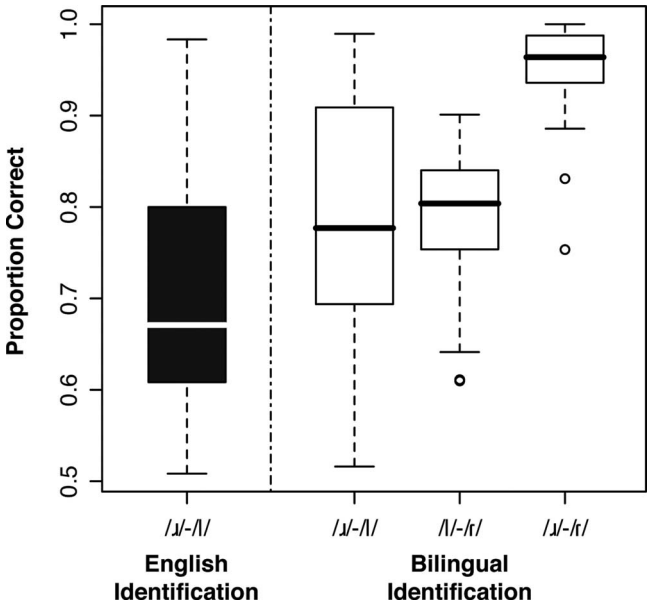


FIG. 1. Boxplots of English /ɹ/-/l/ identification and bilingual /ɹ/-/l/-/ɾ/ identification. The bilingual identification results are presented as pairwise confusions for /ɹ/-/l/, /ɾ/-/l/, and /ɾ/-/ɹ/, with a higher proportion correct indicating less category assimilation. Boxplots display the quartile ranges of scores, with outliers marked by circles.

C. Results

Figure 1 displays the percent-correct English /ɹ/-/l/ identification. As intended, there was a wide range of scores for individual speakers, ranging from near chance to near 100% correct. On average, /ɹ/ was identified correctly on 71% of trials and /l/ was identified correctly on 67% of trials, demonstrating that listeners did not have a strong bias to identify these stimuli as one phoneme or the other.

Table I displays the mean confusion matrix for the bilingual identification task. Japanese listeners demonstrated confusion between English /ɹ/ and /l/; they misidentified /ɹ/ as /l/ on 16% of the trials and /l/ as /ɹ/ on 22% of the trials. They also demonstrated category assimilation between /l/ and Japanese /ɾ/; they identified /l/ as Japanese /ɾ/ on 19% of the trials and Japanese /ɾ/ as /l/ on 17% of the trials. However, they rarely made errors between /ɹ/ and Japanese /ɾ/; they identified /ɹ/ as Japanese /ɾ/ on 2% of the trials and Japanese /ɾ/ as /ɹ/ on 6% of the trials.

In order to examine individual differences in these confusions, we calculated the similarity between individual phoneme pairs (i.e., /ɹ/-/l/, /l/-/ɾ/, and /ɹ/-/ɾ/) from these 3 × 3 confusion data. The calculation of pairwise similarity from larger matrices has been a common issue when analyzing consonant confusion data using multidimensional scaling (e.g., Shepard, 1972; Walden *et al.*, 1980) and hierarchical

TABLE I. Confusion matrix for bilingual identification.

	Response		
Stimulus	/ɹ/	/l/	/ɾ/
/ɹ/	82%	16%	2%
/l/	22%	58%	19%
/ɾ/	6%	17%	77%

cluster analysis (e.g., Iverson *et al.*, 1998) or when applying models of audiovisual integration (e.g., Braida, 1991; Iverson, 2002). Researchers have mostly used straightforward averages of the confusion data to measure similarity [e.g., the distance between /ɹ/ and /l/ can be measured by summing the number of trials in which an /ɹ/ stimulus was identified as /l/ or an /l/ stimulus was identified as /ɹ/; e.g., see Shepard (1972) and Walden *et al.* (1980)]. Alternatively, more theoretical decision models, based on signal detection theory (e.g., Braida, 1991) or Luce's choice theory (e.g., Iverson, 2002), can be applied to correct for response bias. We tried both approaches for the present data and they produced results that were substantively the same. The data reported here are from the more straightforward measure. Specifically, a pairwise percent-correct measure was calculated by summing the correct responses for a pair (e.g., /ɹ/ identified as /ɹ/, and /l/ identified as /l/) and dividing by the total number of correct and incorrect responses for that pair (e.g., /ɹ/ identified as /l/ and /l/ identified as /ɹ/), omitting all trials in which the stimulus or response was from the other category (e.g., /ɾ/).

As displayed in Fig. 1, Japanese speakers demonstrated substantial variability in /ɹ/-/l/ identification of bilingual stimuli, as was found for English /ɹ/-/l/ identification. Identification accuracy was significantly higher for the bilingual than the English stimuli, $t(35)=-7.37$, $p<0.01$, but there was a high correlation between the individual differences for both measures, $r=0.84$, $p<0.01$. Thus, the bilingual stimulus materials seemed to capture most of the individual differences that were measured in the monolingual English task, although the stimuli may have been somewhat easier to identify, at least when those that sounded like /ɾ/ had been excluded.

Individuals confused /l/-/ɾ/ at a rate that was not significantly different than their bilingual /ɹ/-/l/ confusions, $t(35)=-0.03$, $p>0.05$, confirming that English /l/ is perceived as similar to the Japanese /ɾ/. However, there was no significant correlation between individual differences in English /ɹ/-/l/ identification and the degree of /l/-/ɾ/ assimilation, $r=0.22$, $p>0.05$, suggesting that /l/-/ɾ/ category assimilation is not predictive of English /ɹ/-/l/ identification. An inspection of the scatterplot for these two measures in Fig. 2 confirms this view; individuals who were relatively poor at recognizing English /ɹ/ and /l/ had a range of /l/-/ɾ/ category assimilation scores, as did individuals who were better at English /ɹ/-/l/ identification. It is thus questionable whether the degree of assimilation between /l/-/ɾ/ is the cause of difficulties in English /ɹ/-/l/ identification, despite the fact that Japanese adults evidently perceive /l/ and /ɾ/ to be very similar.

Japanese speakers infrequently assimilated English /ɹ/ to Japanese /ɾ/, doing so at a significantly lower rate than for /ɾ/-/l/, $t(35)=-9.25$, $p<0.01$. However, the degree of assimilation was significantly correlated with English /ɹ/-/l/ identification, $r=0.44$, $p<0.01$, suggesting that Japanese listeners whose English /ɹ/ category was more similar to the /ɾ/ category had more problems with English /ɹ/-/l/ identification. Although this may seem to indicate that /ɹ/-/ɾ/ assimilation is the cause of /ɹ/-/l/ identification difficulties, an inspection of the scatterplot in Fig. 2 weakens this view. That is, individuals who were poor at identifying English /ɹ/ and /l/ had a

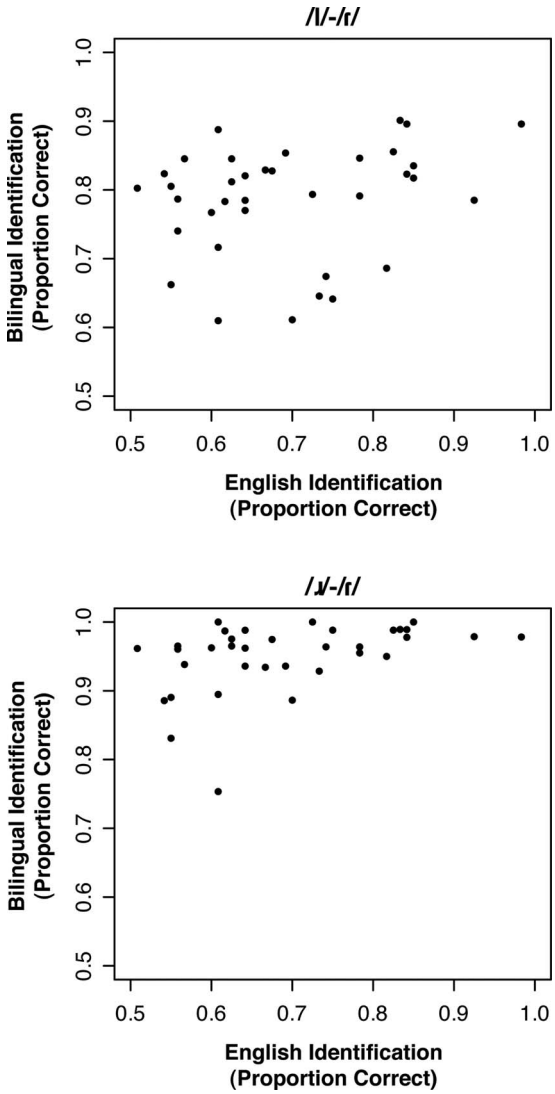


FIG. 2. Scatterplots of the relationships between English /ɾ/-/l/ identification accuracy with the identification of bilingual stimuli for /l/-/ɾ/ and /ɹ/-/ɾ/. Higher bilingual identification scores indicate less confusion between the categories, and a lower degree of category assimilation.

range of /ɹ/-/ɾ/ scores, with many of these individuals rarely confusing /ɹ/ and /ɾ/; it is hard to argue that /ɹ/-/ɾ/ assimilations cause their English /ɹ/-/l/ problems given that these assimilations were not made very frequently, and the more strongly assimilated pair, /l/-/ɾ/, was not correlated with identification accuracy. The frequency of /ɹ/-/ɾ/ assimilation may indicate instead that there is a problem with the underlying /ɹ/ representation, without this problem actually being caused by the proximity of /ɾ/.

Given that English /l/ is similar to both English /ɹ/ and Japanese /ɾ/, it may seem paradoxical that /ɹ/ and /ɾ/ are relatively dissimilar. Part of the answer may be that Japanese listeners use temporal cues to distinguish these consonants (e.g., transition duration) more than they use spectral cues (e.g., F3) (Iverson *et al.*, 2005; Yamada, 1995), and these temporal cues have natural variation, with /ɾ/ having short transitions, /ɹ/ having longer transitions, and /l/ being in between (e.g., see Experiment 3 results). That is, /ɹ/ and /l/ may sometimes be confused because some of these phonemes have overlapping transition durations (e.g., some /ɹ/ have

short enough transitions to sound like /l/), as do /l/ and /ɾ/ (e.g., some /ɾ/ have long enough transitions to sound like /l/), but there may still be little overlap between /ɹ/ and /ɾ/ (i.e., few /ɹ/ have short enough transitions to sound like /ɾ/, and few /ɾ/ have long enough transitions to sound like /ɹ/). In other words, it is possible that /ɹ/ and /ɾ/ are dissimilar even though both are close to /l/, because of the way in which their acoustic distributions overlap; the result is only paradoxical if one conceives of these phonemes simplistically, as single points in a geometric space.

III. EXPERIMENT 2: RATINGS OF SPOKEN ACCENT

Experiment 1 demonstrated that the relationships between category assimilation and English /ɹ/-/l/ identification were weak. However, it is possible that category assimilation could have a greater role in affecting /ɹ/-/l/ production; much of the data supporting SLM has come from production data (e.g., Flege, 1987, 1995, 2003; Bohn and Flege, 1992), and it has often been found that production and perception accuracy for L2 phonemes are only weakly correlated (e.g., Bradlow *et al.*, 1997). We thus decided to include a short test of the subjects' production ability; subjects read an accent-revealing sentence (*The red robin looked across the lovely lake*) and phonetically trained listeners rated the accent of the /ɹ/ and /l/ consonants, as well as the degree of contrast that the talker made between /ɹ/ and /l/.

A. Method

1. Subjects

Same as Experiment 1.

2. Apparatus

Same as Experiment 1.

3. Procedure

Each subject was recorded reading the sentence "The red robin looked across the lovely lake." After all recordings had been made, three phonetically trained British English speakers made three ratings for each recording: degree of /ɹ/ accent, degree of /l/ accent, and the degree of contrast that the speaker made between /ɹ/ and /l/. We have used the contrast rating in previous work (Iverson *et al.*, 2008), and it is useful in cases where speakers do not produce phonemes in a nativelike way, but nonetheless produce a phonetic contrast that could make their speech intelligible to a listener. The ratings were made on an integer scale from 1 (*poor*) to 7 (*good*). The raters were allowed to repeatedly listen to the stimuli before giving their judgment. The rating scores were averaged across the raters for each recording to provide a measure of the degree of accent.

B. Results

As displayed in Fig. 3, the accents of the Japanese speakers varied across the entire range in terms of the degree of contrast and the /ɹ/ accent. They were, as a group, more consistently able to produce /l/ in a way that was closer to the "good" end of the scale.

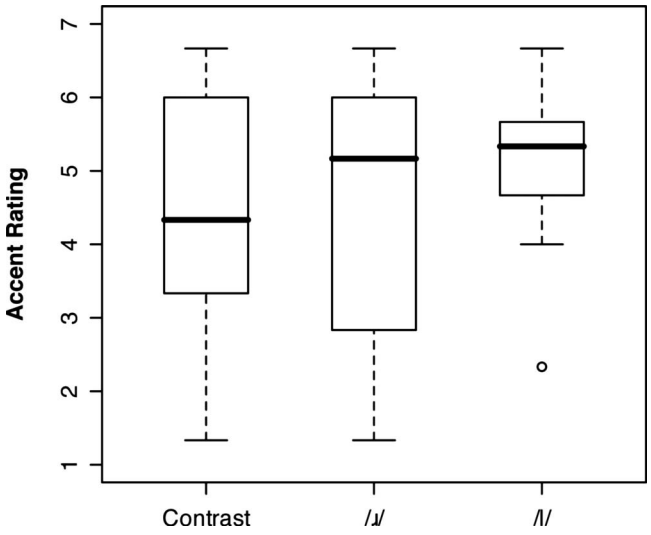


FIG. 3. Boxplots of the rated accent for Japanese speakers saying "The red robin looked across the lovely lake." Three ratings were made for each sentence: the degree of /ɹ/-/l/ contrast, /ɹ/-accent, and /l/-accent. Ratings were made on a scale from 1 (poor, indicating a strong non-native accent or poor contrast) to 7 (good, indicating little non-native accent or high contrast).

The baseline English /ɹ/-/l/ identification accuracy scores of Experiment 1 were significantly correlated with the accent ratings for /ɹ/ accent, $r=0.53$, $p<0.01$, and /ɹ/-/l/ contrast, $r=0.42$, $p<0.01$. However, there was no significant correlation between /l/ accent and English /ɹ/-/l/ identification, $r=0.20$, $p>0.05$. This pattern of results is in accord with previous findings that perceptual identification is moderately related to the degree of spoken accent (e.g., Bradlow *et al.*, 1997).

The degree of assimilation for /l/-/ɾ/ (Experiment 1) was not significantly correlated, $p>0.05$, with /ɹ/ accent, $r=0.13$, /l/ accent, $r=0.19$, or /ɹ/-/l/ contrast, $r=0.08$. Likewise, the degree of assimilation for /ɹ/-/ɾ/ was not significantly correlated, $p>0.05$, with /ɹ/ accent, $r=0.29$, /l/ accent, $r=0.17$, or /ɹ/-/l/ contrast, $r=0.23$. Thus, similar to Experiment 1, there is little evidence that the degree of perceptual assimilation between English /ɹ/ and /l/ with Japanese /ɾ/ has a substantial role in causing the individual differences in /ɹ/-/l/ production abilities.

IV. EXPERIMENT 3: PERCEPTUAL MAPPING OF BEST EXEMPLARS

This experiment measured the cues that the Japanese individuals use to represent /ɹ/, /l/, and /ɾ/, in order to better understand the assimilation process, as well as attempt to measure assimilation in a way that better predicts the ability of individuals to identify English /ɹ/ and /l/. The experiment adapted the goodness optimization method that we have used recently to map best-exemplar locations for vowels in high-dimensional, phonetically detailed, stimulus spaces (Evans and Iverson, 2004, 2007; Iverson and Evans, 2003, 2007; Iverson *et al.*, 2006). On each trial, subjects see a target syllable on the computer screen, hear a synthesized stimulus embedded in a natural carrier sentence, and rate on a continuous scale how close the stimulus that they hear is to

being a good exemplar of the syllable on the screen. The algorithm then adjusts the stimulus that they hear on successive trials to iteratively find a good exemplar of that syllable. In this way, we have been able to find best exemplars in very large sets of stimuli after 35 trials.

This approach has been useful for assessing cross-language differences in vowel spaces (Iverson and Evans, 2007). The best exemplars for 14 English vowels were found for L1 speakers of Spanish, French, German, and Norwegian, and the accuracy with which their best exemplars represented static formant targets and formant movement were highly correlated with their ability to recognize natural vowel stimuli.

The present experiment used a similar method to find best exemplars of /ɹ/, /l/, and /ɾ/ in a five-dimensional stimulus space that included F1, F2, and F3 frequencies during the closure, and the closure and transition duration. Although F3 is the primary cue that native speakers use to distinguish /ɹ/ and /l/, F2 and transition duration are often considered to be secondary cues (e.g., Underbakke *et al.*, 1988); these secondary cues can be particularly important to Japanese adults, who typically have difficulty hearing F3 variation (e.g., Iverson *et al.*, 2003; Underbakke *et al.*, 1988; Yamada, 1995). Closure duration and F1 were included to allow good examples of /ɾ/ to be found; /ɾ/ is often described as being much more rapid than /ɹ/ or /l/ (e.g., Miyawaki *et al.*, 1975), and occasionally has a stoplike quality that could possibly lower F1. We did not include a burst dimension in this space, because our pilot investigations showed that many speakers do not produce /ɾ/ with a burst, and it was possible to synthesize a subjectively good exemplar of /ɾ/ without it.

A. METHOD

1. Participants

The Japanese speakers were the same as in Experiments 1 and 2. We additionally tested 13 native speakers of British English to provide normative data. Their ages ranged between 18 and 62 years (median=24), and all were born and raised in southern England. No listeners reported having hearing problems.

2. Apparatus

The stimuli were played to subjects using Sennheiser HD 280 headphones while subjects sat in a sound-treated room. Responses were entered and recorded using a custom-written program which also controlled the presentation of stimuli.

3. Stimuli

The stimuli were synthetic C-/ɑ/ syllables embedded in naturally spoken English and Japanese carrier sentences (i.e., say [ ] again, and *mata [ ] to itte kudasai*). The same female speaker was used for both carrier sentences, an English/Japanese bilingual who spent her childhood in both the UK and Japan, and is highly fluent in both languages.

The synthetic syllables were modeled based on this natural talker, using a Klatt synthesizer (Klatt and Klatt, 1990). A five-dimensional set of stimuli was created by orthogonally varying F1, F2, F3, closure duration, and transition duration (i.e., from the consonantal articulation to the following vowel). The values were chosen so that they would span an acoustic space that included /ɹ/, /l/, and /ɾ/, as well as match the target speaker's voice. F1 varied from 123 to 603 Hz. F2 was always at least 1 ERB (Glasberg and Moore, 1990) greater than F1, and less than 2489 Hz. F3 was always at least 1 ERB greater than F2, and less than 3951 Hz. The closure and transition durations varied from 5 to 290 ms. The frequency values were quantized in 1-ERB steps and the duration values were quantized with a log spacing (12 steps for each). All other synthesis parameters were the same for all stimuli, and were set so that they closely matched the original F0 contour, amplitude envelope, long-term average spectra, and vowel formants of the original syllables produced by this talker. There were a total of 60 660 stimuli for each carrier sentence.

B. Procedure

On each trial, subjects saw a consonant printed on the computer screen (e.g., R) and heard a stimulus (synthesized CV embedded in a natural carrier sentence). They rated on a continuous scale how far away the CV that they heard was from being a good exemplar of the printed consonant. Their ratings were given by mouse clicking on a continuous bar presented on a computer screen. Japanese subjects found best exemplars for all three consonants and English subjects found best exemplars for only /ɹ/ and /l/. All subjects performed this task in both the Japanese and English carrier sentences, even though the English speakers did not understand Japanese. Both languages were included in case there were highly language-specific best-exemplar locations (e.g., it was possible that Japanese subjects would have trouble finding a good /ɾ/ when listening to English). However, this kind of language variation was not found (i.e., people appeared to choose the same kinds of best exemplars in both sentences), so the results were averaged across carrier sentence for later analysis.

A goodness optimization procedure (Evans and Iverson, 2004, 2007; Iverson and Evans, 2003, 2007; Iverson *et al.*, 2006) was used to iteratively change the stimuli that subjects heard on each trial, to search through the multidimensional stimulus space for better exemplars. This procedure involves searching along individual vectors (i.e., one-dimensional straight-line paths crossing through the five-dimensional stimulus space) and finding the best exemplar on each vector. There were a total of 7 search vectors and 5 trials per vector for each consonant.

Vector 1 covaried all five dimensions along a straight-line path that passed through two points: (1) measurements based on the natural productions of each consonant, and (2) a middle point in the stimulus space (i.e., values averaged for all three stimuli). Vectors 2–6 varied one dimension at a time, in a descending order based on their presumed importance to /ɹ/-/l/ identification (F3, F2, transition duration, closure duration, and F1). Vector 7 varied all dimensions, passing through the best value found thus far on each dimension and the middle point in the space. The endpoints of all vec-

tors were constrained by the boundaries of the stimulus space so that listeners were able to choose stimuli over a wide range.

For each vector, subjects first heard stimuli at the two endpoints (i.e., at the boundaries of the stimulus space), with the order of these two trials randomized. The selection of stimuli on the remaining trials was based on the subjects' judgments, using formulas that were designed to find stimuli along the path that would be perceived as better exemplars. On the third trial, subjects heard a stimulus that was selected by a goodness-weighted average of the first two stimuli. On the fourth and fifth trials, the stimuli were selected by fitting a parabola to the goodness data and finding the minimum of this function. See Iverson and Evans (2003) for exact formulas and procedures. At the completion of the fifth trial, subjects were allowed to repeat the search if it had produced a poor exemplar (i.e., they made an explicit judgment about whether the stimulus was or was not close to sounding good). If the best exemplar was judged to be close, the parameters of the best stimulus found thus far were passed onto the next stage of the search algorithm (i.e., to search along the next vector).

C. Results

Figure 4 displays the best exemplars of English /ɹ/, /l/, and Japanese /ɾ/. The results demonstrate that /ɾ/ is more similar to /l/ than /ɹ/ in terms of F3 and transition duration, in accord with the asymmetric pattern of assimilation found in Experiment 1. However, /ɾ/ was different from both /l/ and /ɹ/ in terms of F2, and was possibly closer to /ɹ/ in terms of closure duration. One-way repeated-measures analyses of variance (ANOVAs) were run separately for each dimension to determine whether these three best exemplars were significantly different for the Japanese listeners. The dependent variables were the five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration), and consonant was coded as a three-level within-subject variable (i.e., English /ɹ/, /l/, and Japanese /ɾ/). There were no significant differences for F1, $p>0.05$. However, the consonants differed significantly in F2, $F(2,70)=28.48$, $p<0.01$, F3, $F(2,70)=129.28$, $p<0.01$, closure duration, $F(2,70)=6.36$, $p<0.01$, and transition duration, $F(2,70)=68.45$, $p<0.01$. Tukey HSD comparisons revealed that /ɹ/ was significantly different from both /l/ and /ɾ/ in terms of F2, F3, and transition duration, $p<0.05$, but not F1 or closure duration; /l/ and /ɾ/ were significantly different in terms of F2, F3, closure duration, and transition duration, $p<0.05$.

The results thus demonstrate that Japanese adults were able to have three distinct representations for these categories (e.g., not assimilating /l/ and /ɾ/), despite the fact that many of these listeners had difficulty recognizing natural /ɹ/ and /l/ stimuli. However, it is clear that there were substantial individual differences in their best exemplars. For example, the closure durations of best exemplars for English /ɹ/ and /l/ varied widely and overlapped with the range of Japanese /ɾ/. In order to examine whether their degree of separation between L1 and L2 categories is predictive of their English /ɹ/-/l/ identification accuracy, the absolute values for their difference for /ɹ/-/ɾ/ and /l/-/ɾ/ were computed for the five acoustic dimensions. On all acoustic measures, there were no significant correlations between /l/-/ɾ/ difference and English /ɹ/-/l/ identification (F1, $r=-0.15$; F2, $r=-0.17$; F3, $r=0.10$; closure duration, $r=0.08$; transition duration, $r=-0.09$, $p>0.05$), and no significant correlations between the /ɹ/-/ɾ/ distance and English /ɹ/-/l/ identification (F1, $r=-0.11$; F2, $r=-0.03$; F3, $r=0.25$; closure duration, $r=-0.30$; transition duration, $r=0.12$, $p>0.05$). The results are thus in accord with the findings of Experiment 1; there is little evidence that the similarity of a listener's L1 and L2 categories causes individual differences in /ɹ/-/l/ identification accuracy.

The best exemplars chosen by Japanese speakers for /ɹ/ and /l/ were surprisingly similar to those of English speakers, considering their levels of /ɹ/-/l/ identification accuracy. ANOVA analyses tested whether these best exemplars varied with language group for each acoustic dimension. The acoustic values were entered in separate analyses as dependent measures, consonant (/ɹ/ or /l/) was a within-subject factor, and language group (Japanese or English) was a between-subject factor. There was a significant interaction between consonant and language group for F3, $F(1,94)=5.03$, $p<0.05$; English speakers systematically chose higher F3 frequencies for /l/ than did Japanese speakers, but their values were similar for /ɹ/. This indicates that Japanese adults did not have completely nativelike representations for F3. There was no significant interaction between language and consonant for the other acoustic dimensions, $p>0.05$, and there were no main effects of language. Unsurprisingly, there were main effects of consonant for F3, $F(1,94)=351.85$, $p<0.01$, closure duration, $F(1,94)=12.73$, $p<0.01$, and transition duration, $F(1,94)=97.83$, $p<0.01$, demonstrating that English /ɹ/ and /l/ differ on these dimensions. There were no significant main effects of consonant for F1 and F2, $p>0.05$.

Pearson correlations were used to determine whether the accuracy with which Japanese listeners represent /ɹ/ and /l/ (i.e., their similarity to the average best exemplars chosen by native English speakers) was predictive of identification accuracy. The accuracy of the representations was measured for each of the five acoustic dimensions, combining /ɹ/ and /l/ using a Euclidean metric. For example, we calculated how far each individual's /ɹ/ and /l/ best exemplars were from the English averages on the F3 dimension, and then combined these two values by calculating the square root of the sum of squares. There was a significant correlation between F3 accuracy and /ɹ/-/l/ identification, $r=-0.46$, $p<0.01$; Japanese individuals who had F3 representations that were similar to those of native English speakers were better at English /ɹ/-/l/ identification. However, there were no significant correlations between the other four dimensions and /ɹ/-/l/ identification, $p>0.05$.

Figure 5 displays the scatterplot of F3 accuracy versus /ɹ/-/l/ identification. The significant correlation occurred because individuals who were accurate in their /ɹ/-/l/ identification tended to have nativelike F3 values, whereas less accurate subjects were more variable in their F3 values, with some of these individuals having nativelike F3 values and

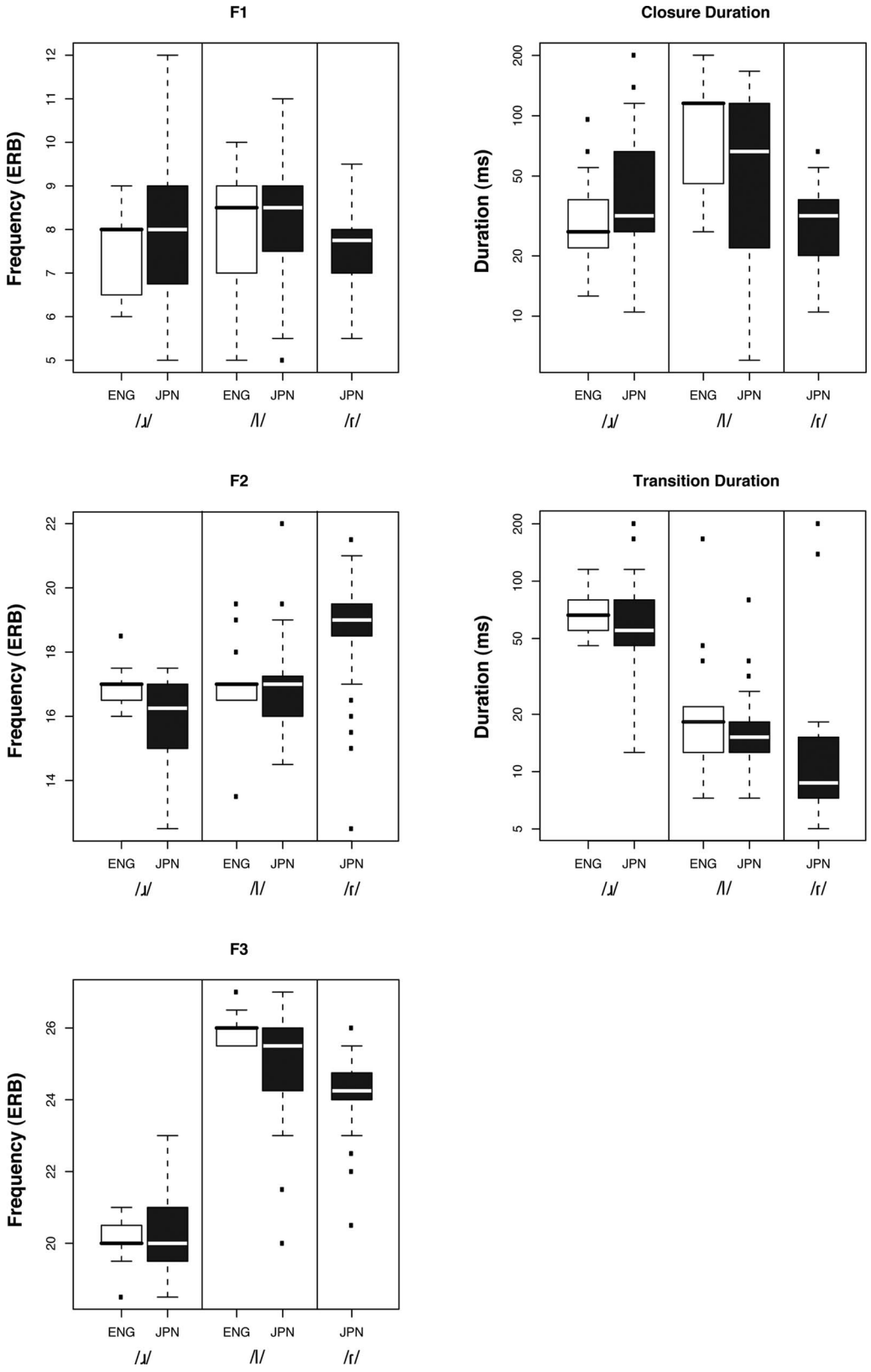


FIG. 4. Boxplots of the acoustic parameters of the best exemplars for /ɹ/, /l/, and /ɾ/ by L1 speakers of English (ENG) and Japanese (JPN).

others having ones that were far away. This may have occurred because the less accurate subjects either had an unclear representation of F3 or had a perceptual difficulty hearing variation along this dimension (e.g., Iverson *et al.*, 2003). That is, less accurate subjects may have simply been more random about which F3 values they said they preferred, rather than systematically preferring values that were far from those of native listeners.

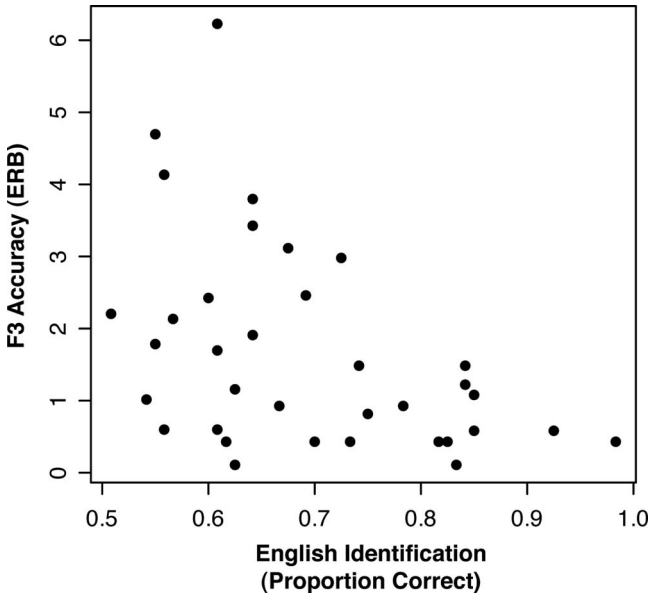


FIG. 5. Scatterplot of the relationship between the accuracy of the English /ɹ/ and /l/ best exemplars for Japanese listeners along the F3 dimension (i.e., distance from English averages) and their English identification accuracy.

V. GENERAL DISCUSSION

The present results confirm that English /l/ is assimilated into the Japanese /ɾ/ category, at least to some degree. That is, listeners confused /l/ and /ɹ/ in a forced-choice task as much as they confused /ɹ/ and /l/, but their best exemplars for /l/ were not identical to those for /ɹ/, even though they were broadly similar on several dimensions (e.g., F3 and transition duration). The assimilation of /ɹ/ and /ɾ/ was comparatively weak; individuals confused /ɹ/ and /ɾ/ infrequently in a forced-choice task and chose best exemplars for /ɹ/ that were different from /ɾ/ in terms of F3 and transition duration. The results thus suggest that /ɹ/ and /l/ differentially assimilate into the Japanese /ɾ/ category, with /l/ being closer, supporting the claims of Aoyama *et al.* (2004). It is possible that previous work using subjective ratings of category goodness (e.g., Guion *et al.*, 2000; Iverson *et al.*, 2003) did not find a clearer asymmetry because it is hard for individuals to rate differences in category goodness when both tokens are poor exemplars overall.

Within the classification scheme of PAM, this appears to be a Category Goodness (CG) difference or an Uncategorized versus Categorized type (UC) (Best, 1995; Best *et al.*, 2001), not the single category assimilation type hypothesized by Best and Strange (1992). The problem for PAM is that CG and UC contrasts are predicted to be relatively easy to discern; in both cases the two phonemes should be perceived as being different because they differentially assimilate into the L1 phonological system. However, the English /ɹ/-/l/ contrast perceived by Japanese listeners is one of the most difficult contrasts reported in the literature (e.g., Goto, 1971; Miyawaki *et al.*, 1975). In combination with previous work, demonstrating that listeners can hear acoustic variation in /ɹ/ and /l/ that does not affect perceived goodness (Iverson *et al.*, 2003), it thus seems clear that PAM does not explain why Japanese listeners have difficulty with English /ɹ/-/l/.

The results of this study are also problematic for SLM. Previous work supporting SLM has suggested that language learners merge (Flege, 2003; MacKay *et al.*, 2001) or dissimilate (Flege, 1995; Flege *et al.*, 2003) L1 and L2 phonetic categories that are close enough to interact, so we expected that Japanese learners who were better at identifying /ɹ/ and /l/ would have assimilation patterns that would be distinct from those who identified /ɹ/ and /l/ less accurately. We found little evidence that such individual differences in assimilation were related to identification performance. There was a significant correlation between /ɹ/-/ɾ/ assimilation and identification performance in Experiment 1, so the possibility that this assimilation makes it hard to learn to recognize these phonemes cannot be completely discarded. However, it seems unlikely that /ɹ/-/ɾ/ assimilation is the actual cause of these learning problems because very few listeners made this confusion frequently. Moreover, the strength of /l/-/ɾ/ assimilation was stronger than /ɹ/-/ɾ/, so SLM predicts that /l/-/ɾ/ assimilation should have an even bigger effect on category learning (Aoyama *et al.*, 2004). Instead, we found no reliable correlation of /l/-/ɾ/ assimilation with the individual differences in identification performance.

In addition, the results of Experiment 3 suggest that listeners were able to maintain three separate categories for /ɹ/, /l/, and /ɾ/, despite the fact that the subject group as a whole was not particularly skilled at identifying /ɹ/ and /l/. For example, their best exemplars demonstrated that they knew that /l/ has a longer initial closure and a lower F2 than /ɾ/; the closeness of /l/ and /ɾ/ did not cause them to assimilate these phonemes into a single category. The only evidence for problems caused by /l/-/ɾ/ similarity was the fact that the best exemplars for /l/ by Japanese speakers were less nativelike along the F3 dimension, compared to their best exemplars for /ɹ/. This fits SLM's predictions (Aoyama *et al.*, 2004) that assimilation causes more problems for learning /l/ than /ɹ/. However, it is unexplained why this should affect only the F3 dimension; the best exemplars of /l/ for Japanese listeners were not significantly different from those of native speakers in any other respect.

If assimilation does not explain individual differences in identification performance, what does? The only positive conclusion from the present study is that the representation of F3 is important; individuals who had more nativelike best exemplars along the F3 dimension were more accurate at identifying /ɹ/ and /l/. F3 has been known to be problematic for Japanese speakers for a long time (e.g., Miyawaki *et al.*, 1975), but this is perhaps the first study to demonstrate that such results on synthetic speech relate to how individuals identify real speech, produced by multiple talkers in multiple word contexts. It is worth noting that this correlation was not particularly high, $r=-0.46$, but this may have occurred because individuals with poor identification ability were fairly random in the F3 values that they preferred.

It is plausible too that this correlation was low because there are other factors that affect individual differences in the ability to identify /ɹ/ and /l/. For example, previous work (Iverson *et al.*, 2003) has suggested that central auditory processing, at a stage prior to phonemic categorization, affects the ability of Japanese listeners to identify /ɹ/ and /l/; Japa-

nese adults are much more sensitive to acoustic variation that is irrelevant to /ɹ/-/l/ categorization (e.g., F2 variation, and within-category F3 variation) than they are to the critical F3 differences that distinguish these categories for native speakers. Such sensitivities could interfere with category learning, making it harder for them to learn which F3 values are correct for each phoneme, as found in the best exemplars measured in Experiment 3. In addition, this pattern of auditory processing could interfere with online speech perception even if listeners manage to form nativelike category representations; it could make it harder for them to focus on the acoustic variation that is important even though, at some level, they may represent that /ɹ/ has a lower F3 than /l/.

The present best-exemplar assessment of English and Japanese categories for /ɹ/ and /l may seem at odds with previous work on category boundaries. That is, previous work (e.g., Iverson *et al.*, 2005; Yamada, 1995) has shown that Japanese and English speakers use very different cues for /ɹ/-/l/ identification (F3 for English speakers; F2, transition, and closure duration for Japanese speakers), but the present results suggest that their best exemplars are broadly similar. Part of this difference may occur because the best exemplars better demonstrate sensitivity to secondary cues. For example, English speakers in the present study showed that they knew that /ɹ/ and /l/ differ in transition and closure duration, even though English identification decisions are mostly driven by F3. Likewise, Japanese listeners appear to have broadly nativelike representations for secondary cues in their best exemplars, but they may rely on them more for identification because their perception of F3 is relatively poor.

To summarize, the results of this study suggest that Japanese adults have an asymmetric pattern of category assimilation for the /ɹ/-/l/ contrast, with /l/ assimilating stronger into the /ɾ/ category than does /ɹ/, but the degree to which individuals assimilate /ɹ/ and /l/ into their L1 categories is not strongly predictive of their ability to identify these phonemes accurately. Rather, their ability to identify /ɹ/ and /l/ is more directly related to their representation of F3 for these phonemes, which may be caused by their auditory/phonetic sensitivities along this dimension.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (**2004**). "Perceived phonetic distance and L2 learning: The case of Japanese /r/ and English /l/ and /r/," J. Phonetics **32**, 233–250.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Language Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 171–204.

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology," J. Acoust. Soc. Am. **1097**, 775–794.

Best, C. T., and Strange, W. (**1992**). "Effects of phonological and phonetic factors on cross-language perception of approximants," J. Phonetics **20**, 305–330.

Boersma, P., and Weenik, D. (**2008**). "Praat: Doing phonetic by computer," retrieved from http://www.praat.org (Last viewed September, 2008).

Bohn, O.-S., and Flege, J. E. (**1992**). "The production of new and similar vowels by adult German learners of English,"Stud. Second Lang. Acquis. **14**, 131–158.

Bradlow, A. R., and Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Braida, L. D. (**1991**). "Crossmodal integration in the identification of consonant segments," Q. J. Exp. Psychol. A **43**, 647–677.

Evans, B. G., and Iverson, P. (**2004**). "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," J. Acoust. Soc. Am. **115**, 352–361.

Evans, B. G., and Iverson, P. (**2007**). "Plasticity in vowel perception and production: A study of accent change in young adults," J. Acoust. Soc. Am. **121**, 3814–3826.

Flege, J. E. (**1987**). "The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification," J. Phonetics **15**, 47–65.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Language Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 233–277.

Flege, J. E. (**2003**). "Assessing constraints on second-language segmental production and perception," *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. Schiller (Mouton de Gruyter, Berlin), pp. 319–355.

Flege, J., Schirru, C., and MacKay, I. (**2003**). "Interaction between the native and second language phonetic subsystems," Speech Commun. **40**, 467–491.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Goto, H. (**1971**). "Auditory perception by normal Japanese adults of the sounds L and 'R'," Neuropsychologia **9**, 317–323.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (**2000**). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," J. Acoust. Soc. Am. **107**, 2711–2724.

Hazan, V., Sennema, A., Iba, M., and Faulkner A. (**2005**). "Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English," Speech Commun. **47**, 360–378.

Iverson, P. (**2002**). "A choice theory method of evaluating audiovisual integration," Speech, Hearing, and Language: Work in Progress **14**, 85–92.

Iverson, P., Bernstein, L. E., and Auer, E. T., Jr. (**1998**). "Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition," Speech Commun. **26**, 45–63.

Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., and Evans, B. G. (**2008**). "Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers,"J. Exp. Psychol. Hum. Percept. Perform. **34**, 1305–1316.

Iverson, P., and Evans, B. G. (**2003**). "A goodness optimization method for investigating phonetic categorization," Proceedings of the 15th International Conference of Phonetic Sciences, Barcelona.

Iverson, P., and Evans, B. G. (**2007**). "Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration," J. Acoust. Soc. Am. **122**, 2842–2854.

Iverson, P., Hazan, V., and Bannister, K. (**2005**). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults," J. Acoust. Soc. Am. **118**, 3267–3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.

Iverson, P., Smith, C. A., and Evans, B. G. (**2006**). "Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration," J. Acoust. Soc. Am. **120**, 3998–4006.

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.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–886.

Lotto, A. J., Sato, M., and Diehl, R. L. (**2004**). "Mapping the task for the second language learner: The case of Japanese acquisition of // and /l/," in *From Sounds to Sense:* 50+Years *of Discoveries in Speech Communication*, edited by J. Slifka, S. Manuel, and M. Matthies (MIT Research Laboratory in Electronics, Cambridge, MA), pp. C-181–C186.

MacKay, I. R. A., Flege, J. E., Piske, T., and Schirru, C. (**2001**). "Category

restructuring during second-language (L2) speech acquisition," J. Acoust. Soc. Am. **110**, 516–528.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., and Fujimura, O. (**1975**). "An effect of language experience:The discrimination of /r/ and /l/ by native speakers of Japanese and English," Percept. Psychophys. **18**, 331–340.

Shepard, R. N. (**1972**). "Psychological representation of speech sounds," in *Human Communication: A Unified View*, edited by E. E. David and P. B. Denes (McGraw-Hill, New York), pp. 67–133.

Underbakke, M., Polka, L., Gottfried, T. L., and Strange, W. (**1988**). "Trading relations in the perception of /r/-/l/ by Japanese learners of English," J. Acoust. Soc. Am. **84**, 90–100.

Vance, T. (**1987**). *An Introduction to Japanese Phonology* (State University of New York Press, Albany, NY).

Walden, B. E., Montgomery, A. A., Prosek, R. A., and Schwartz, D. M. (**1980**). "Consonant similarity judgments by normal and hearing-impaired listeners," J. Speech Hear. Res. **23**, 162–184.

Yamada, R. A. (**1995**). "Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native takers of Japanese," in *Speech Perception and Language Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 305–320.

Yamada, R., and Tohkura, Y. (**1992**). "The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners," Percept. Psychophys. **52**, 376–392.

Automatic source speaker selection for voice conversion

Oytun Turk[a] and Levent M. Arslan[b]
Electrical and Electronics Engineering Department, Bogazici University, Bebek, Istanbul, 34342 Turkey and Sestek Inc., R&D Department, ITU Ayazaga Kampusu, ARI-1 Teknopark Binasi, Maslak, Istanbul 34342 Turkey

(Received 1 August 2007; revised 24 October 2008; accepted 24 October 2008)

This paper focuses on the importance of source speaker selection for a weighted codebook mapping based voice conversion algorithm. First, the dependency on source speakers is evaluated in a subjective listening test using 180 different source-target pairs from a database of 20 speakers. Subjective scores for similarity to target speaker's voice and quality are obtained. Statistical analysis of scores confirms the dependence of performance on source speakers for both male-to-male and female-to-female transformations. A source speaker selection algorithm is devised given a target speaker and a set of source speaker candidates. For this purpose, an artificial neural network (ANN) is trained that learns the regression between a set of acoustical distance measures and the subjective scores. The estimated scores are used in source speaker ranking. The average cross-correlation coefficient between rankings obtained from median subjective scores and rankings estimated by the algorithm is 0.84 for similarity and 0.78 for quality in male-to-male transformations. The results for female-to-female transformations were less reliable with a cross-correlation value of 0.58 for both similarity and quality. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3027445]

PACS number(s): 43.72.Ar, 43.72.Ja [DOS] Pages: 480–491

I. INTRODUCTION

The aim of voice conversion is to transform a source speaker's voice into a target speaker's voice automatically. As far as speaker variability is concerned, voice conversion systems may face two problems: variation in performance on source speakers for a given target speaker and variation in performance on different source-target pairs. Performance assessments and detailed analyses can be carried out in both cases by employing a database that consists of a large number of source-target pairs. This would practically be an impossible task considering all possible speaker variations. However, similar assessments can be performed on a more limited and controlled database focusing on a subset of the first problem: assessment of the performance dependency on a limited number of source speakers for a given target speaker's voice. It would then be interesting to develop a method for automatic selection of the best source speaker alternative.

In practical voice conversion applications, the target speaker is fixed most of the time. Conversion can be performed off-line which enables the search for an appropriate source speaker. As an example, consider a movie dubbing application that involves the transformation of an ordinary voice to a celebrity's voice. In this scenario, the source speaker can be selected from a set of candidates by collecting databases from different source speakers, performing voice conversion from each source speaker's voice to the target speaker's voice, and comparing the outputs. It is desirable to perform subjective comparisons since objective criteria may not always reflect the performance as perceived by human listeners. However, designing and carrying out a subjective listening test for each target voice will be time-consuming and expensive. Therefore, selection of the most suitable candidate using an automatic algorithm will be helpful. It will even be possible to use such a technique to search for the most appropriate source speaker from a portfolio that contains a large number of prerecorded source speaker candidates and to call back only the most promising ones for a given voice conversion task.

Voice conversion has been a popular topic in speech processing research (Abe *et al.*, 1988; Arslan, 1999; Moulines and Sagisaka, 1995; Stylianou *et al.*, 1998). There are two common stages in voice conversion algorithms: training and transformation. In the training stage, the acoustical mapping function between the source and the target speaker's acoustical spaces is estimated using machine learning techniques including vector clustering/quantization (Abe *et al.*, 1988), codebook mapping (Acero, 1993), sentence hidden Markov models (HMMs) and weighted codebook mapping (Arslan, 1999), Gaussian mixture models (GMMs) (Stylianou *et al.*, 1998), radial basis function networks (RBFNs) (Drioli, 1999), artificial neural networks (ANNs) (Narendranath *et al.*, 1995), and self-organizing maps (Knohl and Rinscheid, 1993). The mapping is then used in the transformation stage to predict the target acoustic features corresponding to a given set of source features. Necessary modifications are performed on the source waveform to match the characteristics of the target speaker's voice including the transformation of the vocal tract, glottal source, duration, and energy characteristics. The vocal tract characteristics can be transformed using formant modification (Mizuno and Abe, 1995), interpolation of the line spectral frequencies (LSFs) (Arslan, 1999), and sinusoidal modeling techniques (Laroche *et al.*, 1993). There exist several methods for time scaling

[a]Electronic mail: oytun.turk@sestek.com.tr. Author to correspondence should be addressed. Present address: The Speech Group in DFKI Language Technology Lab, Berlin, Germany.
[b]Electronic mail: arslanle@boun.edu.tr.

and pitch scaling based prosody modifications: time-domain pitch-synchronous overlap-add algorithm (Moulines and Charpentier, 1990), frequency-domain pitch-synchronous overlap-add algorithm (FD-PSOLA) (Moulines and Verhelst, 1999), sinusoidal synthesis (Quatieri and McAulay, 1992), phase vocoding (Flanagan and Golden, 1966), and mixed excitation techniques (Acero, 1998).

Voice conversion provides a useful framework to analyze, model, and transform perceived characteristics of speech. It has applications in text-to-speech synthesis for adapting the synthesizer to new voices (Kain and Macon, 1998; Zhang *et al.*, 2001), voice quality transformation (Turk and Arslan, 2005), transformation of expressive style (Inanoglu and Young, 2007), enhancement of abnormal speech (Bi and Qi, 1997), and speaker adaptation for speech recognition (Shikano *et al.*, 1991). Applications in international movie dubbing (Turk, 2003) and music (Turajlic *et al.*, 2003) have also emerged in recent years.

In this paper, we first examine the performance dependence of a weighted codebook mapping based voice conversion algorithm on source-target pair combinations in a subjective framework. The algorithm was originally proposed by Arslan (1999) and refined by Turk and Arslan (2006). We show that the performance as measured by subjective similarity to target voice and output quality is dependent on the source speakers given the target speaker. Then, we propose an automatic algorithm for source speaker selection among a set of candidates for a given target speaker. This algorithm tests voices of different source speakers using objective criteria to find out the most appropriate candidate for a given target speaker.

In summary, this study focuses on two hypotheses.

- The performance of weighted codebook mapping conversion is dependent on source speakers for a given target speaker.
- Given a target speaker and a set of source speaker candidates, the best source speaker that would result in closest subjective match can be automatically selected using a set of objective distance measures extracted from a small amount of recordings.

This paper extends the results obtained in a previous study by the authors (Turk and Arslan, 2005) where an ANN was trained to obtain estimates of subjective scores from a set of acoustical distance measures among pairs of source and target speakers. In this paper, we present the acoustic features and the distance measures employed in more detail. Statistical analysis is performed for investigating the dependence of subjective scores on source speakers for a given target voice. The performance of the ANN is evaluated in ranking estimation in addition to subjective score estimation tasks.

There exist well-established methods for the analysis and modeling of variability of various acoustic features among speakers in literature. As an example, the vocal tract structure can be modeled using formant estimation (Holmes *et al.*, 1990), linear prediction (LP) analysis (Makhoul, 1975), or sinusoidal transform coding (McAulay and Quatieri, 1995). Pitch is one of the most important speaker-specific dimensions of prosody. It can be estimated using the autocorrelation function, average magnitude difference function, Fourier transform, and harmonic analysis (Rabiner and Schafer, 1978). There has been considerable amount of work on the analysis, modeling, and modification of glottal source characteristics in voice quality research (Childers and Lee, 1991; Childers, 1995; Fant *et al.*, 1985). Spectral tilt (Maddieson and Ladefoged, 1985), H1-H2 (Hanson, 1995), jitter (Kreiman and Gerrat, 2003), shimmer (Kreiman and Gerrat, 2003), and soft phonation index (Bhuta *et al.*, 2004) are other important parameters of voice quality.

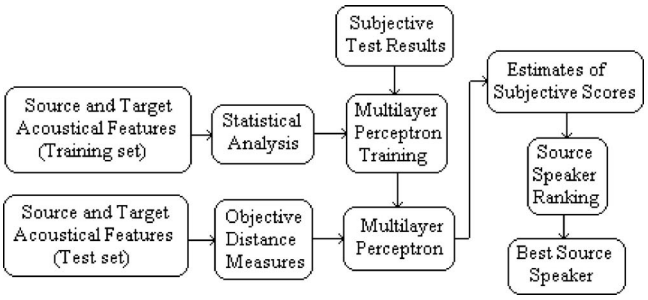


FIG. 1. Flowchart of the proposed automatic source speaker selection algorithm.

Figure 1 shows the flowchart of the proposed automatic source speaker selection algorithm. In Sec. II, the voice conversion algorithm is briefly reviewed and an extended subjective listening test is described. The subjective performance of a weighted codebook mapping based voice conversion algorithm is reported based on output similarity to the target speaker's voice and quality. Statistical test results formally show the dependence of performance on the source speaker given the target speaker. Section III describes the acoustic features employed for estimating objective distances between source and target speakers. The ratio of means and standard deviations as well as the rank-sum values obtained by the Wilcoxon rank-sum testing procedure (Wilcoxon, 1945; Mann and Whitney, 1947) are employed for objective measurement of variability in feature distributions. In Sec. IV, the proposed ANN-based source speaker selection algorithm that estimates the subjective performance and source speaker rankings from a set of objective distance measures is described. In Sec. V, the performance of the proposed algorithm is evaluated in estimating the subjective scores and source speaker rankings using tenfold cross-validation. Finally, the study is concluded with a discussion of the results in Sec. VI.

II. SUBJECTIVE LISTENING TEST

A. Weighted codebook mapping based voice conversion

"Speaker transformation algorithm using segmental codebooks" (STASC) is employed for the voice conversion task (Arslan, 1999) with several refinements as proposed by Turk and Arslan (2006). In the training stage of STASC, the mapping between the source and target acoustical parameters is learned. Identical sentence utterance recordings from the source and target speakers are aligned with an HMM based phoneme recognizer (Woodland *et al.*, 1994). Acoustic fea-

tures including LSFs (Huang *et al.*, 2001), f0, duration, and energy are computed. LSFs in corresponding source and target phoneme pairs are averaged to obtain a smoothed estimate of the source and target vocal tract characteristics for each source-target phoneme pair. Outliers are eliminated based on distance measures obtained from LSFs, pitch, duration, and energy as described by Turk and Arslan (2006). Application of smoothing with averaging and elimination of outliers reduce problems due to automatic alignment mismatches. In the transformation stage, input LSFs extracted on a frame-by-frame basis are matched with the source speaker training LSFs and the corresponding target LSFs are determined. For each input LSF vector, the distances to each averaged source training LSFs are calculated using inverse harmonic mean weighting (Laroia *et al.*, 1991). The best matches (typically four to ten) are selected and a weighted average of the corresponding target training LSF vectors is target. computed. This target LSF estimate is converted to LP coefficients. The target vocal tract spectrum is reconstructed and multiplied with the linear prediction LP-residual spectrum. Prosodic modifications are performed simultaneously using FD-PSOLA (Moulines and Verhelst, 1995) to obtain the time domain waveform. During FD-PSOLA, pitch scaling is applied for matching the mean and the standard deviation of the target pitch. Duration scaling can also be applied. However, this was not required in our case since the source speakers were asked to mimic the reference speaker's timing during database collection. We have not applied intensity transformations since it generally does not contribute to similarity to target voice (Turk, 2003).

B. Database

We have collected a database of twenty utterances and three sustained vowels from ten male and ten female native Turkish speakers in an acoustically isolated recording room. The electro-glottograph (EGG) recordings were collected simultaneously. One of the male speakers was selected as the reference and the remaining speakers were told to mimic the timing of the reference speaker as close as possible to improve overall voice conversion performance (Kain, 2001). We have avoided selecting different reference speakers for male and female sets in order to be able to use the database for cross-gender transformations in the future.

C. Procedure

Male-to-male and female-to-female conversions were considered separately. Intergender transformations were avoided as our previous experience shows that almost always the transformations within same gender are more successful (Turk, 2003). The main reason for quality degradation in intergender transformations is the extensive amount of pitch modification required. It is known that naturalness of PSOLA output decreases for pitch scaling factors less than 0.8 or greater than 1.2 (Pacheco and Seara, 2002). Each speaker was considered as the target and voice conversion was performed from the remaining nine speakers of the same gender to that target speaker. Therefore, 90 source-target pairs were available for each gender, resulting in a total number of 180 different source-target pairs. Eighteen utterances were used in training and two utterances for testing. One of the test utterances was a neutral sentence and the other was a question sentence.

TABLE I. Reference set for the MOS-score.

Coder or recording format	Bit rate (kbps)	MOS
PCM	64	4.4
ADPCM (G.726)	32	4.2
LD-CELP (G.728)	16	4.2
CSA-CELP (G.729)	8	4.2
CELP	4.8	4.0
LPC-10 (FS 1015)	2.4	2.3

A set of twelve native Turkish listeners (seven male, five female) who had no known hearing problems and who were researchers in speech technology were presented with pairs of target and transformation output recordings. The listeners were asked to assign two scores: similarity to target on a discrete 10-point scale (S-score), and quality on a 5-point mean opinion score (MOS) scale (Q-score). The provided range for S-scores was chosen to be finer as compared to Q-scores in parallel with a previous study on voice conversion (Stylianou *et al.*, 1998). Note that the S-score evaluates similarity to target speaker's voice only and lower S-scores do not mean that the output sounds like the source speaker. Q-score corresponds to the standard MOS scale for sound quality (ITU-T Rec. P.800, 1996; ITU-T Rec. P.830 1996): 1=Very bad quality (the noise is very disturbing and the conversation cannot be understood), 2=Poor quality (the noise is disturbing but the conversation can be understood), 3=Fair quality (the noise disturbs a little, the conversation can be understood), 4=Good quality (the noise does not disturb, the conversation is distinctly understood), 5=Excellent quality (there is no noise, the conversation is clearly and distinctly understood). The listeners were presented with six reference recordings for quality along with reference MOS values as shown in Table I. Eight pairs (four male-male, four female-female) that consisted of original recordings were included for reliability assessment which were successfully identified by all listeners.

D. Results

Tables II and III show the average similarity to target scores (S-scores) corresponding to conversions among all source-target speaker pairs and median values obtained for each source and target speaker. Note that transformations from/to the reference speaker's voice are included in all median calculations for male-to-male transformations. Median S-scores were 5.56/10.00 (minimum: 2.75, maximum: 8.08) for male-to-male transformations and 7.27/10.00 (minimum: 4.92, maximum: 8.75) for female-to-female transformations.

We have applied one-way analysis of variance (ANOVA) (Witte and Witte, 2006) to show that the mean S-scores for each target vary with the source speaker. For this purpose, for each target, the S-scores from each listener are provided as the dependent variable and the corresponding source speaker identity as the independent variable. *F*-ratios

TABLE II. Average S-scores for all male source-target pairs. Speaker 8 is the reference speaker. The best S-score along each row is shown inside round brackets, i.e., (5.46), and the best S-score along each column is shown inside square brackets, i.e., [6.88]. An entry inside both round and square brackets corresponds to the score which is a maximum both in its row and column, i.e., ([6.75]). When the best S-score belongs to the reference speaker, the second best S-score in the corresponding row or column is also shown using the same bracket notation. In the last column and row, the median of average scores is given. The best median scores is shown in bold characters. The bottom-right entry shows the median of all entries in rows 1–10 and columns 1–10.

		Source speaker										
		1	2	3	4	5	6	7	8	9	10	median
Target Speaker	1	⋯	4.75	4.88	4.21	5.08	7.04	4.21	(7.96)	4.25	5.38	4.88
	2	2.75	⋯	[6.88]	4.88	6.54	[7.25]	([7.58])	([8.08])	[7.17]	[6.25]	**6.88**
	3	5.42	4.63	⋯	3.25	4.33	(6.38)	4.58	5.92	5.17	4.83	4.83
	4	5.38	4.42	4.29	⋯	(6.58)	5.17	4.38	(7.58)	4.63	4.67	4.67
	5	6.13	6.38	5.33	5.75	⋯	(7.21)	5.63	(7.63)	5.96	6.08	6.08
	6	(5.46)	3.17	5.04	3.13	5.17	⋯	3.83	5.33	4.17	4.17	4.17
	7	5.46	(6.08)	6.04	5.67	5.29	5.00	⋯	(7.21)	5.92	5.88	5.88
	8	5.50	5.29	6.67	4.00	5.08	6.33	(7.25)	⋯	6.13	4.54	5.50
	9	6.21	([6.75])	5.08	4.13	4.63	5.00	5.79	(7.71)	⋯	5.71	5.71
	10	([7.00])	6.58	6.21	[5.88]	[6.92]	5.83	6.46	(7.63)	6.71	⋯	6.58
	median	5.46	5.29	5.33	4.21	5.17	6.33	5.63	**7.63**	5.92	5.38	**5.56**

and the corresponding p-values for each target are given in Table IV. Similar analysis has been performed for Q-scores as well. We have observed that for all target speakers, both scores change significantly depending on the source speaker for a confidence level of 95%.

For male speaker pairs, almost always the highest S-scores and Q-scores were obtained when the reference speaker was used as the source speaker. This indicates that the performance of voice conversion improves when the source prosody matches the target prosody. Excluding the reference speaker from the set of source speakers, the source speaker that results in the best voice conversion performance varies as the target speaker varies. Therefore, our fundamental hypothesis that the performance of the voice conversion algorithm is dependent on the specific source-target pair chosen is supported.

Average female-to-female similarity scores have been compared with the male-to-male case using a pairwise t-test (Witte and Witte, 2006). The mean of the average female-to-female similarity scores (7.18) was found to be significantly higher than the mean of the average male-to-male similarity scores (5.62) with a p-value of 1.9077×10^{-19}. The reason might be due to the difficulty in identifying female speakers when they mimicked the overall prosody of the reference speaker according to listeners' comments. The listeners have been more critical in the case of male speakers whom they identified better and provided lower similarity scores. This comment is also supported by the informal feedback from the listeners that they had more difficulty in discriminating between female speakers as compared to male speakers. A similar pairwise t-test was applied for comparison of the mean of the average male-to-male and female-to-female quality scores. The mean of female-to-female Q-scores (3.96) was found to be significantly higher than the mean of male-to-male Q-scores (3.68) with a p-value of 2.3601×10^{-8}.

The median scores in the last rows of Tables II and III show that part of the source speakers is not generally appropriate for generating the voices of target speakers. Transformations using these speakers as source speakers resulted in lower median scores. As an example, the median S-score is lowest for both male and female speakers No. 4. The last

TABLE III. Average S-scores for all female source-target pairs. Please refer to the caption of Table II for details on scores marked with round and/or square brackets or shown in bold characters. Note that since the reference speaker was male and we did not consider cross-gender transformations, there are no scores corresponding to the reference speaker in this table. The bottom-right entry shows the median of all entries in rows 1–10 and columns 1–10.

		Source speaker										
		1	2	3	4	5	6	7	8	9	10	Median
Target speaker	1	⋯	6.17	(7.83)	5.04	7.54	7.25	7.46	(7.83)	7.08	6.33	7.25
	2	6.88	⋯	7.92	[7.96]	5.71	7.75	7.25	([8.08])	6.96	6.88	7.25
	3	7.58	6.25	⋯	6.13	7.96	7.46	7.04	7.96	(8.04)	6.17	7.46
	4	7.08	([8.04])	6.33	⋯	5.17	7.38	7.58	8.75	7.13	6.25	7.13
	5	7.08	7.29	7.88	5.67	⋯	8.00	6.96	7.67	([8.67])	5.83	7.29
	6	[7.63]	6.58	([8.67])	6.08	7.79	⋯	[8.08]	8.04	7.63	6.38	7.63
	7	6.88	6.58	7.92	6.67	6.92	([8.38])	⋯	8.00	6.21	6.54	6.88
	8	7.00	6.08	(8.58)	7.21	5.67	8.25	7.42	⋯	7.54	[7.00]	7.21
	9	[7.63]	7.17	7.58	4.92	([8.13])	7.92	6.96	7.33	⋯	6.42	7.33
	10	7.54	6.96	(8.25)	6.88	6.50	8.21	6.29	7.50	7.42	⋯	7.42
	median	7.08	6.58	7.92	6.13	6.92	7.92	7.25	7.96	7.42	6.38	**7.27**

TABLE IV. *F*-ratios and *p*-values after one-way ANOVA of S-scores for each target speakerin male and female speaker sets.

Target speaker (male)	*F*-ratio	*p*-value
1	3.24	0.003 0
2	3.52	0.001 5
3	2.89	0.007 0
4	3.17	0.003 6
5	2.43	0.020 8
6	5.09	0.000 038
7	4.11	0.000 376
8	3.61	0.001 2
9	3.06	0.004 6
10	6.18	0.000 003
Target speaker (female)	*F*-ratio	*p*-value
1	4.09	0.000 4
2	2.27	0.030 4
3	6.91	0.000 0
4	2.55	0.015 7
5	2.70	0.011 0
6	3.06	0.004 6
7	3.19	0.003 4
8	4.19	0.000 3
9	2.99	0.005 5
10	3.77	0.000 844

columns in Tables II and III show that it is more difficult to obtain similar output to the target speakers' voices for part of the target speakers. Male speaker No. 6 and female speaker No. 1 are examples. Similar observations also hold for the Q-scores given in Tables V and VI. For example, male speaker No. 10 and female speaker No. 5 are not good source speaker choices if one is interested in expected output quality. Male speaker No. 10 and female speaker No. 3 are the target speakers for which the median quality is lower when compared with the rest of the target speakers.

III. ACOUSTIC FEATURE ANALYSIS

As a preprocessing step, all recordings were phonetically labeled using phonetic HMMs trained with the hidden Markov model toolkit (Woodland *et al.*, 1994). Pitch marks were extracted from the EGG waveforms using a simple peak detecting algorithm which uses autocorrelation based pitch detection results as a starting point. Phonetic labels and pitch marks were visually inspected and gross errors were corrected manually. For each source-target speaker pair, phonetic labels were used in determining the time alignment.

A. Acoustic features

A set of acoustic features that are relevant for describing the differences among speakers in terms of physiology, articulation, prosody, and voice quality was selected. As it was impossible to collect a database for all possible ranges of a large number of acoustic features, we have decided to extract various common features and to use an algorithm that can automatically learn from these features for the selection of the best source speaker for a given target speaker. Vocal tract characteristics have been modeled with LSFs since they are known to relate well with perceptual phenomena, i.e., closer LSF pairs are likely to correspond to formant locations which are perceptually significant in characterizing speakers and articulation patterns. For the objective evaluation of similarity/dissimilarity of prosody characteristics, various dimensions related to prosody are considered including pitch, duration, and energy. Two issues were taken into account considering prosody related problems in voice conversion.

- Automatic alignment accuracy. Training stage of voice conversion requires accurate alignments of the source and target recordings for reliable estimation of the transformation functions. Large mismatches in prosodic characteristics may cause alignment mismatches and reduce voice conversion performance.
- Reduction in naturalness with large amounts of prosody modifications. Large prosodic differences between the source and the target speaker lead to the requirement of excessive amounts of prosody modifications which in turn reduce quality.

We have also selected a set of voice quality features and features that compare the background acoustic environment.

TABLE V. Average Q-scores for all male source-target pairs. Speaker 8 is the reference speaker. Please refer to the caption of Table II for details on scores marked with round and/or square brackets or shown in bold characters. The bottom-right entry shows the median of all entries in rows 1–10 and columns 1–10.

		Source speaker										
		1	2	3	4	5	6	7	8	9	10	Median
Target speaker	1	⋯	3.67	[3.71]	3.50	[3.96]	([4.25])	3.75	(4.29)	4.17	3.33	3.75
	2	3.83	⋯	3.17	3.54	3.88	4.08	([4.46])	4.29	[4.42]	3.21	3.88
	3	3.54	3.29	⋯	3.13	3.08	3.38	3.21	(4.08)	3.17	3.00	3.21
	4	3.46	3.46	3.58	⋯	3.79	3.71	3.83	(4.29)	(4.21)	2.96	3.71
	5	(3.92)	3.71	3.29	3.58	⋯	3.63	3.46	3.88	(3.92)	3.21	3.63
	6	3.96	3.75	3.54	3.58	3.88	⋯	3.75	3.92	(4.21)	3.25	3.75
	7	3.92	4.00	3.42	3.71	3.71	3.92	⋯	4.17	(4.33)	3.29	**3.92**
	8	3.29	3.67	[4.21]	3.71	3.46	3.83	(4.29)	⋯	3.83	3.17	3.71
	9	([4.25])	[4.13]	3.17	[3.79]	3.92	3.92	3.92	([4.33])	⋯	[3.54]	**3.92**
	10	3.54	3.04	2.92	3.08	3.17	3.13	2.83	(3.79)	(3.71)	⋯	3.13
	Median	3.83	3.67	3.42	3.58	3.79	3.83	3.75	**4.17**	**4.17**	3.21	3.71

TABLE VI. Average Q-scores for all female source-target pairs. Please refer to the caption of Table II for details on scores marked with round and/or square brackets or shown in bold characters. Note that since the reference speaker was male and we did not consider cross-gender transformations, there are no scores corresponding to the reference speaker in this table.

		Source speaker										
		1	2	3	4	5	6	7	8	9	10	Median
Target Speaker	1	⋯	3.79	4.25	3.83	3.58	4.08	3.79	(4.29)	3.92	4.00	3.92
	2	3.88	⋯	4.08	3.79	3.63	(4.17)	4.04	4.04	3.75	3.75	3.88
	3	(4.13)	3.71	⋯	3.92	3.75	3.88	4.00	3.58	3.92	3.50	3.88
	4	3.79	[4.25]	3.75	⋯	3.79	3.88	4.04	(4.42)	3.71	3.75	3.79
	5	[4.21]	3.88	3.75	3.83	⋯	4.04	3.79	3.88	([4.25])	3.67	3.88
	6	4.17	4.00	[4.29]	3.67	3.92	⋯	[4.25]	([4.50])	[4.25]	[4.25]	**4.25**
	7	3.92	4.13	4.17	3.92	3.71	4.21	⋯	(4.25)	3.67	4.04	4.04
	8	4.17	3.88	(4.25)	4.00	3.71	3.79	4.04	⋯	3.88	4.04	4.00
	9	4.04	4.17	4.21	([4.25])	[4.13]	([4.25])	3.46	4.00	⋯	4.00	4.13
	10	4.00	3.75	3.96	3.92	3.63	3.88	4.00	(4.21)	3.96	⋯	3.96
	Median	4.04	3.88	4.17	3.92	3.71	4.04	4.00	**4.21**	3.92	4.00	**3.96**

Variations in these features may lead to further problems in automatic alignment, acoustical space mapping, and accurate estimation of target features from training data.

1. Vocal Tract

LSFs were computed on a frame-by-frame basis using a LP analysis order of 20 at 16 kHz. A fixed window size of 20 ms was used with a fixed skip rate of 10 ms. Preemphasis with a digital filter of the form $P(z)=1-0.97z^{-1}$ was applied. Inverse harmonic mean weighting based LSF distance (Laroia *et al.*, 1991) was employed for calculating the LSF distances between the source and target speech frames. The weighting procedure assigns relatively higher weights to closer LSF pairs which are known to correspond to formant locations resulting in a better perceptual representation as compared to using no weights at all. The LSF distances employed in vocal tract distance computation in various settings are shown in Table VII.

Note that frame based LSF distances account for the instantaneous variations in the vocal tract whereas phoneme-based LSF distances represent the more global, phoneme-level differences in vocal tract characteristics.

2. Pitch

Fundamental frequency (f0) contours for all source and target recordings were extracted using the pitch marks obtained from the corresponding EGG signals. The mean and the standard deviation of f0 values were computed for each speaker. Figure 2 shows interpolated histograms of f0 values for two source-target speaker pairs. Note that all histograms presented in this paper were obtained using appropriately small bins. Cubic spline interpolation was then performed to obtain smoothed versions to enhance visibility, especially for regions where two histograms overlap. Histogram smoothing also helped to reduce the effects of noise and outliers since histogram values were estimated from limited amount of data. The acoustical distance measures employed for pitch characteristics are given in Table VII.

TABLE VII. A summary of the acoustic features employed, which part of the recordings they were extracted from, and the distance measures employed for each of them.

Acoustic features	Applied to	Distance measures employed
Frame based LSFs	Sentences	LSF distance (mean, standard deviation, rank-sum value)
Averaged LSFs for phonemes	Sentences	LSF distance (mean, standard deviation, rank-sum value)
f0	Sentences	Mean ratio, standard deviation ratio, rank-sum value
f0 (phoneme level)	Sentences	Mean and standard deviation of f0 difference
Phoneme durations	Sentences	Mean ratio, standard deviation ratio, rank-sum value
Utterance durations	Sentences	Mean ratio, standard deviation ratio, rank-sum value
Silence durations (in the beginning)	Sentences	Mean ratio, standard deviation ratio, rank-sum value
Silence durations (atthe end)	Sentences	Mean ratio, standard deviation ratio, rank-sum value
Pause durations	Sentences	Mean ratio, standard deviation ratio, rank-sum value
Energy	Sentences	Mean ratio, standard deviation ratio, rank-sum value
Spectral tilt	Sustained vowels	Mean ratio, standard deviation ratio, rank-sum value
Open quotient	Sustained vowels	Mean ratio, standard deviation ratio, rank-sum value
Jitter	Sustained vowels	Mean ratio, standard deviation ratio
Shimmer	Sustained vowels	Mean ratio, standard deviation ratio
Voiced frame ratio	Sentences	Ratio
SPI	Sentences	Mean ratio, standard deviation ratio, rank-sum value
H1-H2	Sentences	Mean ratio, standard deviation ratio, rank-sum value
EGG shape parameters	Sustained vowels	Mean ratio, standard deviation ratio, rank-sum value

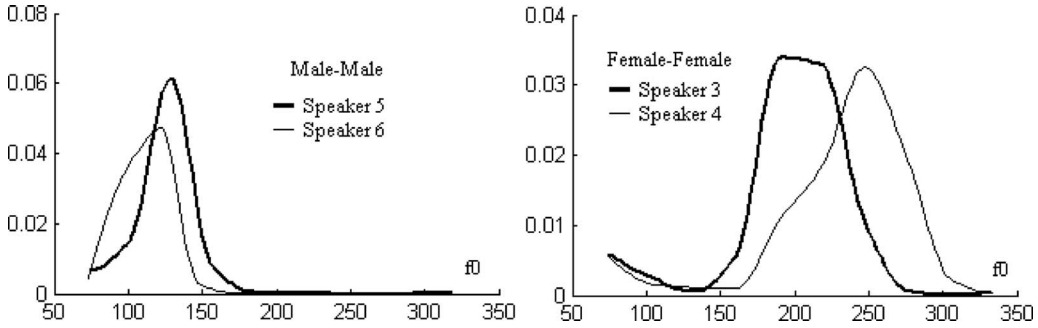


FIG. 2. Sample distributions of f0 values for a male speaker pair and a female speaker pair.

3. Duration

The duration of each phoneme was determined from the phonetic labels. The distance measures used for durational characteristics are shown in Table VII.

4. Energy

The energy contours for all source and target speech signals were extracted using a root-mean-square (rms) energy estimator on a frame-by-frame basis. The window size and the skip rate were set to 20 ms and 10 ms, respectively. The mean and the standard deviation of rms energy values were computed for each speaker. The average rms energy value for each phoneme was also computed. Table VII shows the acoustical distance measures employed for energy characteristics.

5. Spectral tilt

The spectral tilt was computed by fitting a LP model of order 2 to sustained vowel recordings. The corresponding LP spectrum estimate was calculated and spectral amplitude values were converted to decibels. A least-squares line was fitted between the global peak value and the value at 3 kHz. The slope of this line was used as a simple parameter for spectral tilt. Figure 3 shows an example for spectral tilt estimation. Figure 4 shows interpolated histograms of spectral tilt values for two source-target speaker pairs. The spectral tilt values were estimated using fixed window size and skip rate frames from two recordings of the sustained vowel /aa/. The distance measures for spectral tilt include ratio of means, ratio of standard deviations, and rank-sum values estimated using source and target spectral tilt values.

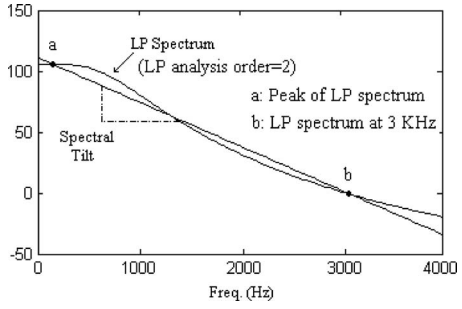


FIG. 3. Spectral tilt estimation. Point (a) labels the spectral peak of the LP spectrum obtained with an analysis order of 2, and point (b) labels the LP spectrum at 3 kHz.

6. Open quotient

For each period of the EGG signal, the open quotient was estimated as the ratio of the positive segment of the signal to the length of the signal (Childers and Lee, 1991). The open quotient values were estimated on a frame-by-frame basis from two recordings of the sustained vowel /aa/ for each speaker pair. Figure 5 shows an example for open quotient estimation from one period of the EGG signal. Figure 6 shows interpolated histograms of open quotient values for two source-target speaker pairs. The distance measures employed for open quotient included ratio of means, ratio of standard deviations, and rank-sum values estimated using source and target: open quotient values.

7. Jitter and shimmer

Jitter and shimmer are the relative evaluation of the period-to-period variability of pitch and intensity in the speech signal. These features were computed for sustained vowels using the equations provided in the Kay Elemetrics multidimensional voice program manual (Operations Manual, 1993).

8. Voiced frame ratio

The total number of voiced frames was determined from pitch contours and the ratio of voiced frames to total speech frames was computed. The ratio of the voiced frame ratio values for the source and the target speakers is used as a distance measure between the two speakers.

9. Soft phonation index

Soft phonation index (SPI) is the ratio of the lower-frequency harmonic energy in the range 70–1600 Hz to the upper-frequency harmonic energy in the range 1600–4500 Hz. SPI was calculated using discrete fourier transform (DFT) spectrum estimated from pitch-synchronous frames of approximately three pitch periods in length. SPI

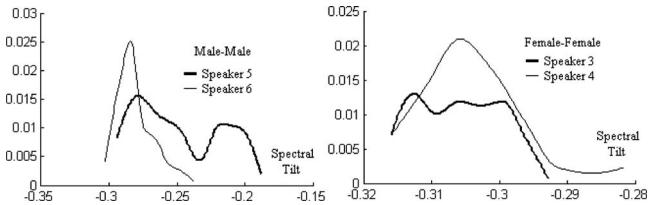


FIG. 4. Sample distributions of spectral tilt values for a male speaker pair and a female speaker pair.

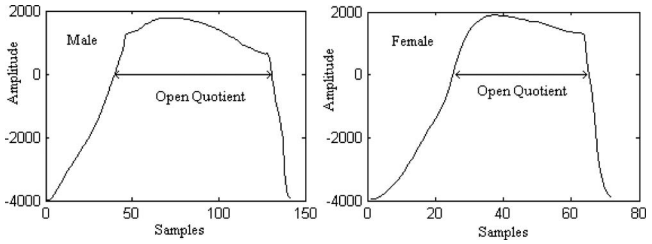


FIG. 5. Open quotient estimation from EGG signals.

values were computed by peak picking on the DFT spectrum for the lower- and upper-frequency ranges and averaging respective sum of harmonic amplitudes. Figure 7 shows the spectrum of a speech frame and the corresponding harmonics. The harmonics in the lower-frequency range are marked with a "+" sign and those in the higher-frequency range are marked with a "○" sign. The objective distance measures employed for SPI included ratio of means, ratio of standard deviations, and the rank-sum values estimated using source and target SPI values.

10. H1-H2

H1-H2 is the amplitude difference of the first and the second harmonic in the spectrum (Hanson, 1995). These two parameters can be used in identifying breathiness or creakiness in the voice characteristics. For example, breathy voices tend to produce stronger H1, and creaky voices tend to produce weaker H1. H1-H2 values were estimated frame by frame and the resulting samples are used as the sample distribution. Figure 8 shows sample distributions of H1-H2 values for two source-target pairs for a sustained vowel. For the male speaker pair, the H1-H2 distributions look more similar as compared to the female speaker pair. For the female speaker pair, we observe that the distribution for "female speaker 4" is flatter, indicating that her voice alters more. The objective distance measures employed for H1-H2 include ratio of means, ratio of the standard deviations, and rank-sum values estimated using source and target H1-H2 values.

11. Electroglottograph signal period shape

We have used three parameters to characterize one period of the EGG signals, as shown in Fig. 9. The first parameter, α, is the slope of the least-squares line fitted from the glottal closure instant to the peak of the EGG signal. The second parameter, β, is the slope of the least-squares line fitted to the segment of the EGG signal when the vocal folds are open. Finally, the third parameter, γ, is the slope of the least-squares line fitted to the segment of the EGG signal

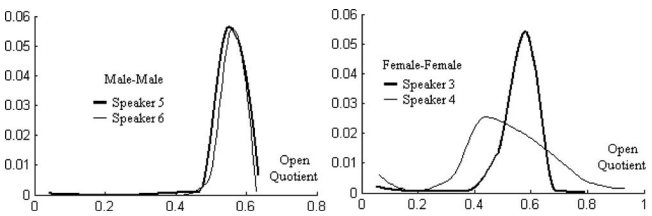


FIG. 6. Sample distributions of open quotient values for a male speaker pair and a female speaker pair.

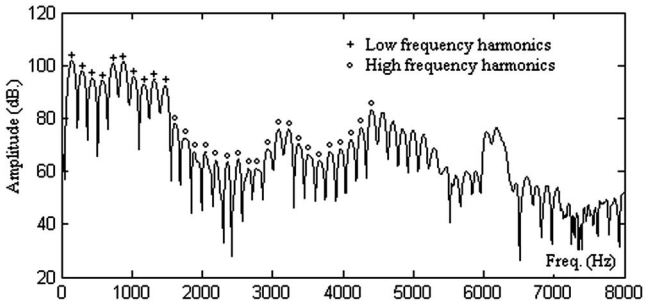


FIG. 7. Estimation of the ratio of lower harmonic energy to the higher harmonic energy from the DFT spectrum using lower f0 harmonics (+) and higher f0 harmonics (○).

corresponding to the case when the vocal folds are close. Sample distributions for these parameters are shown for two female speakers in Fig. 10.

The algorithm simply fits three lines to the following segments of each EGG period: from start of the period to 25% of the period in duration, from 25% to 75% of the period in duration, and from 75% of the period to the end of the period. The objective distance measures for EGG shape included the ratio of means, the ratio of standard deviations, and rank-sum values estimated using source and target EGG shape parameters.

12. Acoustic background

LSF parameters for non-speech frames were used in order to represent the acoustic background. Our previous experience has shown that differences in acoustic background result in significant performance degradation in voice conversion (Turk and Arslan, 2006). Therefore, a measure of the acoustic difference of acoustic backgrounds might serve as a useful feature for a source speaker selection algorithm. The distance measures corresponding to acoustic background differences included the mean, the standard deviation, and the rank-sum value estimated using source and target LSF distances corresponding to silent frames in the beginning and at the end of the utterances.

B. Objective distance measures

Three major distance measures were employed for different acoustic features to provide objective distance measures as input to the ANN: rank-sum values as calculated by the Wilcoxon rank-sum test (Wilcoxon, 1945), the ratio of means, and the ratio of standard deviations. Wilcoxon rank-sum test is a nonparametric alternative to the two-sample *t*-test (Wild and Seber, 1999). It is valid for data from any distribution and is much less sensitive to outliers as com-

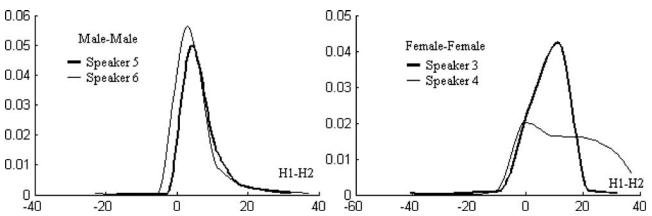


FIG. 8. Sample distributions of H1-H2 values for a male speaker pair and a female speaker pair.

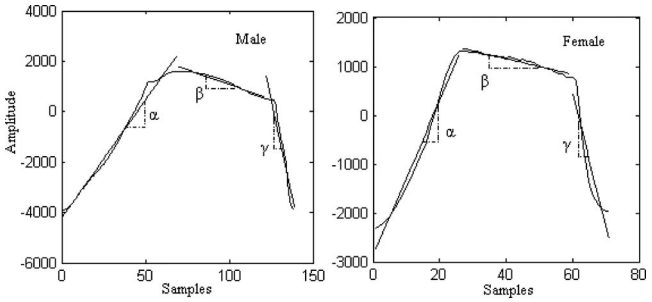


FIG. 9. Simple model for EGG shape.

pared to the two-sample *t*-test. These properties make it a useful tool for comparing acoustic features which may not be well modeled by parametric distributions. The test reacts not only to the differences in the means of distributions but to the differences between the shapes of the distributions. A lower rank-sum value indicates that the two distributions under comparison are closer.

We have used the rank-sum value as the objective distance between two sample distributions in comparing the acoustical characteristics of source-target pairs for most of the features. Several features were already based on averages and rank-sum computation was not used for them including jitter, shimmer, voiced frame ratio, and phoneme-based average f0 difference. The ratios of means and standard deviations were used as additional objective distance measures for f0 (frame-by-frame f0 values and average absolute f0 difference in corresponding source and target phonemes), duration (phoneme duration, word duration, utterance duration, and silence duration), energy (frame-by-frame energy values and average absolute energy difference in corresponding source and target phonemes), spectral tilt, open quotient, jitter, shimmer, voiced frame ratio, SPI, H1-H2, and EGG shape parameters. Table VII shows a summary of the acoustic features employed, which part of the recordings they were extracted from, and the distance measures employed for each of them.

C. Correlation Analyses

In order to investigate the linear relationship among the distance measures and the subjective test scores, we have performed correlation analyses. We have selected six utterances (five neutral, one question) and two /aa/ vowels from the collected database. The statistics of the acoustical distance measures described Sec. III A are estimated for all combinations of source-target pairs. The correlation coefficients of each acoustic parameter with the subjective scores (S-score and Q-score) are computed. The results show that the greatest correlation value was 0.502 for the distance between the standard deviations of source and target pitch values. However, even the highest correlation value does not imply a strong linear relationship. Therefore, a nonlinear learning scheme was required for the estimation of subjective scores and rankings of the source speakers for a given target speaker.

IV. AUTOMATIC SOURCE SPEAKER SELECTION ALGORITHM

In order to estimate subjective scores without actually carrying out a subjective listening each time a source speaker needs to be selected among a set of possible candidates for a given target speaker, we have trained an ANN. The main objective was to provide a large number of acoustical distance measures between a given target speaker and a source speaker candidate and let the ANN learn which parameters contribute to what extent to the expected subjective quality of transformations from that source speaker. Our hypothesis was that the performance of weighted codebook mapping based voice conversion can be dependent on complex interactions between the source and the target speaker's vocal tract and prosody characteristics. Therefore, we have provided the ANN with a large number of common acoustic features and trained it to estimate the similarity and quality scores.

The scores obtained in the subjective listening test described in Sec. II and the distance values obtained for each feature were used in training an ANN. A three-layer network structure with ten units in the hidden layer was used. The input vectors were 49-dimensional and the output vectors were 2-dimensional (S-score and Q-score). The optimum number of hidden units was determined by an exhaustive search of candidate values between 5 and 20 on a separate validation set. The training algorithm used the gradient-descent algorithm (Duda *et al.*, 2001) to update the network weights iteratively. Random initialization of the network weights, learning-rate adaptation, and momentum were also employed (Duda *et al.*, 2001). Each input dimension was normalized to have zero mean and unity variance. Each output dimension was scaled linearly in the range [0,1] before training the ANN and rescaled to its original range when the ANN was used for estimating the subjective scores after training.

Figure 1 shows the flowchart of the automatic source speaker selection algorithm. The source and target recordings are acoustically analyzed. The distributions and statistical

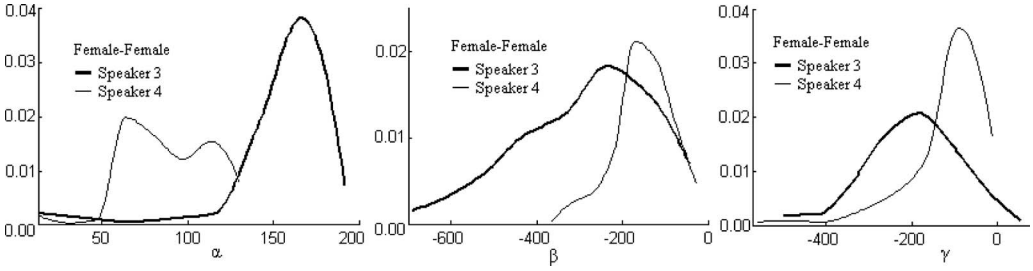


FIG. 10. Sample distributions of EGG shape parameters α, β, and γ for a male speaker pair and a female speaker pair.

TABLE VIII. Mean (μ) and standard deviation (σ) values of subjective scores estimated by the ANN after tenfold cross-validation and testing.

	Tenfold cross validation set		Test set	
Performance Measure	E_S (μ,σ)	E_Q (μ,σ)	E_S (μ,σ)	E_Q (μ,σ)
Male to male	(0.26, 0.51)	(0.09, 0.11)	(0.28, 0.12)	(0.13, 0.19)
Female to female	(1.40, 1.63)	(0.33, 0.33)	(1.26, 1.84)	(0.23, 0.39)
Overall	(0.83, 1.06)	(0.21, 0.26)	(0.77, 1.03)	(0.18, 0.23)

characteristics of the acoustical parameters are estimated. The objective distances between the distributions are calculated using the Wilcoxon rank-sum test and/or the means and the standard deviations. Then, the objective distances are used as the input to the ANN that outputs the estimates of the subjective similarity score (S-score) and the subjective quality score (Q-score). The algorithm outputs the best source speaker candidates that have relatively higher S-scores and Q-scores.

V. EVALUATIONS

The performance of the proposed algorithm was evaluated using tenfold cross-validation. For this purpose, two male and two female speakers were reserved as the test set. The training and validation steps at each cross-validation were as follows.

- Two male and two female speakers were reserved as the validation set. The objective distances among the remaining male speaker pairs and female speaker pairs were used as the input to the ANN and the corresponding subjective scores as the output. Therefore, input vectors were 49-dimensional (all objective distance measures) and the output vectors were 2-dimensional (i.e., the S-score and the Q-score).
- After training, the subjective scores were estimated for the target speakers in the validation set and the error for the S-score and the Q-score was calculated. The error on each cross-validation step was defined as the absolute difference between the output of the ANN and the subjective test results.

The two steps described above were repeated ten times by using randomly selected target speakers as the validation set. Note that none of these target speakers were used in training the ANN. Since we had 12 listeners, 180 source-target pairs, and 2 transformation outputs for each source-target pair, there were a total of 4320 S-score and Q-score pairs and corresponding 49-dimensional distance vectors available for training and testing. As two male and two female target speakers were reserved for testing, 1248 S-score and Q-score pairs were used at each cross-validation step, leaving 3072 pairs for training.

The average cross-validation errors were computed as the average of the errors in the individual steps. Finally, the ANN was trained using all speakers excluding those in the test set. Table VIII shows the means and the standard deviations of error values as obtained on the validation and test sets. Note that E_S and E_Q correspond to the error in S-scores and Q-scores, respectively. Means and standard deviations of errors for male-to-male transformations and female-to-female transformations are reported along with the overall error statistics. We observe that the score estimates are significantly better for male-to-male transformations as compared to female-to-female transformations. This is also verified with a statistical comparison of E_S and E_Q values for the two cases. The mean errors for both similarity and quality scores for male-to-male transformations were found to be lower than that of female-to-female transformations with a p-value of 0.0. On contrary to what one would expect, the algorithm performed slightly better on the test set as compared to the validation set for female-to-female transformations and for the overall case.

TABLE IX. Normalized cross-correlation coefficient statistics for tenfold cross-validation and testing for rankings obtained according to the subjective listening test and estimated using the ANN score estimates.

	Tenfold cross-validation set		Test set	
Performance measure	R_S (μ,σ)	R_Q (μ,σ)	R_S (μ,σ)	R_Q (μ,σ)
Male to male	(0.88, 0.11)	(0.79, 0.08)	(0.84, 0.12)	(0.78, 0.21)
Female to female	(0.61, 0.29)	(0.56, 0.26)	(0.58, 0.25)	(0.58, 0.16)
Overall	(0.74, 0.25)	(0.68, 0.19)	(0.71, 0.21)	(0.68, 0.20)

In order to evaluate the performance of the proposed algorithm further, the rankings predicted by the algorithm were compared with the preferences of the listeners as follows.

- For each target speaker, source speakers were sorted in decreasing order according to median similarity and quality scores given in Tables II–V.
- For each target speaker in the cross-validation and test stages, source speakers were sorted in decreasing order according to the similarity and quality scores as estimated by the ANN.
- For each target speaker in the cross-validation and test stages, normalized cross-correlation coefficient between the original and ANN-estimated versions of sorted source speaker indices was calculated.

Table IX shows the means and the standard deviations of normalized cross-correlation coefficients between the original and ANN-based rankings of the source speakers for tenfold cross-validation and test steps including male-to-male transformations, female-to-female transformations, and overall cases separately. We observe that the performance of the automatic source selection algorithm is significantly better for estimating both similarity and quality based rankings for male-to-male transformations as compared to female-to-female transformations. This is also verified by pairwise t-tests where the following p-values are obtained: 0.0012 (similarity, cross-validation), 0.0003 (quality, cross-validation), 0.0026 (similarity, test), and 0.0033 (quality, test). Therefore, estimating the source speaker rankings for female-to-female transformations war in general harder.

VI. CONCLUSIONS AND DISCUSSIONS

In this study, the dependence of a weighted codebook mapping based voice conversion algorithm is investigated in a subjective listening test. We have shown that both similarity to target speaker's voice and output quality vary depending on the source speaker for a given target voice. The second goal of the study is to develop an automatic source speaker selection algorithm using a small amount of recordings from the source and target speakers. An algorithm is proposed which estimates the subjective scores for voice conversion from a set of objective distance measures between the source and target speaker's acoustic features. The algorithm learns the relationship of the subjective scores and the objective distance measures through nonlinear regression with an ANN. Once the ANN is trained, the algorithm can be used in ranking of a set of source speakers in terms of the expected similarity to the target voice and output quality. The average correlation between source speaker rankings obtained in the subjective listening test and rankings estimated by the ANN was found to be quite reliable for male-to-male transformations. Specifically, an average correlation of 0.84 was observed for similarity and 0.78 for quality. However, the results for female-to-female transformations indicate requirement for improvements since the average correlation score for both similarity and quality was only 0.58. The lower performance in the case of female-to-female conversions can be explained by the informal feedbacks of the listeners that they had found it harder to discriminate among female speakers in the test. As a result, the subjective scores for female-to-female case had less variation making it harder for the ANN to extract ranking patterns. This is confirmed by comparing the variances of male-to-male and female-to-female scores. The variances of female-to-female conversion scores (0.7187 and 0.0482 for similarity and quality, respectively) were significantly smaller as compared to the variances in the male-to-male conversion cases (1.4082 and 0.1616 for similarity and quality, respectively).

One should note that the results presented can be dependent on the database as well as the voice conversion algorithm employed. Therefore, developing similar automatic source speaker selection methods for different voice conversion algorithms, carrying out tests using different speakers, and comparing the results with the results reported in this study remain as future research tasks. A number of possibilities exist for improving the performance of the proposed automatic source speaker selection algorithm as well. First of all, different ANN architectures can be employed and parameters optimized separately for source speaker selection. Other regression algorithms based on RBFNs, decision trees, or support vector machines (Kecman, 2001) can be tested. The algorithm can be modified to estimate the rankings of source speakers directly instead of estimating the subjective scores and performing the ranking according to the subjective scores. For this purpose, the subjective listening test can be performed in the form of a source speaker ranking evaluation. The subjects can be asked to rank a number of source speaker candidates in terms of overall voice conversion performance for a given target speaker. This kind of testing would provide direct rankings rather than indirect rankings obtained from similarity and quality scores.

In an initial attempt to use decision trees instead of ANNs to perform the ranking among source speakers, we have observed pitch and voice quality features to be highly discriminative in estimating the subjective scores for some source speakers. As an example, male speaker No. 4 who had a deep and low-pitched voice and who was one of the worst source speaker candidates according to average scores could be easily separated from the rest of the source speakers by using H1-H2 and pitch features in the upper leaves of the decision tree. Such information can be used to investigate a subset of optimal features to be used in training improved ANNs for the source speaker selection task.

The set of acoustic features can be expanded, for example, using cepstral coefficients or other perceptually motivated representations for the vocal tract characteristics, and more style specific features for prosody characteristics. Parametric modeling of acoustic feature distributions using appropriate models may also improve robustness for different speakers. Acoustic distance measurement can also be based on differences from a world model of speakers trained using a large number of speakers. Using such models resulted in significant performance improvements in speaker identification (Fredouille *et al.*, 1999). In order to further improve the performance, new sources of information can be incorporated. As an example, the set of source-target utterances used in source speaker selection can be used for assessing the HMM based alignment performance. Fast voice conversion can be performed using this restricted set and the transformation output can be analyzed in order to estimate new acoustical distance measures including transformed-to-target and transformed-to-source acoustical distances.

ACKNOWLEDGMENTS

Part of this research is supported by The Scientific and Technological Research Council of Turkey TUBITAK (TEYDEB 3040039).

Abe, M., Nakamura, S., Shikano, K., and Kuwabara, H. (**1988**). "Voice conversion through vector Quantization," in Proceedings of the IEEE ICASSP, pp. 565–568.

Acero, A. (**1993**). *Acoustical and Environmental Robustness in Automatic Speech Recognition* (Kluwer Academic, Norwell, MA).

Acero, A. (**1998**). "A mixed-excitation frequency domain model for time-scale pitch-scale modification of speech," in Proceedings of the ICSLP 1998, Sydney, Australia, pp. 1923–1926.

Arslan, L. M. (**1999**). "Speaker transformation algorithm using segmental codebooks," Speech Commun. **28**, 211–226.

Bhuta, T., Patrick, L., and Garnett, J. D. (**2004**). "Perceptual evaluation of voice quality and its correlation with acoustic measurements," J. Voice **18**, 299–304.

Bi, N., and Qi, Y. (**1997**). "Application of speech conversion to alaryngeal speech enhancement," IEEE Trans. Speech Audio Process. **5**, 97–105.

Childers, D. G. (**1995**). "Glottal source modeling for voice conversion," Speech Commun. **16**, 127–138.

Childers, D. G., and Lee, C.-K. (**1991**). "Vocal quality factors: Analysis, synthesis, and perception," J. Acoust. Soc. Am. **90**, 2394–2410.

Drioli, C. (**1999**). "Radial basis function networks for conversion of sound spectra," in Proceedings of the Second COST G-6 Workshop on Digital Audio Effects (DAFx99), NTNU, Trondheim.

Duda, R. O., Hart, P. E., and Stork, D. G. (**2001**). *Pattern Classification* (Wiley, New York).
Fant, G., Liljencrants, J., and Lin, Q. (**1985**). "A four-parameter model of the glottal flow," Speech Transmission Laboratory Quarterly Progress and Status Report No. 4, Royal Institute of Technology, Stockholm, Sweden.
Flanagan, J. L., and Golden, R. M. (**1966**). "Phase vocoder," Bell Syst. Tech. J. **45**, 1493–1500.
Fredouille, C., Bonastre, J.-F., and Merlin, T. (**1999**). "Similarity normalization method based on world model and a posteriori probability for speaker verification," in Proceedings of the Eurospeech, pp. 983–986.
Hanson, H. M. (**1995**). "Individual variations in glottal characteristics of female speakers," in Proceedings of the IEEE ICASSP 1995, pp. 772–775.
Holmes, W., Holmes, J., and Judd, M. (**1990**). "Extension of the bandwidth of the JSRU parallel-formant synthesizer for high quality synthesis of male and female speech," in Proceedings of the IEEE ICASSP, Vol. **90**, pp. 313–316.
Huang, X., Acero, A., and Hon, H.-W. (**2001**). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development* (Prentice-Hall, Upple Saddle River, NJ).
Inanoglu, Z., and Young, S. (**2007**). "A system for transforming the emotion in speech: Combining data-driven conversion techniques for prosody and voice quality," in Proceedings of the Interspeech 2007, Antwerp, Belgium, pp. 490–493.
ITU-T Rec. P.800 (**1996**). "Methods for subjective determination of transmission quality," International Telecommunication Union, Geneva, Switzerland, August.
ITU-T Rec. P.830, (**1996**). "Subjective performance assessment of telephone-band and wideband digital codecs," International Telecommunication Union, Geneva, Switzerland, February.
Kain, A. (**2001**). "High resolution voice transformation," Ph.D. thesis, Oregon Health and Science University, Portland, OR.
Kain, A., and Macon, M. (**1998**). "Personalizing a speech synthesizer by voice adaptation," Proceedings of the Third ESCA/COCOSDA International Speech Synthesis Workshop, pp. 225–230.
Kecman, V. (**2001**). *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models* (The MIT Press, Cambridge, MA).
Knohl, L., and Rinscheid, A. (**1993**). "Speaker normalization with self-organizing feature maps," in Proceedings of the IJNN-93-Nagoya, International Joint Conference on Neural Networks, pp. 243–246.
Kreiman, J., and Gerratt, B. R. (**2003**). "Jitter, shimmer, and noise in pathological voice quality perception," in Proceedings of the VOQUAL 2003, pp. 57–62.
Laroche, J., Stylianou, Y., and Moulines, E. (**1993**). "HNS: Speech modification based on a harmonic+noise model," in Proceedings of the IEEE ICASSP-93, Minneapolis, MN, pp. 550–553.
Laroia, R., Phamdo, N., and Farvardin, N. (**1991**). "Robust and efficient quantization of speech LSP parameters using structured vector quantizers," in Proceedings of the IEEE ICASSP, pp. 641–644.
Maddieson, I., and Ladefoged, P. (**1985**). "Tense and lax in four minority languages of China," J. Phonetics **13**, 433–454.
Makhoul, J. (**1975**). "Linear prediction: A tutorial review," Proc. IEEE **63**, 561–580.
Mann, H. B., and Whitney, D. R. (**1947**). "On a test of whether one of two random variables is stochastically larger than the other," Ann. Math. Stat. **18**, 50–60.
McAulay, R. J., and Quatieri, T. F. (**1995**). "Sinusoidal coding," in *Speech Coding and Synthesis*, edited by W. B. Kleijn, and K. K. Paliwal (Elsevier Science B.V., The Netherlands), pp. 121–173.
Mizuno, H., and Abe, M. (**1995**). "Voice conversion algorithm based on piecewise linear conversion rules of formant frequency and spectrum tilt," Speech Commun. **16**, 153–164.
Moulines, E., and Charpentier, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun. **9**, 453–467.
Moulines, E., and Sagisaka, Y. (**1995**). "Voice conversion: State of the art and perspectives," Speech Commun. **16**, 125–126.
Moulines, E., and Verhelst, W. (**1995**). "Time-domain and frequency-domain techniques for prosodic modification of speech," in *Speech Coding and Synthesis*, edited by W. B. Kleijn and K. K. Paliwal (Elsevier Science B.V., The Netherlands), pp. 519–555.
Narendranath, M., Murthy, H. M., Rajendran, S., and Yegnanarayana, B. (**1995**). "Transformation of formants for voice conversion using artificial neural networks," Speech Commun. **16**, 207–216.
Operations Manual (**1993**). Multi-dimensional (MDVP) model 4305, pp. 93–131, Kay Elemetrics Corp.
Pacheco, F. S., and Seara, R. (**2002**). "Prosodic speech modification using RELP," in Proceedings of the IEEE International Telecommunications Symposium (ITS) 2002.
Quatieri, T. F., and McAulay, R. J. (**1992**). "Shape invariant time-scale and pitch modification of speech," IEEE Trans. Signal Process. **40**, 497–510.
Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ).
Shikano, K., Nakamura, S., and Abe, M. (**1991**). "Speaker adaptation and voice conversion by codebook mapping," in Proceedings of the IEEE International Symposium on Circuits and Systems 1991, Vol. **1**, pp. 594–597.
Stylianou, Y., Cappe, O., and Moulines, E. (**1998**). "Continuous probabilistic transform for voice conversion," IEEE Trans. Speech Audio Process. **6**, 131–142.
Turajlic, E., Rentzos, D., Vaseghi, S., and Ho, C.-H. (**2003**). "Evaluation of methods for parametric formant transformation in voice conversion," Proceedings of the IEEE ICASSP, Vol. **1**, pp. 724–727.
Turk, O. (**2003**). "New methods for voice conversion," MS thesis, Bogazici University, Instabul, Turkey.
Turk, O., and Arslan, L. M. (**2005**). "Donor selection for voice conversion," in Proceedings of the 13th European Signal Processing Conference (EUSIPCO 2005), Antalya, Turkey, pp. 65–68.
Turk, O., and Arslan, L. M. (**2006**). "Robust processing techniques for voice conversion," Comput. Speech Lang. **20**, 441–467.
Wilcoxon, F. (**1945**). "Individual comparisons by ranking methods," Biometrics Bulletin **1**, 80–83.
Wild, C. J., and Seber, G. A. F. (**1999**). *Chance Encounters: A First Course in Data Analysis and Inference* (Wiley, New York).
Witte, R. S., and Witte, J. S. (**2006**). *Statistics* (Wiley, Hoboken, NJ).
Woodland, P. C., Odell, J. J., Valtchev, V., and Young, S. J. (**1994**). "Large vocabulary continuous speech recognition using HTK," in Proceedings of the IEEE ICASSP 1994, pp. 125–128.
Zhang, W., Shen, L. Q., and Tang, D. (**2001**). "Voice conversion based on acoustic feature transformation," in Proceedings of the Sixth National Conference on Man-Machine Speech Communications.

Detection of time-varying harmonic amplitude alterations due to spectral interpolations between musical instrument tones

Andrew B. Horner[a)]
Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

James W. Beauchamp[b)]
School and Music and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Richard H. Y. So[c)]
Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

(Received 18 April 2008; revised 16 October 2008; accepted 22 October 2008)

Gradated spectral interpolations between musical instrument tone pairs were used to investigate discrimination as a function of time-averaged spectral difference. All possible nonidentical pairs taken from a collection of eight musical instrument sounds consisting of bassoon, clarinet, flute, horn, oboe, saxophone, trumpet, and violin were tested. For each pair, several tones were generated with different balances between the primary and secondary instruments, where the balance was fixed across the duration of each tone. Among primary instruments it was found that changes to horn and clarinet timbres were most easily discriminable, while changes to saxophone and trumpet timbres were least discriminable. Among secondary instruments, the clarinet had the strongest effect on discrimination, whereas the bassoon had the least effect. For primary instruments, strong negative correlations were found between discrimination and their spectral incoherences, suggesting that the presence of dynamic spectral variations tends to increase the difficulty of detecting time-varying alterations such as spectral interpolation.
 [DOI: 10.1121/1.3025916]

PACS number(s): 43.75.Cd, 43.66.Jh [DD] Pages: 492–502

I. INTRODUCTION

Musical instruments are recognizable even when the sound has been substantially altered by a spectrum equalizer or room acoustics. Previous work has shown that the detection of time-invariant spectral alterations in musical instrument tones is more difficult when the original tone has significant time-varying spectral fluctuations, while spectral "jaggedness" has relatively little effect (Horner *et al.*, 2004). Do the same results hold for time-variant spectral alterations such as spectral interpolation? Answering this question is the primary objective of this paper.

In small amounts, time-invariant spectral alterations can change a clarinet sound into other clarinetlike sounds. In larger amounts, they can change a clarinet sound into unknown synthetic sounds. But, due to their time-invariance, they cannot produce all possible clarinet sounds, or sounds from other instruments, such as the violin. An alternative alteration method that includes time-varying changes is spectral interpolation. Spectral interpolation combines two or more sounds to create a new sound with an intermediate spectrum. For example, a spectrally interpolated sound could be formed by resynthesizing a spectral mix consisting of 60% of a viola spectrum (primary) and 40% of a clarinet spectrum (secondary).

Some related previous studies have considered the effect of simple spectral alterations of static spectra, speech, and audio signals. Plomp (1970) considered the correspondence between an error metric and discrimination databased on static musical instrument and vowel spectra. He concluded that differences in timbre can be predicted well from spectral differences. Toole and Olive (1988) explored the effect of adding a single resonance to the spectra of noise, music recordings, and speech. They found that resonances are more easily heard in noise input signals than in speech and music input signals. Watkins (1991) and Watkins and Makin (1996) investigated the effect of spectral-envelope distortion on vowel sounds in speech, and found that the perception of a sound can be influenced by the sound that precedes it. This suggests that spectral variations might also have a significant influence on discrimination.

A related area of research is *spectral profile analysis* (Green, 1988), which studies the ability of listeners to discriminate an original stimulus from a spectrally altered version of the stimulus. However, there are significant differences between profile analysis and the approach of the current study. Most importantly, spectral profile analysis usu-

[a)]Electronic mail: horner@cs.ust.hk
[b)]Electronic mail: jwbeauch@uiuc.edu
[c)]Electronic mail: rhyso@ust.hk

ally only considers static spectra while the current study is concerned with time-varying alterations in dynamic spectra. In addition, spectral profile analysis has typically used log-spaced rather than harmonic-spaced components, which Versfeld and Houtsma (1991) found to produce very different results. Another difference is that profile analysis spectra are usually flat unlike acoustic instrument spectra, though some researchers have also used "perturbed" or "jagged" spectra (Kidd *et al.*, 1991) and noted that spectral jaggedness increases the threshold of detection (Lentz and Richards, 1998). Yet another basic difference is that profile analysis studies have usually only attempted to determine the threshold of discriminating a change in a single spectral component, or at most a few components. Bernstein *et al.* (1987) observed that thresholds for single-component changes cannot be used to accurately predict thresholds for multiple-component changes.

It is difficult to generalize profile analysis results to dynamic harmonic spectra, although the above results suggest that multiple-component changes to acoustic spectra are more difficult to detect than are single-component changes. Also, comodulation-masking-release results (Mendoza *et al.*, 1996) suggest that alterations to coherent spectra are easier to detect than are alterations to incoherent spectra. This suggests that time-varying alterations may be still more difficult to detect.

The most relevant previous work studied time-invariant (static) alteration of musical instrument spectra, where each harmonic amplitude was multiplied by a time-invariant random scalar (Horner *et al.*, 2004). This work found that listeners had more difficulty discriminating alterations to instrument sounds containing more pronounced spectral variations. Spectral incoherence (SI) and normalized centroid deviation (NCD) were both found to have strong negative correlations with discrimination scores. This suggested that dynamic spectral variations increase the difficulty of detecting spectral alterations. However, the same study found relatively low correlation for spectral irregularity (SIR), a measure of the jaggedness of a spectrum.

In a recent related study, Gunawan and Sen (2008) tested discrimination thresholds for perturbing musical instrument spectral envelopes by attenuating bands within them, as a function of center frequency and bandwidth. Perturbing center frequencies varied from 689 to 19 294 Hz, while bandwidths ranged from 1378 to 11 025 Hz, depending on the center frequency. Using tones of three musical instruments [each performed at E_4^b (~311.1 Hz)], they found that spectral sensitivity was governed by only the first few harmonics and sensitivity did not improve when extending the bandwidth any higher. On the other hand, sensitivity was found to decrease if changes were made only to the higher frequencies and continued to decrease as the bandwidth was widened. Note that this study, like the study of Horner *et al.* (2004), was based on static processing of the time-varying spectral envelope.

While the studies mentioned above have considered the effect of simple spectral alterations, relatively little work has addressed the effect of time-variant spectral alterations such as those created by spectral interpolation of dynamic musical instrument spectra. Grey (1975) studied the effect of instrument time-variant interpolation in his Ph.D. thesis, cross-fading pairs of instrument tones to create new hybrid tones. For each instrument pair 11 tones were presented to the listener with interpolation levels increasing in steps of 10% going from the primary instrument (0%) to the secondary instrument (100%). The interpolation level was fixed across the duration of each tone. The listener was told to identify the point at which they identified the initial appearance of the secondary instrument in the sound. Results indicated that listeners tended to delay the perception of the secondary instrument well beyond the 50% interpolation point, thus demonstrating a hysteresis effect in timbre perception.

Related to spectral interpolation is the issue of *blend* (Kendall and Carterette, 1993; Sandell, 1995), which is defined as whether the instruments fuse into a single composite timbre, segregate into distinct timbral entities, or fall somewhere in between the two extremes. Sandell (1995) investigated spectral centroid and other factors in determining blended instrument pairings and found that lower average values of both centroid and onset duration for a pair of tones led to increased blends, as did closeness in value for the two factors.

This study will investigate the listeners' ability to discriminate changes to the time-varying spectral amplitudes of musical sounds caused by different degrees of spectral interpolation. Are measures of spectral variation (SI and NCD) strongly correlated with discrimination in spectrally interpolated tones? Does spectral jaggedness (irregularity) have more of an effect in interpolated tones? We will address these questions. We will also determine which instruments are least and most affected by spectral interpolation. Conversely, we will determine which instruments have the least and most effect as secondary instruments in the interpolation pair.

Section II presents the techniques used for analysis and synthesis of the stimuli, followed by a discussion of the discrimination experiment in Sec. III. Section IV reviews various spectral correlate measures, such as spectral incoherence and irregularity. Finally, in Sec. V, we present the discrimination results in terms of the effect of instrument, interpolation level, and spectral variations.

II. STIMULUS PREPARATION

Eight sustained musical instrument sounds were selected as prototype signals for stimulus preparation. Sounds of a bassoon, clarinet, flute, horn, oboe, saxophone, trumpet, and violin, performed at E_4^b (~311.1 Hz), were used to represent the several Western wind and the bowed string instrument families. Five of the sounds were taken from the McGill University Master Samples recordings (CD version), two from the Prosonus Sound Library (bassoon and oboe), and one (trumpet) was recorded at the University of Illinois Urbana-Champaign School of Music. All sounds were recorded at 16 bits, 44.1 kHz. Except for the bassoon and horn, these sounds were also used by McAdams *et al.* (1999), who gave more details about their characteristics. All eight sounds were also used by Horner *et al.* (2004). After parameter equalization (see Sec. II A), each sound was used as a stimu-

lus, including attack, sustain, and decay segments of the sounds. Sounds were chosen to be representative of each instrument's timbre, in that typical expressive elements were included.

The sounds were first subjected to time-variant spectrum analysis using a computer-based phase-vocoder method (Beauchamp, 2007). This phase vocoder is different from most in that it allows a fixed analysis frequency (311.1 Hz) to be tuned to the estimated fundamental frequency of the input signal. Beauchamp (2007) and Horner *et al.* (2004) gave further details on the phase-vocoder analysis method. Briefly, the method uses a fast Fourier transform (FFT) with Hamming window whose duration (~6.4 ms) is set to approximately twice the period of the input signal in order to minimize leakage between harmonics. The FFT bin frequencies are thus integer multiples of approximately 155.5 Hz. However, only the even multiples, corresponding to the input signal's harmonics, are retained. Prior to FFT analysis, the signal is up-sampled in order to create a power-of-2 number of samples for the window.

A. Temporal and loudness equalization and frequency flattening

Sound duration is a potential factor in discrimination. For example, a sound lasting 5 s might well be easier to discriminate than a sound lasting 0.3 s. In order that duration would not be a factor in the study, the sounds were standardized to a 2 s duration by interpolating the analysis data. Briefly, this was done by first identifying the attack and decay portions of each sound and then cross-fading the beginning of interior of the sound (using a cubic spline cross-fade function) with a later portion of the sound so as to reduce the duration to 2 s. This process was performed on the analysis data prior to resynthesis. Details are given by McAdams *et al.* (1999).

Attack and decay duration are also potential factors in discrimination. For example, if a tone with a short attack is interpolated with a tone with a long attack, the listener may be able to detect a difference compared to the original tone simply based on the longer attack time. In order that attack and decay times would not be factors in the study, the sounds were standardized to 0.05 s attack times and 0.15 s decay times through interpolation of the analysis data during these segments. Most attack and decay times were very close to the standardization values. The authors noted no significant perceptual differences between the standardized tones and the original 2 s tones. Where minor audible differences did occur, the standardized tone was judged as realistic as the original.

Next, amplitude multipliers were determined by a loudness program (Moore *et al.*, 1997) in order that each sound had a loudness of 87.4 phons. An iterative procedure adjusted the amplitude multiplier starting from a value of 1.0 until the resulting phons were within 0.1 phons of 87.4 [which corresponded to the trumpet sound played through headphones at 78 dB sound pressure level (SPL)].

In addition, frequency variations and inharmonicity were eliminated from the sounds in order that they would not be factors in this study. This was done by setting each harmonic's frequency equal to the product of its harmonic number and the fixed analysis frequency (311.1 Hz) in the analysis data prior to resynthesis, resulting in flat, equally spaced frequency versus time envelopes.

Sounds produced by the equalization and frequency flattening methods described above are referred to as *reference sounds* from here on in this paper.

B. Spectral interpolation and resynthesis

Spectral interpolation was performed on the analysis data of each pair of instruments, after which the sounds were regenerated by additive synthesis. Interpolation was accomplished in the frequency domain by calculating the weighted sum of the instrument pairs' harmonic amplitudes as follows:

$$A'_k(t_n) = (1 - l)A_{p,k}(t_n) + lA_{s,k}(t_n), \quad (1)$$

where $A'_k(t_n)$, $A_{p,k}(t_n)$, and $A_{s,k}(t_n)$ are the kth harmonic (linear) amplitudes of the interpolated, primary, and secondary instrument tones at time t_n, t_n is the time corresponding to the nth frame (i.e., $t_n = n\Delta t$ and $\Delta t = 6.4$ ms), and $0 < \ell \leq 0.5$ is the interpolation level. For example, if we wanted an interpolation of 60% viola with 40% clarinet, the viola would be the primary instrument, the clarinet would be the secondary instrument, and the interpolation level ℓ would be 0.4. The interpolation level remains fixed across the duration of the tone. It should be noted that since the instantaneous phases and frequencies corresponding to the primary and secondary instruments are equal, this method is equivalent to *mixing* frequency-flattened versions of the equalized tones in the time domain.

Interpolated sounds were synthesized for each pair of the eight instruments for interpolation levels of 5%, 10%, 15%, 20%, 25%, 30%, 40%, and 50%, yielding a total of 56 interpolated sounds for each primary instrument. These increments were chosen to provide adequate resolution of the interpolation process. However, going beyond 50% is not necessary because doing so reverses the role of primary and secondary instruments.

Both the reference sounds and the interpolated sounds were resynthesized by additive synthesis with strictly fixed harmonic frequencies before being compared in the listening experiment.

III. EXPERIMENTAL METHOD

A. Subjects

Twenty listeners participated in the experiment. They were undergraduate students at the Hong Kong University of Science and Technology (HKUST), ranging in age from 19 to 23 years, who reported no hearing problems. The listeners were paid to compensate for their time spent in the experiment.

B. Design of experiment

The experiment uses an unbalanced factorial design with three independent factors: primary instrument (eight levels: bassoon, clarinet, flute, horn, oboe, saxophone, trumpet, and

violin), secondary instrument (eight levels: bassoon, clarinet, flute, horn, oboe, saxophone, trumpet, and violin), and interpolation level (eight levels: 5%, 10%, 15%, 20%, 25%, 30%, 40%, and 50%). The participant is the random factor (20 levels), and the data collected from the 20 participants are used as repeated measures. In a full factorial design, the primary and secondary instruments can be the same. However, this may cause confusion because the resulting interpolated sound will then be the same as the uninterpolated reference sound, regardless of the interpolation level. To avoid this confusion, the authors decided not to present conditions where the primary and secondary instruments are the same. In summary, each participant will take part in 448 ($8\times7\times8=448$) conditions (or 8 blocks of 56 conditions). An unbalanced general linear model is used in the analysis of variance (ANOVA) analyses and type III sums of squares are used to calculate the *F*-values (see Sec. V).

C. Test procedure

Within each of the 448 conditions, a two-alternative forced-choice discrimination paradigm was used. The listener heard two pairs of sounds and chose which pair was "different." Each trial structure was one of *AA-AB*, *AB-AA*, *AA-BA*, or *BA-AA*, where *A* represents the reference sound and *B* represents one of the 56 interpolated sounds. (Note that the reference sound is always the primary sound.) This paradigm had the advantage of not being as susceptible to variations in listeners' criteria across experimental trials compared to the simpler *A-B* method. All four combinations were presented for each interpolated sound. The two 2 s sounds of each pair were separated by a 500 ms silence, and the two pairs were separated by a 1 s silence. For each trial, the user was prompted with "which pair is different, 1 or 2?" and the response was given by using a keyboard. The computer would not accept a response until at least the first pair had been played. A custom program written at HKUST ran on an Intel PC to control the experiment.

For each instrument, a block of 224 trials was presented to each of the listeners (four trial structures $\times$ 56 interpolated sounds). The order of presentation of these 224 trials was randomized. For each interpolation, discrimination performance was averaged using the results of the four trials for each listener. Because these four trials were presented in random order within the 224 trials, the effects of possible learning were averaged out. The same trials were presented to each listener, although in a different random order. The duration of each block was about 45 min, and listeners took 5–10 min breaks after each block. Eight blocks were presented to each listener, corresponding to the eight instruments. The order of presentation of the instruments was also randomized for each listener. Sessions of four blocks were scheduled on two different days. The average time to complete a block was 50 min (including breaks). The total duration of the experiment was about 6.5 h.

Listeners were seated in a "quiet room" with less than 40 dB SPL background noise level (mostly due to computers and air conditioning). Sound signals were converted to analog by a SoundBlaster Audigy soundcard and then presented through Sony MDR-7506 headphones at 87.4 phons. The Audigy DAC utilized 24 bits with a maximum sampling rate of 96 kHz and a 100 dB signal-to-noise ratio. The sounds were actually played at 44.1 kHz.

At the beginning of the experiment, the listener read the instructions and asked any necessary questions of the experimenter. Five test trials (chosen at random from the instruments) were presented before the data trials for each instrument.

IV. SPECTRAL CORRELATE MEASURES

Spectral centroid has been shown to be strongly correlated with one of the most prominent dimensions of timbre as derived by multidimensional scaling (MDS) experiments (Grey and Gordon, 1978; Wessel, 1979; Krumhansl, 1989; Iverson and Krumhansl, 1993; Krimphoff *et al.*, 1994; Kendall and Carterette, 1996; Lakatos, 2000). The time-varying normalized spectral centroid based on a sound's harmonic amplitudes, $\{A_k(t_n)\}$, is defined as

$$\mathrm{NSC}(t_n)=\frac{\Sigma_{k=1}^{K} kA_k(t_n)}{\Sigma_{k=1}^{K} A_k(t_n)}, \tag{2}$$

where K is the number of harmonics and $\mathrm{NSC}(t_n)$ can be thought of as the amplitude-averaged harmonic number at time t_n. It is "normalized" because spectral centroid is frequently given in frequency units, and, in this case, frequency has been normalized out of the definition by division by the fundamental frequency (311.1 Hz). Discrimination of spectral alterations may be affected by some aspect of the spectral variations of the reference sounds. Several different measures of these variations are possible. Grey (1977) referred to "spectral fluctuation" as an interpretation of a dimension revealed by MDS of musical instrument dissimilarity judgments. "Spectral flux" was qualitatively described by Krumhansl (1989) as "the temporal evolution of the spectral components" and by McAdams *et al.* (1999) as "the change in shape of a spectral envelope over time."

Krimphoff (1993) defined "flux" as the root-mean-squared deviation of the normalized spectral centroid over time. We call it NCD, which is given by

$$\mathrm{NCD}=\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}(\mathrm{NSC}(t_n)-\mathrm{NSC}_{xx})^2}, \tag{3}$$

where NSC_{xx} could be the time-averaged, rms, or maximum value of NSC. Time-averaged measurements of NCD were used in the current study.

Beauchamp and Lakatos (2002) measured spectral fluctuation in terms of SI, a measure of how much a spectrum differs from a coherent version of itself. Larger incoherence values indicate a more dynamic spectrum, and smaller values indicate a more static spectrum. A perfectly static spectrum has an incoherence of zero.

We define the coherent version of a time-varying spectrum to be one that has the same average spectrum and the same instantaneous rms amplitude as the reference sound, but unlike the reference, all harmonic amplitudes vary in time proportional to the rms amplitude and, therefore, in

TABLE I. ANOVA table illustrating the main effects and two-way interactions of primary instrument (eight primary instruments), secondary instrument (eight secondary instruments), and interpolation level (eight levels: 5%, 10%, 15%, 20%, 25%, 30%, 40%, and 50%) on the data collected from 20 listeners participating in the discrimination experiment. Data are the percentage of correct discrimination scores (100%, 75%, 50%, 25%, and 0%) over each group of four trials containing the same reference and altered sounds. For each interpolation level, there were 28 trials (four repetitions with seven different secondary instrument interpolations in the test sounds). Because the primary and secondary instruments are never the same, an unbalanced ANOVA was conducted. The *p*-values generated from ANOVAs for all the main effects and two-way interactions are the same when using type I sum of squares and type III sum of squares. The ANOVAs using type III sum of squares are shown (Johnson and Wichern, 1992).

Source	DF	Sum of squares	Mean square	*F*-value	Pr>*F*
Primary (P) instrument	7	19.75	2.82	73.51	<0.0001 [a]
Secondary (S) instrument	7	13.83	1.97	51.49	<0.0001 [a]
Interpolation level	7	102.75	14.68	382.49	<0.0001 [a]
P-instrument and S-instrument	41	12.52	0.31	7.96	<0.0001 [a]
P-instrument and interpolation level	49	5.31	0.11	2.83	<0.0001 [a]
S-instrument and interpolation level	49	6.04	0.12	3.21	<0.0001 [a]
Measurement error	8799	337.66	0.04		
Corrected total	8959	398.40			

[a]*p*-values are consistent with the results of nonparametric statistical tests.

fixed ratios to each other. Put another way, the coherent spectrum's harmonic amplitudes normalized by the rms amplitude are fixed.

The coherent version of the *k*th harmonic amplitude is defined by

$$\hat{A}_k(t_n) = \frac{\bar{A}_k A_{\text{rms}}(t_n)}{\sqrt{\Sigma_{k=1}^{K} \bar{A}_k^2}}, \tag{4}$$

where $\bar{A}_k$ is the time-averaged amplitude of the *k*th harmonic and

$$A_{\text{rms}}(t_n) = \sqrt{\sum_{k=1}^{K} A_k^2(t_n)} \tag{5}$$

is the instantaneous rms amplitude.

Then, the SI of the reference spectrum is defined as

$$\text{SI} = \left(\frac{\Sigma_{n=0}^{N-1} \Sigma_{k=1}^{K} (A_k(t_n) - \hat{A}_k(t_n))^2}{\Sigma_{n=0}^{N-1} (A_{\text{rms}}(t_n))^2} \right)^{1/2}, \tag{6}$$

where *N* is the sound's total number of frames.

With the definitions of Eqs. (4) and (6), SI varies between 0 and 1 with higher values indicating more incoherence (more dynamic). SI is a measure of spectral fluctuation, while NCD is a measure of the normalized centroid change over time. Since it is possible for NCD to be significant while SI is relatively small and for SI to be large while NCD is small, SI and NCD are approximately independent measures, although they may be correlated for a particular set of musical sounds.

Krimphoff (1993) also introduced the concept of SIR, which was defined by Beauchamp and Lakatos (2002) as

$$\text{SIR} = \frac{1}{N} \sum_{n=0}^{N-1} \times \frac{\Sigma_{k=2}^{K-1} A_k(t_n) |A_k(t_n) - (A_{k-1}(t_n) + A_k(t_n) + A_{k-1}(t_n))/3|}{A_{\text{rms}}(t_n) \Sigma_{k=2}^{K-1} A_k(t_n)}. \tag{7}$$

This formula defines the average difference between a spectrum and a spectrally smoothed version of it. First, for each frame, it is amplitude-averaged over harmonics, then it is normalized by rms amplitude, and finally it is time-averaged over all of the frames.

V. RESULTS

This section analyzes the discrimination data with respect to interpolation level and *relative-amplitude spectral error* (Horner *et al.*, 2004, 2006). The latter error measure (also known as *relative spectral error*) is necessary since interpolations between spectrally similar instruments such as the violin and viola are much more difficult to detect than interpolations between very different instruments such as the violin and bassoon. Therefore, interpolation level is not a good independent predictor of discrimination, and an alternative predictor such as relative-amplitude spectral error is needed to explain the results. This error is given by the formula

$$\varepsilon = \frac{1}{N} \sum_{n=0}^{N-1} |A_k(t_n) - A'_k(t_n)|, \tag{8}$$

where $A_k(t_n)$ and $A'_k(t_n)$ are the *k*th harmonic amplitudes of the reference and interpolated tones, respectively.

A. Data analysis methods

A test of normality indicated that the discrimination data were not normally distributed even after various transforma-

TABLE II. The effects of primary instrument on the ability of listeners to discriminate interpolated tones with nearest-neighbor error levels (5%, 10%, 15%, 20%, 25%, 30%, 40%, and 50%). Within each column, primary instruments with the lowest discrimination scores are listed at the top, and those with the highest discrimination scores are listed at the bottom. The ranking is based on the results of Student–Newman–Keuls tests, and is consistent with the results of nonparametric tests. When two or more primary instruments produced similar discrimination scores (i.e., not significantly different at the $p=0.05$ level), they are grouped and labeled with the same capital letter (A–D). Only the group with the smallest (labeled as A) and the largest discrimination scores (labeled as either B or C or D) are shown.

	Secondary instruments															
	Bs		Cl		Hn		Sx		Fl		Ob		Tp		Vn	
Primary instruments	Sx	A	Sx	A	Sx	A	Vn	A	Tp	A	Tp	A	Sx	A	Sx	A
	Tp		Ob	A	Bs	A	Fl	A	Sx	A	Vn	A	Vn	A	Cl	A
	Fl		Fl		Tp	A	Tp	A	Bs	A	Sx	A	Ob		Fl	A
	Vn		Vn		Vn		Cl	B	Ob		Fl		Cl		Tp	
	Ob		Tp		Ob		Bs	B	Vn		Cl		Fl		Ob	
	Cl	D	Bs	D	Fl		Ob	B	Hn		Bs		Bs		Bs	D
	Hn	D	Hn	D	Cl	C	Hn	B	Cl	D	Hn	D	Hn	D	Hn	D

tions (e.g., sqrt[arcsin(data)]) (Stevens, 2002). Therefore, the data were analyzed using both parametric and nonparametric statistical tests (parametric: ANOVA, Student–Newman–Keuls tests; nonparametric: Friedman ANOVA by rank). The results of parametric and nonparametric tests were found to be consistent. In the rest of the paper, the main effects and the two-way interactions were tested with both ANOVA and nonparametric tests.

B. Effects and interactions of interpolation level and instrument

Discrimination scores, given in terms of percent correct, were computed for each interpolation across the four trial structures for each listener. Because the presentation order of each of the four trials was randomized, any potential effects of learning were averaged out. Table I shows the results of the ANOVA studying the main effects of interpolation level, primary instrument, secondary instrument, and their two-way interactions effects. Inspection of Table I indicates that all main effects and interactions are significant ($p<0.001$). These significant results are also consistent with the results of nonparametric analyses (Kruskal–Wallis H tests, Mann–Whitney U, and Friedman two-way ANOVA were tested).

The interactions among the effects of primary instrument and secondary instruments have been studied using Friedman ANOVAs, and the results are shown in Tables II and III.

Table II shows that the saxophone (Sx) as a primary instrument resulted in the significantly lowest discrimination scores when mixed with all other instruments. This means that when other instruments are mixed with the saxophone, listeners have trouble detecting any difference. On the other hand, the horn (Hn) as a primary instrument resulted in the significantly highest discrimination scores when mixed with the other instruments (the only exception was with the flute). This suggests that when other tones are mixed with the horn, it is relatively easy to detect the difference.

Table III shows that the bassoon (Bs) as a secondary instrument resulted in the significantly lowest discrimination scores when mixed with the other instruments (the only exception was with the clarinet). On the other hand, the clarinet (Cl) as a secondary instrument resulted in the significantly highest discrimination scores when mixed with the other instruments (the only exception was with the oboe). This suggests that the bassoon easily blends with other instrument tones, while the clarinet tends to stand out.

TABLE III. The effects of secondary instrument on the ability of listeners to discriminate interpolated tones with nearest-neighbor error levels (5%, 10%, 15%, 20%, 25%, 30%, 40%, and 50%). Within each column, secondary instruments with the lowest discrimination scores are listed at the top, and those with the highest discrimination scores are listed at the bottom. The ranking is based on the results of Student–Newman–Keuls tests and is consistent with the results of nonparametric tests. When two or more secondary instruments produced similar discrimination scores (i.e., not significantly different at the $p=0.05$ level), they are grouped and labeled with the same capital letter (A–E). Only the group with the smallest (labeled as A) and the largest discrimination scores (labeled at either C or D or E) are shown.

	Primary instruments															
	Bs		Cl		Hn		Sx		Fl		Ob		Tp		Vn	
Secondary instruments	Fl	A	Vn	A	Bs	A	Bs	A	Bs	A	Bs	A	Bs	A	Tp	A
	Hn	A	Ob	A	Fl	A	Tp		Ob		Tp		Ob	A	Ob	A
	Ob	A	Tp		Ob	A	Hn		Vn		Vn		Fl		Bs	A
	Vn		Bs		Vn		Vn		Sx		Cl		Hn		Sx	
	Tp	C	Sx		Sx		Ob		Tp	D	Fl		Vn		Hn	C
	Sx	C	Fl	E	Tp		Fl		Cl	D	Hn		Sx		Fl	C
	Cl	C	Hn	E	Cl	D	Cl	C	Hn	D	Sx	C	Cl	D	Cl	C

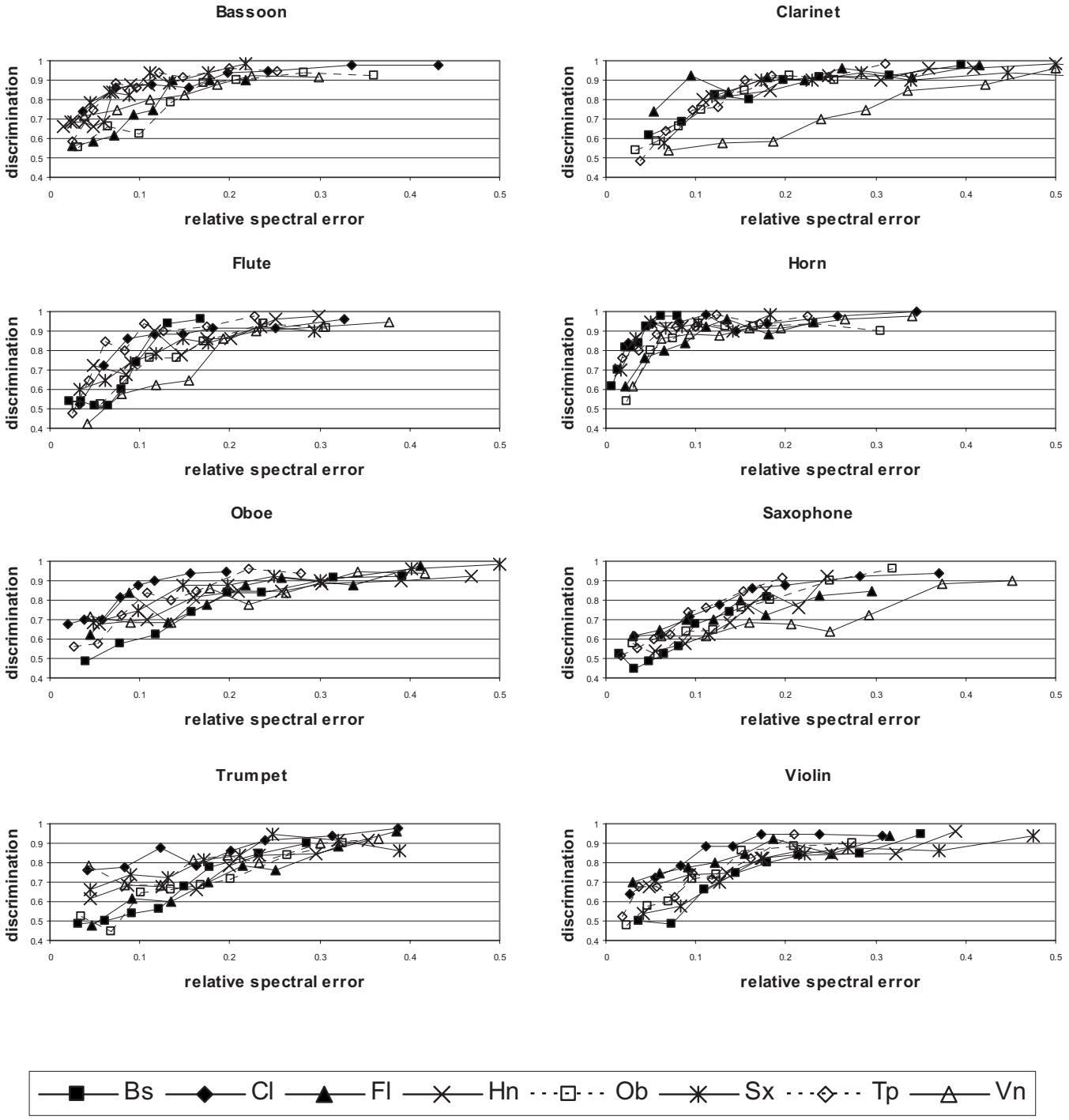


FIG. 1. Mean discrimination scores for the eight primary instruments. Each line of points corresponds to interpolations with one of the other seven secondary instruments (some outlying lines are labeled).

C. Use of relative-amplitude spectral error to explain the results

Although the interpolation levels have significant effects on discrimination, these effects have significant interactions both with the effects of primary instrument and secondary instrument (see Tables II and III). Intuitively, this is easy to understand since interpolations between spectrally similar instruments such as the violin and viola are much more difficult to detect than interpolations between very different instruments such as the violin and saxophone. These differences are better reflected in spectral error metrics such as relative-amplitude spectral error [see Eq. (8)], which is a more "universal" measure than spectral interpolation because it works for time-variant cases other than spectral interpolation. Recent previous work has shown that relative-amplitude spectral error captures about 90% of the variance in discrimination data for instruments similar to those presented in this paper (Horner *et al.*, 2006).

Therefore, relative spectral errors have been calculated for each combination of primary instrument, secondary instrument, and interpolation level. Tests of correlation indicate that the calculated relative spectral errors are significantly

correlated with interpolation level ($p<0.001$, Spearman correlation test) as well as discrimination ($p<0.001$, Spearman). Results of regression analyses indicate that $\log_{10}$ (relative error) can explain 48% of the data variance in the mean discrimination scores of the 20 listeners ($R^2=0.489$). Because relative spectral error accounts for the relative difference between two interpolated tones, it is a better predictor variable than interpolation level. In the subsequent analyses, discrimination scores will be examined as a function of relative spectral error.

For all instruments, larger interpolation levels always result in larger relative spectral error levels. It can also be trivially observed that as the relative spectral level increases, the mean discrimination increases asymptotically to between 90% and 100%. Figure 1 shows discrimination scores averaged over the 20 listeners on the eight primary instruments, where each line of points corresponds to interpolations with one of the other seven secondary instruments.

For example, the graph labeled "clarinet" shows its interpolation with the other seven instruments, and the "Vn" line corresponds to the clarinet-violin interpolations. In particular, the leftmost point on the Vn line (at 7% relative spectral error) is an interpolation with 95% clarinet and 5% violin. Moving rightward, the next point along the line (at 13% error) is an interpolation with 90% clarinet and 10% violin. The points farther to the right on this line increase the violin to 15%, 20% 25%, 30%, 40%, and 50%, and decrease the clarinet correspondingly downward.

Interestingly, the horn shows a consistently faster convergence to a ceiling value than the other primary instruments in Fig. 1 no matter which secondary instrument it is interpolated with. The slowest convergence occurs when the clarinet is the primary instrument and the violin is the secondary instrument—a pronounced outlier among the curves. The other curves in the clarinet graph are very similar (with the exception of the two leftmost flute points), with very little deviation. Perhaps this is due to the prominence of the odd harmonics in the clarinet, which results in fairly uniform discrimination changes.

Figure 2 shows a graph of average discrimination as a function of relative spectral error for each primary instrument, where points have been grouped into error bins of 5% (e.g., 5%, 10%, 15%, 20%, etc.) using nearest-neighbor grouping. Bins with only one element have been dropped from the figure. The curves vary considerably from instrument to instrument, with a range of 25% for error levels up to 15%, and a range of 15% for error levels between 20% and 30%.

Inspection of Fig. 2 indicates that the horn has significantly higher average discrimination scores than the other primary instruments ($p<0.05$, Student–Newman–Keuls tests), meaning that listeners found it relatively easy to detect interpolations to the horn, in agreement with the results in Table II. The bassoon also had significantly higher discrimination scores ($p<0.05$, Student–Newman–Keuls tests). These two instruments tend to blend or fuse with others in a smooth way, allowing other instruments to transparently shine through them quite readily. The saxophone and trumpet have significantly lower average discrimination scores than the other primary instruments ($p<0.05$, Student–Newman–Keuls tests), which basically agrees with the results in Table II. These instruments tend to be more opaque and dominate other instruments, making it more difficult to detect interpolations to them.

These observations are also in agreement with the advice given by Rimsky-Korsakov (1964) in his classic orchestration book, where relative strengths of the instruments are given for the various instruments in order to balance them when playing at the same dynamic level. While the other woodwinds are given a weight of 1, Rimsky-Korsakov (1964) assigned the saxophone a weight of 3 and the trumpet a weight of 4 reflecting their relative strength and ability to project.

Figure 3 displays the same data as Fig. 1, but it is arranged by secondary instrument. For example, in the bassoon graph, the "Hn" line represents interpolations with the bassoon as a secondary instrument and the horn as a primary instrument. The other lines in the bassoon graph represent interpolations with the other primary instruments (relative spectral error is given with respect to each of the primary instruments).

Figure 4 shows a graph of average discrimination for each secondary instrument. There is not as much variation between the instruments as in Fig. 1. The clarinet has significantly higher average discrimination scores than the other primary instruments ($p<0.05$, Student–Newman–Keuls tests), meaning that listeners found it relatively easy to detect the prominent odd harmonics of the clarinet when they were interpolated with the primary instrument. This agrees with the results in Table III. The trumpet is also significantly higher ($p<0.05$, Student–Newman–Keuls tests). On the other hand, the violin has significantly lower average discrimination scores than the other primary instruments ($p<0.05$, Student–Newman–Keuls tests), though there is considerable variation among the violin scores. This differs from the results in Table III, which found the bassoon to be the lowest with respect to interpolation level. The unusually

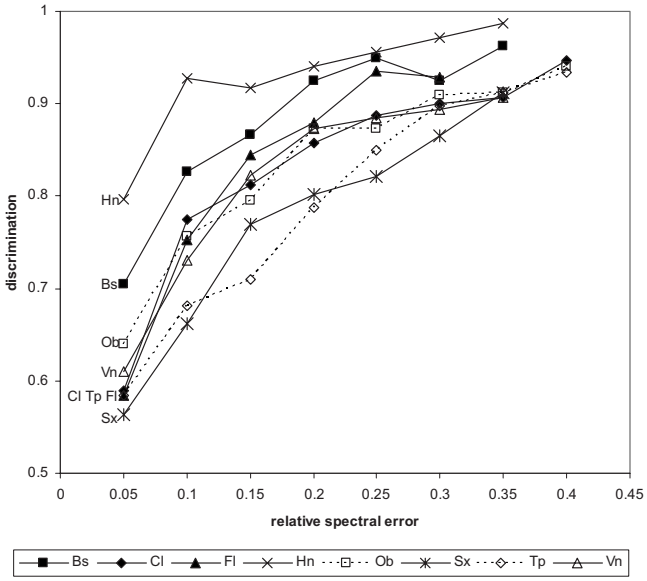


FIG. 2. Average discrimination scores for the eight primary instruments.

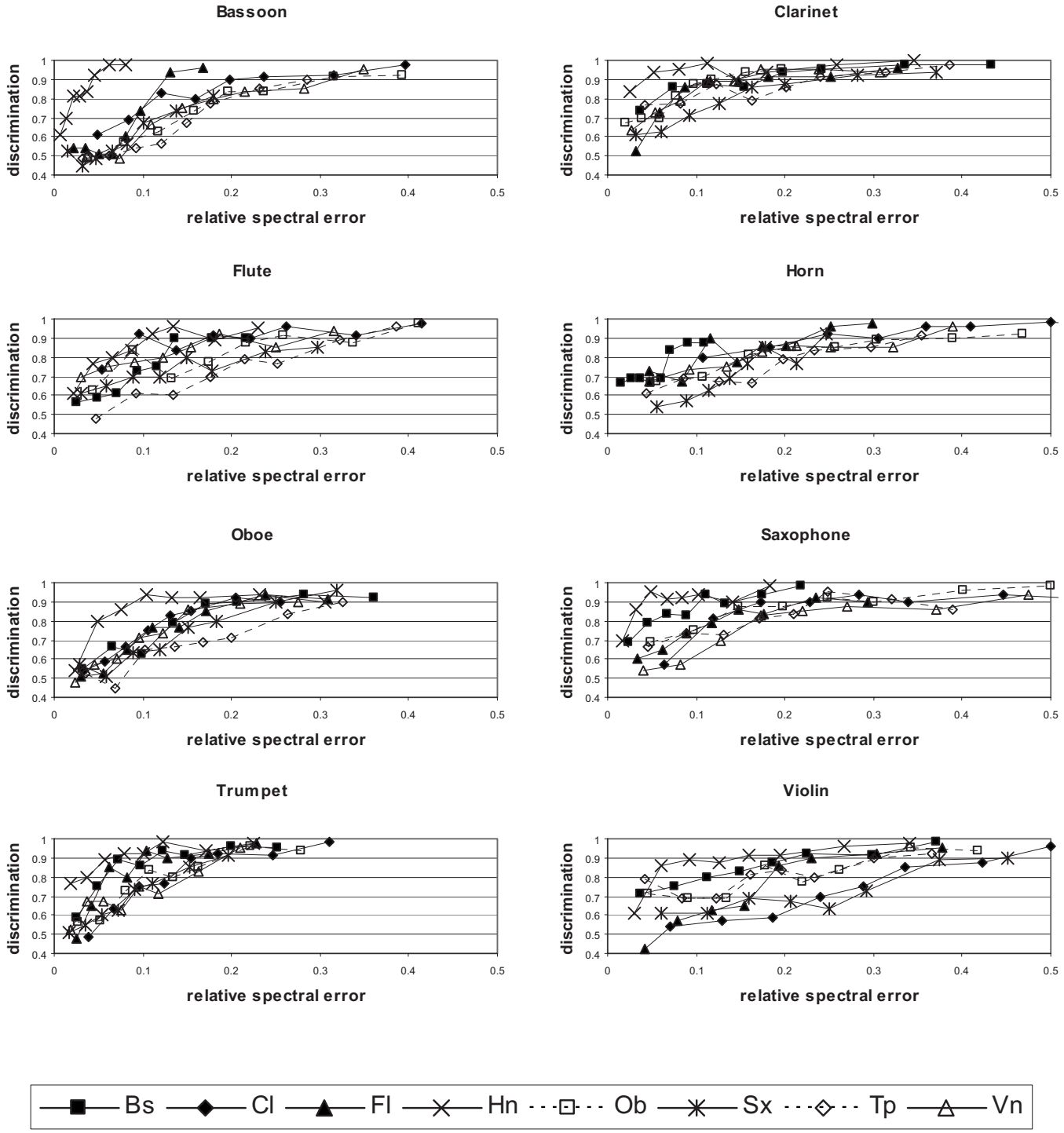


FIG. 3. Mean discrimination scores for the eight secondary instruments. Each line of points corresponds to interpolations with one of the other seven primary instruments (some outlying lines are labeled).

strong variations probably account for the difference. The saxophone also exhibits some unusual behavior, having among the highest discrimination scores for low-error levels and the lowest scores for high-error levels.

D. Correlation of spectral incoherence, normalized centroid deviation, and spectral irregularity with discrimination

Discrimination scores were correlated with physical measures SI, NCD, and SIR [see Eqs. (3), (6), and (7)] of the reference sounds to determine whether these measures of spectral variation were significantly related to discrimination. Table IV gives SI, spectral centroid deviation, SIR, and average discrimination scores for error levels of 10%, 15%, and 20% for each of the eight primary instruments. Table V gives the same values for each of the eight secondary instruments.

For primary instruments, strong negative correlations were found between the discrimination scores and SI with 10%, 15%, and 20% error levels ($r(6)=-0.76$, $p<0.05$ at 10%; $r(6)=-0.74$, $p<0.05$ at 15%; and $r(6)=-0.74$, p

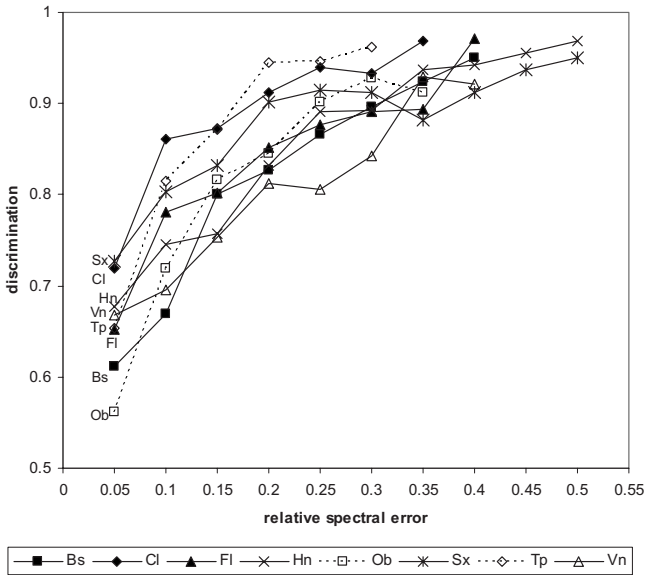


FIG. 4. Average discrimination scores for the eight secondary instruments.

<0.05 at 20%, Spearman correlation). No statistically reliable correlation relationship was found for NCD or SIR. For secondary instruments, no statistically reliable correlation relationship was found for SI, NCD, or SIR.

VI. DISCUSSION

The results presented in Sec. V show that, for the instruments tested, SI is strongly correlated with difficulty of detecting spectral interpolations. It seemed reasonable to expect that increased incoherence would make detection of a time-varying spectral change such as spectral interpolation more difficult, especially since our previous work had found that SI was strongly correlated with difficulty of detecting static random spectral alterations.

On the other hand, based on our previous work, we also expected NCD to affect discrimination, but correlations between this parameter and discrimination were found to be not significant. Though SI and centroid deviation are theoretically quite independent, in our previous work they appeared to be tightly correlated measures of time-variant spectral variation. Apparently, centroid deviations in the primary instrument have little affect on the detection of spectral interpolations, perhaps because such deviations would be somewhat averaged out by interpolation with the secondary instrument. There is no such averaging in random spectral alteration, perhaps accounting for the difference.

Both our previous and current studies found SIR to have no significant correlation to discrimination. Time-varying spectral variations, especially those of SI, seem to be more important than jaggedness of a spectrum with respect to difficulty of detecting spectral alterations and interpolations.

The results of this study can potentially lead to methods of estimating the perception of timbral difference based on spectral difference. For the spectral interpolation case, relative spectral error is by no means a perfect predictor of discrimination. R^2 correspondence was about 48% for this case as opposed to 90% found for static spectral variations (Horner *et al.*, 2006). In the future, it would be interesting to compare different measures of spectral distance in terms of their correlations with discrimination results as well as to results of listener judgments of timbral distance. Higher correspondences may be obtained from timbral distance judgments rather than discrimination, and if so, in the opinion of these authors, such judgments would yield just as meaningful, if not more meaningful results as those based on discrimination. Another way to get at the problem of spectral distance for the time-varying case would be to measure correspondences for the case of data-reduced time-varying synthesis of musical sounds. Three possibilities, where the fidelity of synthesis can be easily varied, are (1) frequency-modulation synthesis (Horner *et al.*, 1993a), (2) principal-component synthesis (Horner *et al.*, 1993b), and (3) critical-band wavetable synthesis (Beauchamp and Horner, 1995). In each of these cases, unlike the static spectral variation method, ratios between original and synthetic harmonic amplitudes vary with time.

Our investigation provides an excellent framework for timbre hybridization. In future work, it will be interesting to see which instrument spaces yield the most promising sounds. For example, it is easy to imagine that the Australian didgeridoo would produce a fascinating blend of timbres with almost any other instrument in small amounts. Mongolian throat singing could also produce some amazingly rich timbres as primary instrument, and a fascinating way of "muting" other instruments when used as the secondary instrument.

These results may also be relevant for orchestration theory, as well as for the identification of salient timbre dimensions, an important topic in music information retrieval.

TABLE IV. SI, NCD, SIR, and average 10%, 15%, and 20% error level discrimination scores for the eight primary instruments.

Instrument	Spectral incoherence	Normalized centroid deviation	Spectral irregularity	10% discrim.	15% discrim.	20% discrim.
Bs	0.075	0.400	0.093	0.826	0.866	0.925
Cl	0.085	0.700	0.174	0.775	0.812	0.857
Fl	0.118	0.600	0.129	0.752	0.845	0.88
Hn	0.057	0.200	0.073	0.927	0.917	0.94
Ob	0.069	0.800	0.137	0.756	0.795	0.873
Sx	0.101	0.600	0.195	0.662	0.769	0.801
Tp	0.184	1.600	0.039	0.681	0.710	0.787
Vn	0.193	1.400	0.131	0.730	0.822	0.873

TABLE V. SI, NCD, SIR, and average 10%, 15%, and 20% error level discrimination scores for the eight secondary instruments.

Instrument	Spectral incoherence	Normalized centroid deviation	Spectral irregularity	10% discrim.	15% discrim.	20% discrim.
Bs	0.075	0.400	0.093	0.668	0.800	0.827
Cl	0.085	0.700	0.174	0.860	0.872	0.912
Fl	0.118	0.600	0.129	0.780	0.801	0.851
Hn	0.057	0.200	0.073	0.745	0.756	0.832
Ob	0.069	0.800	0.137	0.719	0.815	0.845
Sx	0.101	0.600	0.195	0.802	0.831	0.901
Tp	0.184	1.600	0.039	0.814	0.871	0.944
Vn	0.193	1.400	0.131	0.695	0.753	0.812

ACKNOWLEDGMENTS

We wish to express our appreciation to Xavier Rodet for his suggestion to undertake this spectral interpolation detection study as a follow-up to our earlier work on spectral alteration detection. We would like to thank Cammy Zhuang, for running subjects in the listening experiment, and Simon Cheuk-Wai Wun, for writing the program for Intel personal computers used for the listening tests. This work was supported in part by the Hong Kong Research Grants Council's CERG Project Nos. 613806 and 613508.

Beauchamp, J. W., and Horner, A. (**1995**). "Wavetable interpolation synthesis based on time-variant spectral analysis of musical sounds," presented at the 98th AES Convention, Audio Engineering Society Preprint No. 3960, pp. 1–17.

Beauchamp, J. W., and Lakatos, S. (**2002**). *New Spectro-Temporal Measures of Musical Instrument Sounds Used for a Study of Timbral Similarity of Rise-Time- and Centroid-Normalized Musical Sounds*, Proceedings of the Seventh International Conference on Music Perception and Cognition (University of New South Wales, Sydney, Australia), pp. 592–595.

Beauchamp, J. W. (**2007**). in *Analysis, Synthesis, and Perception of Musical Sounds*, edited by J. W. Beauchamp (Springer, New York), pp. 1–89.

Bernstein, L. R., Richards, V. M., and Green, D. M. (**1987**). in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 6–15.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University Press, New York).

Grey, J. M. (**1975**). "An exploration of musical timbre," Ph.D. thesis, Stanford University, Stanford, CA; also available as Stanford Department of Music Report No. STAN-M-2, Stanford University, Stanford, CA.

Grey, J. M. (**1977**). "Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. **61**, 1270–1277.

Grey, J. M., and Gordon, J. W. (**1978**). "Perceptual effects of spectral modification on musical timbres," J. Acoust. Soc. Am. **63**, 1493–1500.

Gunawan, D., and Sen, D. (**2008**). "Spectral envelope sensitivity of musical instrument sounds," J. Acoust. Soc. Am. **123**, 500–506.

Horner, A., Beauchamp, J. W., and Haken, L. (**1993a**). "Machine tongues XVI: Genetic algorithms and their application to FM matching synthesis," Comput. Music J. **17**, 17–29.

Horner, A., Beauchamp, J. W., and Haken, L. (**1993b**). "Methods for multiple wavetable synthesis of musical instrument tones," J. Audio Eng. Soc. **41**, 336–356.

Horner, A., Beauchamp, J. W., and So, R. (**2004**). "Detection of random alterations to time-varying musical instrument spectra," J. Acoust. Soc. Am. **116**, 1800–1810.

Horner, A., Beauchamp, J. W., and So, R. (**2006**). "A search for best error metrics to predict discrimination of original and spectrally altered musical instrument sounds," J. Audio Eng. Soc. **54**, 140–156.

Iverson, P., and Krumhansl, C. L. (**1993**). "Isolating the dynamic attributes of musical timbre," J. Acoust. Soc. Am. **94**, 2595–2603.

Johnson, R. A., and Wichern, D. W. (**1992**). *Applied Multivariate Statistical Analysis*, International ed. (Prentice-Hall, Englewood Cliffs, NJ).

Kendall, R. A., and Carterette, E. C. (**1993**). "Identification and blend of timbres as a basis for orchestration," Contemporary Music Review **9**, 51–67.

Kendall, R. A., and Carterette, E. C. (**1996**). *Difference Thresholds for Timbre Related to Spectral Centroid*, Proceedings of the Fourth International Conference on Music, Perception and Cognition (Faculty of Music, McGill University, Montreal, Canada), pp. 91–95.

Kidd, G., Jr., Mason, C. R., Uchanski, R. M., Brantley, M. A., and Shah, P. (**1991**). "Evaluation of simple models of auditory profile analysis using random reference spectra," J. Acoust. Soc. Am. **90**, 1340–1354.

Krimphoff, J. (**1993**). "Analyse acoustique et perception du timbre (Acoustic analysis and timbre perception)," DEA thesis, Université du Maine, Le Mans, France.

Krimphoff, J., McAdams, S., and Winsberg, S. (**1994**). "Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique (Timbre characteristics for complex sounds. II. Acoustical analysis and psychophysical measurements)," J. Phys. **4**, 625–628.

Krumhansl, C. L. (**1989**). in *Structure and Perception of Electroacoustic Sounds and Music*, edited by S. Nielzén and O. Olsson (Excerpta Medica, Amsterdam), pp. 43–53.

Lakatos, S. (**2000**). "A common perceptual space for harmonic and percussive timbres," Percept. Psychophys. **62**, 1426–1439.

Lentz, J. J., and Richards, V. M. (**1998**). "The effects of amplitude perturbation and increasing numbers of components in profile analysis," J. Acoust. Soc. Am. **103**, 535–541.

McAdams, S., Beauchamp, J. W., and Meneguzzi, S. (**1999**). "Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters," J. Acoust. Soc. Am. **105**, 882–897.

Mendoza, L., Schultz, M. L., and Schulz, R. A. (**1996**). "Comodulation masking release as a function of masking noise-band temporal envelope similarity in normal hearing and cochlear impaired listeners," J. Acoust. Soc. Am. **99**, 2565.

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.

Plomp, R. (**1970**). in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden, The Netherlands), pp. 405–408.

Rimsky-Korsakov, N. (**1964**). *Principles of Orchestration* (Dover, New York), pp. 22–23, 33.

Sandell, G. J. (**1995**). ""Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration," Music Percept. **13**, 209–246.

Stevens, J. (**2002**). *Applied Multivariate Statistics for the Social Sciences*, 4th ed. (Lawrence Erlbaum Association, Hillsdale, NJ), pp. 264.

Toole, F. E., and Olive, S. E. (**1988**). "The modification of timbre by resonances: perception and measurement," J. Audio Eng. Soc. **36**, 122–142.

Versfeld, N. J., and Houtsma, A. J. M. (**1991**). "Perception of spectral changes in multi-tone complexes," Q. J. Exp. Psychol. **43A**, 459–479.

Watkins, A. J. (**1991**). "Central auditory mechanisms of perceptual compensation for spectral-envelope distortion," J. Acoust. Soc. Am. **90**, 2942–2955.

Watkins, A. J., and Makin, S. J. (**1996**). "Effects of spectral contrast on perceptual compensation for spectral-envelope distortion," J. Acoust. Soc. Am. **99**, 3749–3757.

Wessel, D. L. (**1979**). "Timbre space as a musical control structure," Comput. Music J. **3**, 45–52.

Voice classification and vocal tract of singers: A study of x-ray images and morphology

Friederike Roers
Voice Research Laboratory, University of Music Carl Maria von Weber, D-01001, Dresden, Germany

Dirk Mürbe
Voice Research Laboratory, University of Music Carl Maria von Weber, D-01001, Dresden, Germany and Department of Otorhinolaryngology, Technical University of Dresden, Dresden, Germany

Johan Sundberg[a)]
Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH, Stockholm SE 10044, Sweden

(Received 8 April 2008; revised 22 October 2008; accepted 22 October 2008)

This investigation compares vocal tract dimensions and the classification of singer voices by examining an x-ray material assembled between 1959 and 1991 of students admitted to the solo singing education at the University of Music, Dresden, Germany. A total of 132 images were available to analysis. Different classifications' values of the lengths of the total vocal tract, the pharynx, and mouth cavities as well as of the relative position of the larynx, the height of the palatal arch, and the estimated vocal fold length were analyzed statistically, and some significant differences were found. The length of the pharynx cavity seemed particularly influential on the total vocal tract length, which varied systematically with classification. Also studied were the relationships between voice classification and the body height and weight and the body mass index. The data support the hypothesis that there are consistent morphological vocal tract differences between singers of different voice classifications.
 [DOI: 10.1121/1.3026326]

PACS number(s): 43.75.Rs, 43.70.Aj, 43.70.Gr [BHS] Pages: 503–512

I. INTRODUCTION

Voice timbre characteristics are determined by the voice source and the formants. While the source is determined by the transglottal airflow, the formant frequencies are defined by the shape of the vocal tract (VT).

The dimensions of the VT are affected by two factors. One is functional, and is determined by the subject's articulatory settings and maneuvers. The other is the individual morphology of the VT, which sets limits regarding the formant frequency variability available by articulation.

Some voice quality differences between male and female adults and children exemplify the relevance of VT dimensions to formant frequencies. Peterson and Barney (1952) revealed the systematic formant frequency differences between these voice categories. Chiba and Kajiyama (1941) and Fant (1966) interpreted these differences in morphological terms. More specifically, Fant (1966) related the formant frequency differences between male and female adults to the fact that particularly the pharyngeal part of the female VT is considerably shorter than that of the male VT. The relationship between VT length and body dimensions was studied by Fitch and Giedd (1999) who used magnetic resonance imaging (MRI) data from children and adults and found significant correlation between VT length and both body height and bodyweight. Nordström (1977) used simulation for exploring the formant frequency effects of the VT difference between adult male, adult female, and infant VTs.

VT dimensions should be relevant also to the voice characteristics of singer voices and possibly even to their classification. Cleveland (1977) analyzed sustained vowels sung by professional singers. His results showed systematic formant frequency differences between tenors, baritones, and basses. He concluded that these differences were due to different dimensions of the oral and the pharyngeal cavities. Also, the center frequency of the singer's formant cluster,[1] a spectrum peak in male operatic singer voices, has been found to vary between the male voice classifications tenor, baritone, and bass (Sundberg, 2001). For instance, tenors tend to have a higher center frequency than baritones. These findings support the assumption that there are differences in VT dimensions between different voice classifications.

This assumption was further supported by x-ray data presented by Dmitriev and Kiselev (1979). They measured the VT length as the sum of three straight lines, one pharyngeal, one velar, and one oral. The results revealed a correlation between voice classification and VT length and also the center frequencies of peaks in the long term-average spectra of singing. Their method of measuring the VT length was further developed and accurately described by Sulter *et al.* (1992).

Not only VT length but also other properties of the VT, e.g., the height of the maxillary dental arch, have been assumed to be relevant to voice classification. Thus, Gutzmann

a)Author to whom correspondence should be addressed.

(1949, 1958) reported that he had noted a clear correlation, a high palate typically occurring in sopranos and a flat and wide palate in basses. In addition, he pointed out that a high palate is characteristic of singers with a "light" timbre while a "dark" timbre is mostly associated with a flat palate. However, Ackermann (1967) measured the impressions of the palate of 164 singers and failed to observe any relationship with voice classification. His results were later corroborated by Rudolph (1969) in his analysis of 72 palate impressions. Marunick and Menaldi (2000) analyzed the relationship between various measures of the maxillary dental arch and voice classification in nine female singers but failed to find any significant covariation between classification and the height of the arch.

The above suggests that singers of different genders and voice classifications differ with regard to VT morphology. The classification of singers is quite important. It is a common experience among voice teachers and singers that a mistake in the classification may even lead to voice disorders. The relationship between VT morphology and voice classification of singers therefore seems an important area of research.

Since 1959 all students admitted to the solo singing education at the University of Music Carl Maria von Weber, Dresden, Germany were submitted to a detailed examination of a variety of factors of obvious or potential relevance to voice function (Mürbe *et al.*, 1999; Roers *et al.*, 2009). The examination, made at the beginning of the conservatory education, was performed by a phoniatrician and a group of singing teachers. The resulting admission records included not only assessment of a number of voice parameters but also data on some body characteristics, such as height and weight. Between 1959 and 1991 they also included x-ray imaging of the entire VT during tidal breathing. This unique Dresden material was stored at the Voice Research Laboratory at the University of Music Carl Maria von Weber, Dresden, Germany.

In the present investigation we used this data bank for examining the relationship between experts' classification of singer voices and VT measures. In particular, we asked if there were consistent VT differences between singers of different voice classifications.

II. METHOD

A. Materials

The material analyzed consisted of the admission records of 340 students, all of whom met the following inclusion criteria: (1) They were unanimously classified by the singing teacher panel according to the traditional voice classification system, soprano, mezzosoprano, alto, tenor, baritone, and bass (Titze, 1994), (2) they kept the same classification throughout their education, and (3) they graduated after uninterrupted education as solo singers at the institute. Of these, x-ray images were available for 132 individuals (38.8%), 71 females (40 sopranos, 22 mezzosopranos, and 9 altos), and 61 males (19 tenors, 23 baritones, and 19 basses). The age range was 16–31 years, and the mean age 18.8 and 19.7 years, respectively. However, for various reasons described below, it was impossible to measure all parameters in all x-ray images.

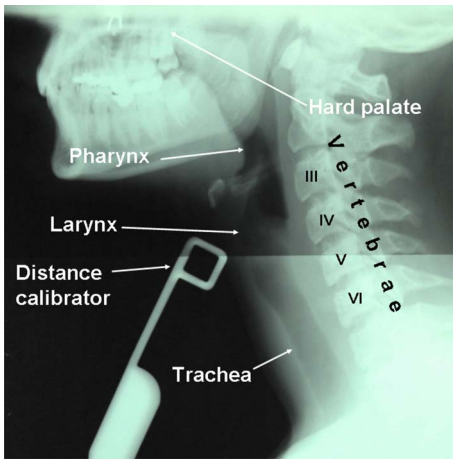


FIG. 1. (Color online) Sagittal x-ray image of the upper airways including pharynx, larynx, proximal trachea, cervical vertebrae, and parts of the skull. Also shown is the metal distance calibrator.

B. Measurements

All x-ray images, taken under identical technical conditions at the Radiological Department of the University Hospital in Dresden, Germany, showed a sagittal projection of the upper airways, see Fig. 1. The profiles included pharynx, larynx, proximal trachea, cervical vertebrae, the skull base, and mostly also the anterior part of the skull. Also included was a metal distance measurement calibrator, inner diameter 1×1 cm^2, mounted on a stand and located in front of the laryngeal midline.

The x rays were taken with the subject in standing position, and with closed mouth. The central ray was about 3–5 cm dorsal to the laryngeal prominence. The tube–photo plate distance was always 1 m, and the distance between the tube and the pharyngeal midline was 25 cm. A six-valve x-ray equipment (D800, TUR, Dresden, Germany) was used with a voltage of 40–45 kV and a 40–75 mA current.

For practical reasons the x-ray images were measured manually by means of a caliper. The x-ray images were placed on a light box. When needed contrast improvement was obtained by using a Cool Brite Illuminator (CB2, RADX, Houston, TX). Measurement reliability, as estimated from the average absolute deviation from the mean of repeated attempts to measure identical distances, was 0.44 mm. This value is comparable to what has been found in other investigations (Fitch, 1997; Lieberman *et al.*, 2001).

Basically adopting the methods used by Dmitriev and Kiselev (1979) and by Damsté (1968) the following distances were measured. The total VT length was measured as the sum of three sections, a pharyngeal, a velar, and an oral; see Fig. 2(a). The pharyngeal section was defined as the distance from the anteriormost point of the laryngeal prominence to the lowermost anterior edge (Margo anterior inferior) of the second vertebra. The velar section was defined as the distance between the latter point and the anteriormost part of the contour of the eighth molar in the maxilla. The

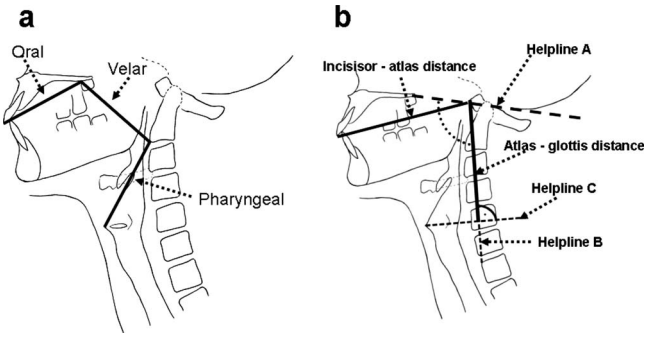


FIG. 2. (a) Schematic illustration of the oral, velar, and pharyngeal straight-line sections of the VT. The oral line runs between the lowermost part of the upper incisor contour and the anteriormost part of the contour of the eighth molar in the maxilla, the velar line from the latter point to the lowermost anterior edge of the second vertebra, and the pharyngeal line from the latter point to the anteriormost point of the laryngeal prominence. (b) Illustration of the measurement of head tilt, defined as the angle between the incisor-atlas line and the atlas-lottis line, which, in turn, were defined by means of helplines A, B, and C.

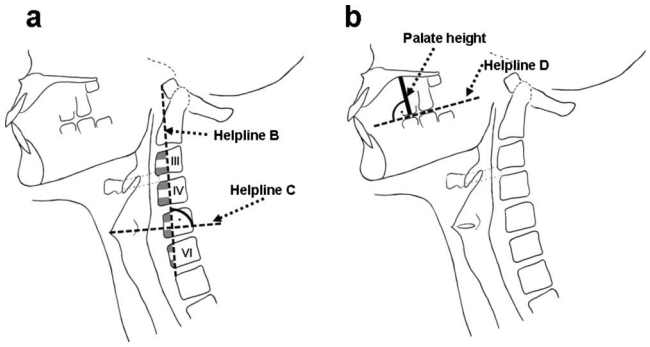


FIG. 3. (a) Illustration of the measurement of the relative position of the glottis by means of helplines B and C. Each vertebra was divided into three equal regions shown by the shadowed and white parts. (b) Illustration of the measurement of the height of the palatal arch by means of helpline D drawn on the lower contour of the molars.

oral section was defined as the distance between the last mentioned point and the lowermost part of the upper incisor contour. Total VT length could not be determined in several images, mostly because they lacked the contour of the hard palate, thus preventing measurement of the oral section (see Table I). Moreover, the angle of the head relative to the spine varied and this affected the velar section and hence also the total VT length. To measure this angle two helplines were drawn, see Fig. 2(b). Helpline A was the so-called McGregor line, which connects the posterior nasal spine with the Os occipital, i.e., the lowermost point on the posterior base of the skull (Hentschel *et al.*, 1999). Helpline B connected the anteriormost part of the contour of the atlas (*Tuberculum anterius*) and the Margo anterior superior of the seventh vertebra, i.e., the uppermost anterior contour of that vertebra. Head tilt was defined as the angle between helplines A and B, averaging to 110° for the subjects. To reduce the influence of head tilt on total VT length, only those images that did not depart more than 15° from this average were accepted. Because of these various limitations of the x-ray view the total VT length could be determined only in 59 out of the 132 cases, see Table I.

The sagittal length of the mouth cavity was measured as the distance from the upper incisors to the atlas. Similarly, the distance along helpline B from its upper end to a line normal to this helpline at the glottal level, helpline C, was used as an estimate of the length of the pharynx cavity.

Since the dimensions of the cervical vertebrae varied between individuals, the vertical position of the larynx was assessed using a relative rather than an absolute measure. Each vertebra was divided into three equal segments. The intervertebral space tended to correspond to one such segment, see Fig. 3(a). The larynx position relative to the vertebrae was determined as the intersection of helpline C with helpline B.

For determining the height of the palate an additional helpline D was constructed, see Fig. 3(b). This helpline was tangent to the lower contour of the upper molars. The height of the palate was measured as the maximum distance normal to this line up to the contour of the palate.

Comparison with vocal fold length is obviously of interest. This parameter was measured in an earlier investigation (Roers *et al.*, 2009) of the same material as was analyzed in the present study. There, it was found that vocal fold length can be accurately predicted from the maximum sagittal diameter of the subglottal trachea; a coefficient of determination of $r2=0.718$ and a trendline slope of 0.98 were found for a material of 29 singers. This method of estimating vocal fold length was used also in the present study for the purpose of comparison with VT dimensions.

TABLE I. X-ray images available and allowing measurement after applying the various inclusion criteria concerning head tilt, measurable oral, velar and pharyngeal sections, and total VT length. F and M in the classification column refer to female and male gender.

		Accepted after application of criterion concerning				
Classification	X rays available	Head tilt	Oral section measurable	Velar section measurable	Pharyngeal section measurable	Total VT length measurable
Soprano (F)	40	39	12	21	39	12
Mezzosoprano (F)	22	22	8	13	22	8
Alto (F)	9	9	5	6	8	5
Tenor (M)	19	18	12	14	18	12
Baritone (M)	23	23	10	16	22	10
Bass (M)	19	17	12	14	18	12
Sum	132	127	59	84	127	59

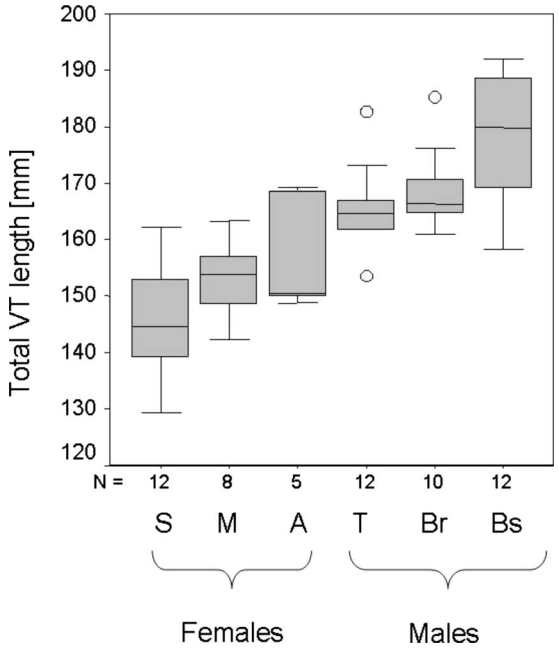


FIG. 4. Box plot of the total VT length for the voice classifications soprano (s), mezzosoprano (m), alto (a), tenor (t), baritone (Br), and bass (Bs). The upper and lower limits of the shadowed areas represent the first and the third quartiles, the line is the median, and the bars show the highest and lowest values that were not classified as outlyers, which, in turn, are shown in terms of open circles.

All measurements were made in the absence of information about the subject's voice classification.

C. Statistical analyses

Statistical analyses were carried out by comparison of means by analysis of variance (ANOVA)-function of SPSS 11.5 for Windows® (Chicago, IL). Significant differences were identified by applying a Student-Newman-Keuls-Procedure (*a posteriori*—testing for multiple comparisons of averages). Also Pearson's r values were calculated by the SPSS software. The significance level applied was at least $\alpha = 0.05$.

III. RESULTS

Figure 4 shows in terms of box plots the relationship between voice classifications and the total VT length as measured in terms of the sum of the three straight-line sections. According to the ANOVA the differences between most of the classifications were significant, except for the mezzosopranos who did not differ significantly from sopranos or altos, and for the altos who did not differ significantly from tenors. Also, the tenors did not differ significantly from baritones.

The variation of the total VT length was more dependent on the length of the pharynx cavity than on the length of the mouth cavity. This can be seen in Fig. 5. For the length of the mouth cavity (left panel) defined as the upper incisor–atlas distance, the correlation was significant ($r2=0.323$, $p<0.01$). This was also true both for the group of females alone ($r2=0.654$, $p<0.01$) and for the group of males alone ($r2=0.161$, $p<0.05$). The correlation between the total VT length and the length of the pharynx cavity, measured as the atlas-glottis distance (right panel), was stronger ($r2=0.602$, $p<0.01$). Again, this correlation applied also to the group of females alone ($r2=0.314$, $p<0.01$) and to the group of males alone ($r2=0.335$, $p<0.01$). Also, the slope of the trendline is much closer to 1.0 for the length of the pharynx cavity than for the length of the mouth cavity.

Significant correlation with total VT length was found also for some other parameters, although some of them failed to differ significantly between voice classifications. The left panel of Fig. 6 shows the relationship between classification and the height of the palate. None of the classifications differed significantly to any other classification. On the other hand, as can be seen in the right panel of Fig. 6, the palate height itself showed a significant variation with total VT length (all singers: $r2=0.493$, $p<0.01$; females alone: $r2=0.406$, $p<0.01$; males alone: not significant). Most female

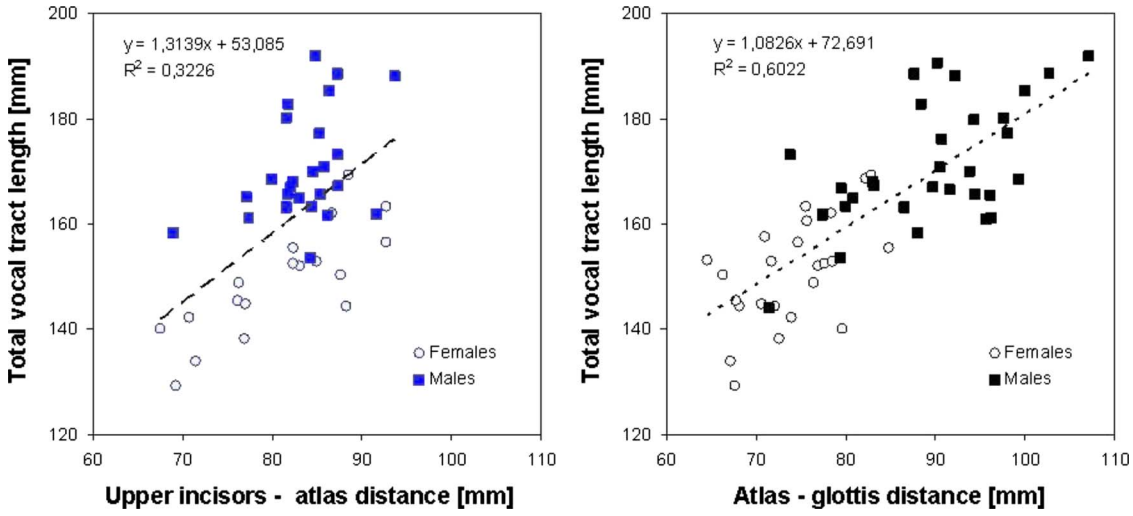


FIG. 5. (Color online) Total VT length vs length of mouth cavity, defined as distance between the upper incisors and atlas, and vs length of pharynx cavity defined as atlas-glottis distance (left and right panels). The dashed line represents the trendline for the female and male subjects pooled with equation and $r2$ given in the upper left corner.

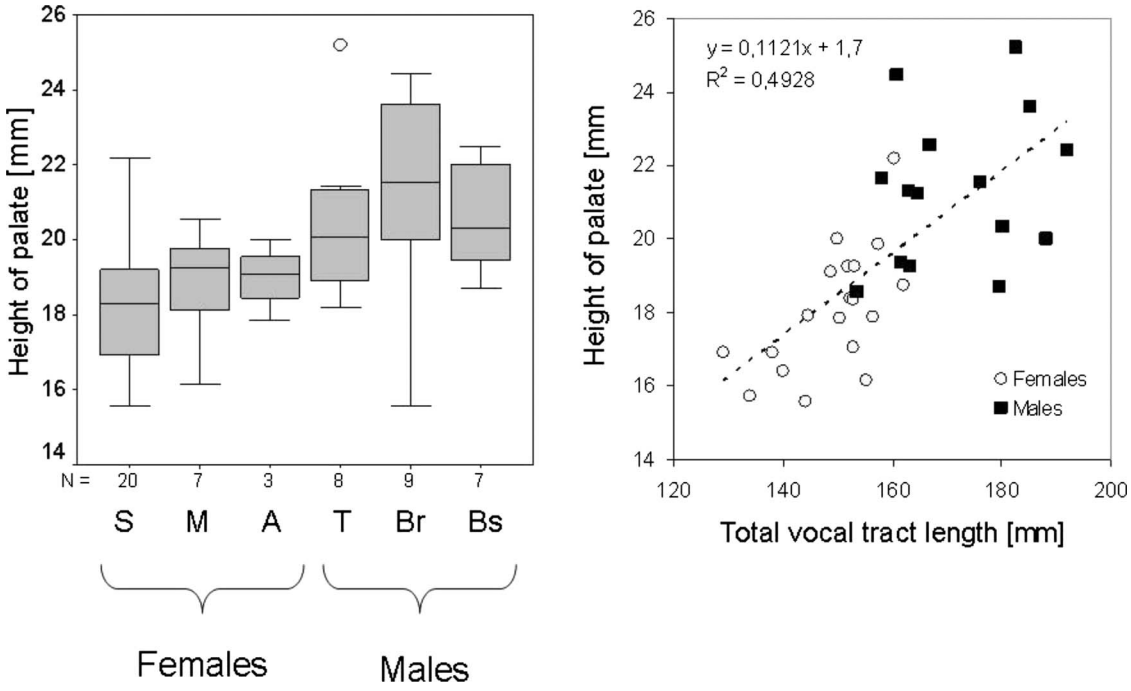


FIG. 6. The left panel shows a box plot of the height of the palatal arch for the indicated voice classifications (acronyms as in Fig. 4). The right panel shows the same parameter as a function of total VT length. The dashed line represents the trendline for the female and male subjects pooled with equation and $r2$ given in the upper left corner.

singers had a lower palate than the male singers. The length of the mouth cavity and the height of the palate were significantly correlated (all singers: $r2=0.323$, $p<0.01$; females alone: $r2=0.654$, $p<0.01$; males alone: $r2=0.161$, $p<0.05$).

In a previous investigation it was revealed that vocal fold length is strongly correlated with the maximum anterior-posterior diameter of the trachea, henceforth referred to as the maximum tracheal diameter. This measure could be determined in 117 singers in our material (Roers *et al.*, 2009). The left panel of Fig. 7 shows a box plot for this measure for various classifications. The ANOVA revealed that each classification differed significantly from all other classifications, except for mezzosoprano versus alto. As can be seen in the

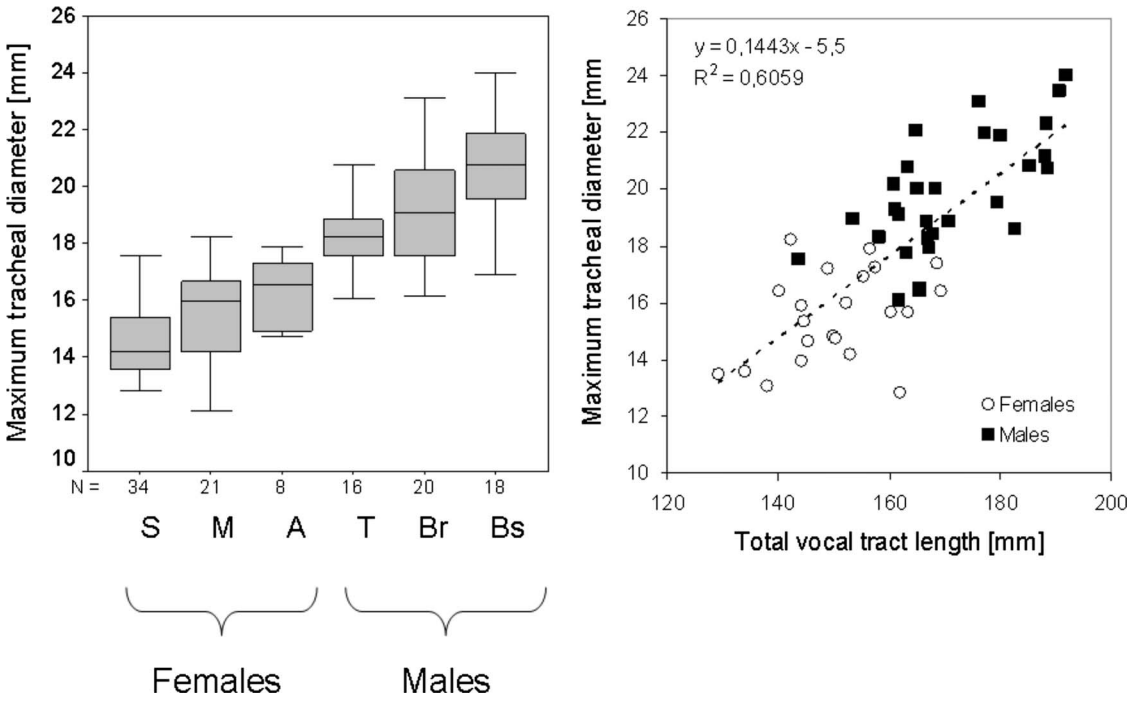


FIG. 7. The left panel shows a box plot of the maximal sagittal diameter of the trachea for the indicated voice classifications (acronyms as in Fig. 4). The right panel shows the same parameter as a function of total VT length. The dashed line represents the trendline for the female and male subjects pooled with equation and $r2$ given in the upper left corner.

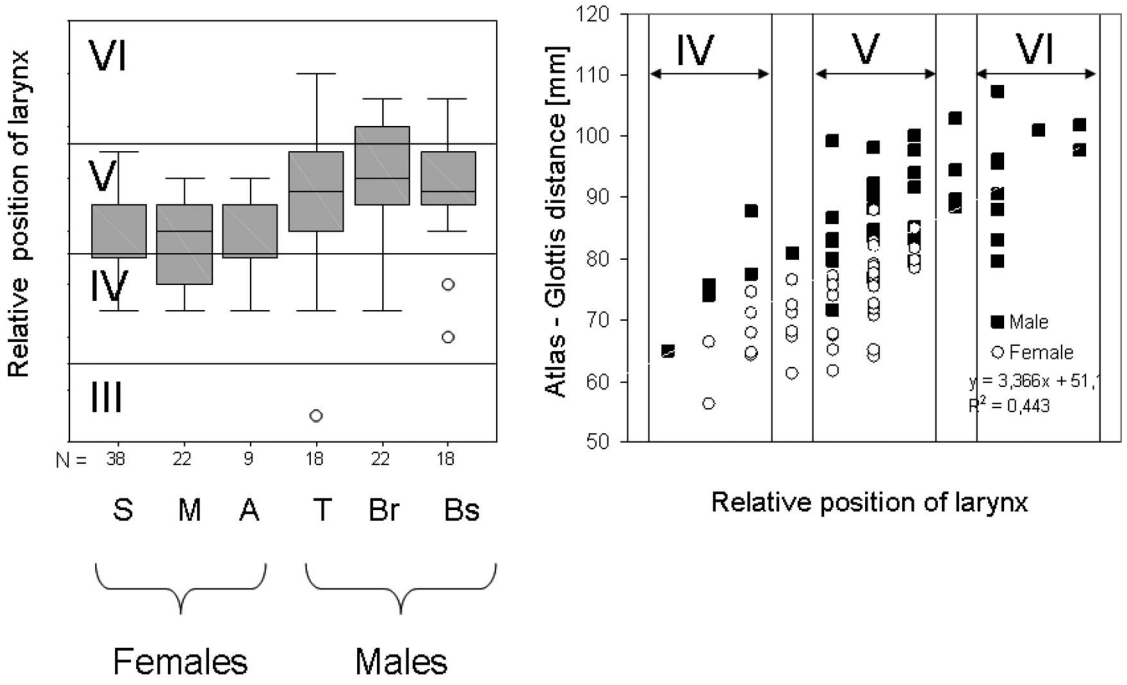


FIG. 8. The left panel shows a box plot of the relative position of the larynx for the indicated voice classifications (acronyms as in Fig. 4). The right panel shows the length of the pharynx cavity as a function of the same parameter. In both panels, Roman numbers represent vertebrae.

right panel of Fig. 7 this measure was also correlated with total VT length (all singers: $r2=0.606$, $p<0.01$; females alone: not significant; males alone: $r2=0.429$, $p<0.01$). Thus, this distance tended to be smaller in the female singers, and long VTs appeared in combination with long vocal folds in the male singers.

The larynx position varied systematically with total VT length, as mentioned. However, also the relative position of the larynx showed some systematic variation, as illustrated in the left panel of Fig. 8. The classifications did not differ significantly although the mean values for the female singers tended to be somewhat lower than that of the male singers. Thus, in the females, the glottis was located at a higher relative level than in the male singers. Even though the relative larynx position correlated with the length of the pharynx cavity (all singers: $r2=0.443$, $p<0.01$; females alone: $r2=0.419$, $p<0.01$; males alone: $r2=0.441$, $p<0.01$), as illustrated in the right panel of Fig. 8, no significant relationship was found between relative larynx position and total VT length.

Mean body height showed differences between classifications, see left panel in Fig. 9. Thus, according to the ANOVA, all classifications differed significantly except sopranos and mezzosopranos and baritones and basses. The right panel of Fig. 9 illustrates the relationship between body height and total VT length. There was a significant correlation for the entire group of singers and for the males (all singers: $r2=0.543$, $p<0.01$; females alone not significant; males alone: $r2=0.325$, $p<0.01$).

With respect to bodyweight, only genders differed significantly according to the ANOVA (see Fig. 10). Bodyweight showed a significant correlation with VT length for the entire singer group, but not within genders (all singers: $r2=0.421$, $p<0.01$; females alone: not significant; males alone: not significant); see the right panel of Fig. 10.

Figure 11 shows the data for the body mass index. The differences between classifications were small although the correlation with total VT length (right panel of Fig. 11) was significant for the entire group of singers (all singers: $r2=0.114$, $p<0.05$; females alone: not significant; males alone: not significant).

IV. DISCUSSION

The main finding of our investigation was that the total VT length differed systematically between most voice classifications, so sopranos tended to have the shortest VTs and basses the longest. The effect was more dependent on the length of the pharynx cavity than on the length of the mouth cavity, a tendency which seems to apply also to nonsingers (Vorperian *et al.*, 2005).

Regarding gender differences in total VT length, it is well known that female adults have shorter VTs than male adults. According to Chiba and Kajiyama (1941) the female VT length is on average 0.87 of the male VT length. The ratio of our average VT length for female and male singers was similar, 0.89. Using sagittal MRI-scans Fitch and Giedd (1999) observed a mean VT length of 150 mm in 6 females, aged 17–25 years, and 161 mm for 13 males in the same age group. The mean VT lengths for our female and male singers amounted to 152 and 171 mm, respectively. However, their values are not fully comparable with ours, since Fitch and Giedd (1999) measured the length as the sum of five straight-line segments. This will shorten the length. Also they included the lips, which will increase the length.

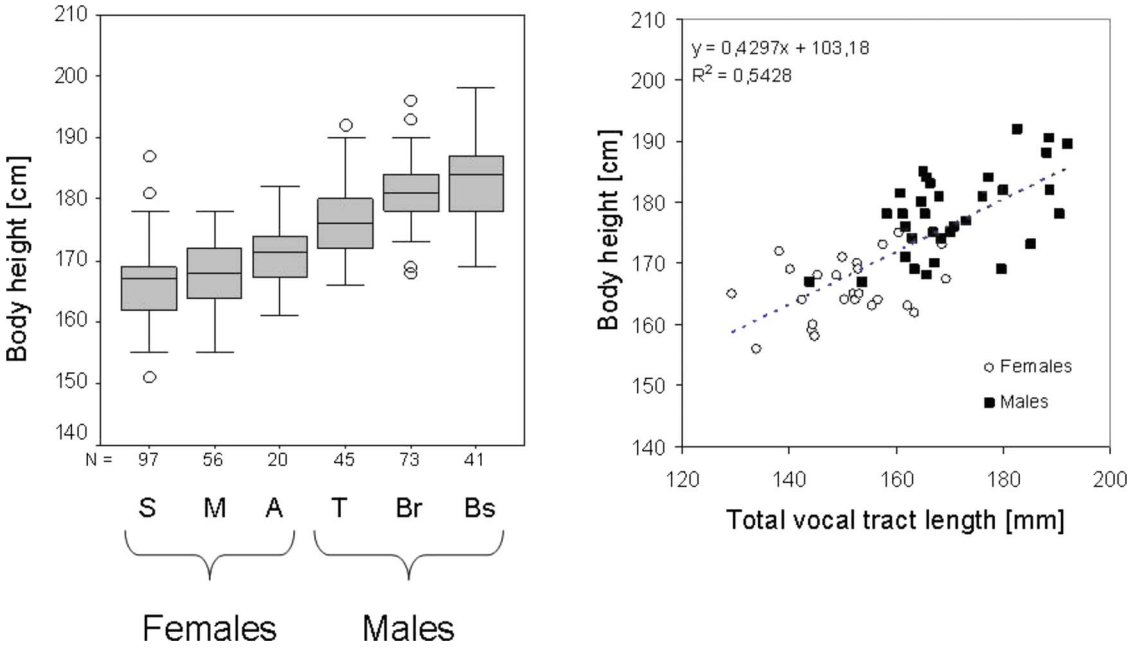


FIG. 9. (Color online) The left panel shows a box plot of body height for the indicated voice classifications (acronyms as in Fig. 4). The right panel shows the same parameter as a function of total VT length. The dashed line represents the trendline for the female and male subjects pooled with equation and $r2$ given in the lower right corner.

On the basis of formant frequency differences between female and male speakers Fant (1966) postulated that the mouth and pharynx cavities were 13 and 23 mm shorter in female than in male adults. Smaller differences were found in our material. For our male and female singers the average length of the incisor-atlas distance differed by no more than 2 mm. This small difference may be due to a selection criterion. The difference between the female and male average lengths of the pharyngeal cavity was greater, 13 mm. This, too, is considerably smaller than the value postulated by Fant (1966). The discrepancy may be due to the fact that his data referred to speech while our measurements refer to resting position. Indeed, Sulter *et al.* (1992) found that changes of the larynx position could cause a change of VT length of almost 4 cm.

Total VT length can be measured in several different manners. The acoustically relevant measure is the VT midline as used by Sulter *et al.* (1992) and Fitch and Giedd

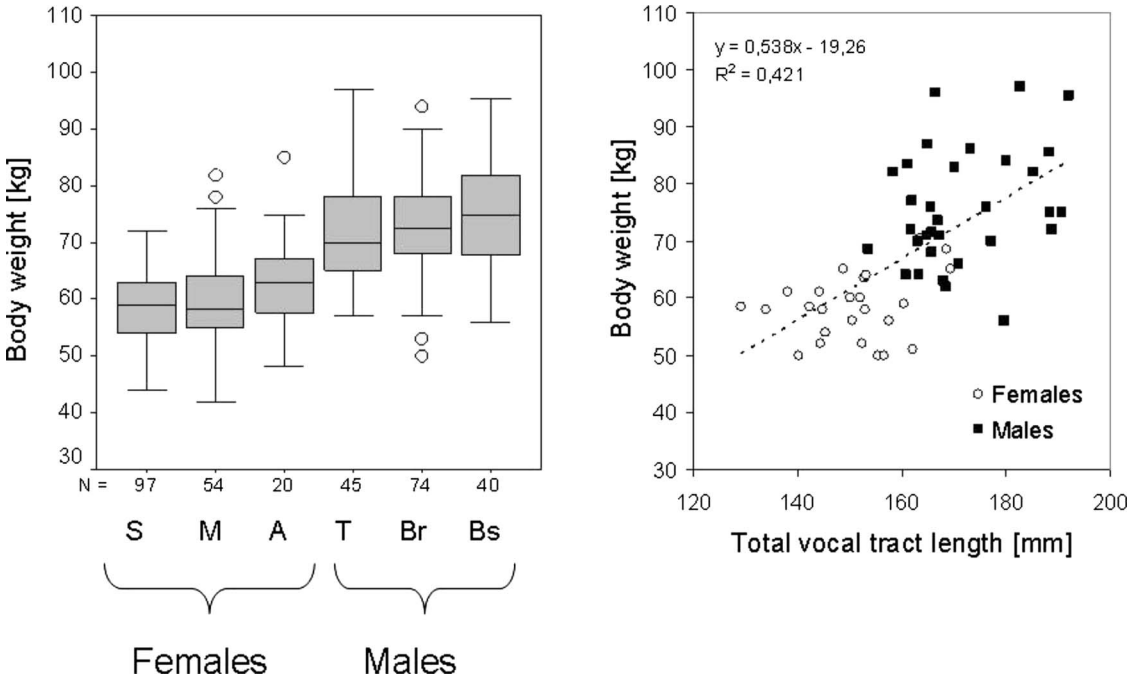


FIG. 10. The left panel shows a box plot of bodyweight for the indicated voice classifications (acronyms as in Fig. 4). The right panel shows the same parameter as a function of total VT length. The dashed line represents the trendline for the female and male subjects pooled with equation and $r2$ given in the upper left corner.

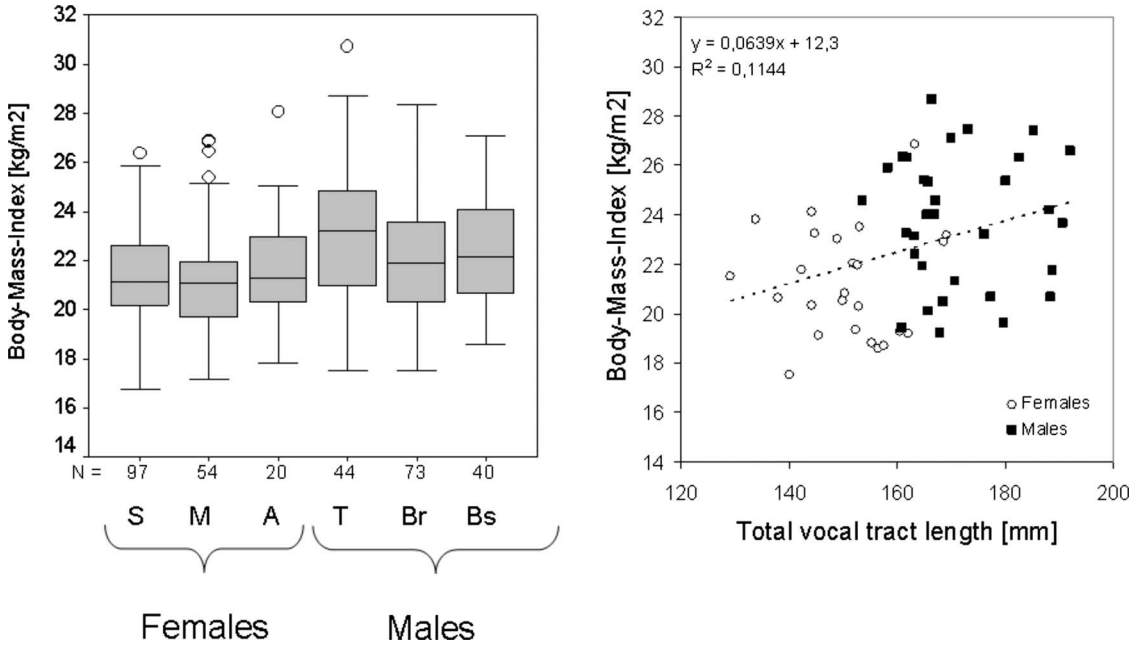


FIG. 11. The left panel shows a box plot of body mass index for the indicated voice classifications (acronyms as in Fig. 4). The right panel shows the same parameter as a function of total VT length. The dashed line represents the trendline for the female and male subjects pooled with equation and $r2$ given in the lower left corner.

(1999). In the present material, however, this was not feasible since the VT walls that are constituted by soft tissues were rarely visible clearly enough in x-ray images. Instead we used a three line segmentation system, based on stable structures that could easily be identified in the material. Our method was basically identical with the straight-line alternative described by Sulter *et al.* (1992). The difference was that our oral section ran between the lower edge of the upper incisor to the anteriormost part of the eighth molar while their ran from the caudal part of the upper lip to the highest part of the hard palate, which, incidentally, may be affected by head tilt.

Our results are easier to compare with those published by Dmitriev and Kiselev (1979) in the sense that our measurement method was similar to theirs. They reported that VT lengths varied systematically with singer classification, just as our results show. However, the lengths that they found were greater for all classifications. This was particularly true for the male singers, the difference amounting to 35% for the baritones and basses and only 8% for the sopranos. The reason for this discrepancy is difficult to find since they do not describe exactly how they obtained their results. However, their subjects were x-rayed while singing a loud tone, while our data refer to rest position, as mentioned. This supports the assumption that their subjects used a larynx position lower than at rest.

Summarizing, our data deviate somewhat from those published in previous investigations. Some discrepancies would be due to different definitions in VT length. Some may be due to differences in head tilting, since a forward tilting will shorten and a backward tilting will lengthen the VT when measured as in the present study. In addition it cannot be excluded that singers' VT morphology deviates in some systematic manners from those of a nonsinger population.

VT dimensions determine formant frequencies. Our results have shown systematic VT length differences between singer voice classifications, particularly with respect to the length of the pharynx cavity. Fant (1960) showed that the second formant frequency F2 of the vowel /i/ can be regarded as a standing wave in the pharynx, so pharynx length should be related to this formant. Cleveland (1977) analyzed the formant frequencies of professional male operatic singers. He showed that tenors and bass singers' F2 in /i/ differed considerably, baritone singers' value lying in between. Our findings regarding the pharynx length in singers seem to explain Cleveland's (1977) observations. Sundberg (2001) showed that the center frequency of singer's formant was slightly lower in bass singers than in baritone singers, tenors having the highest values. This cluster is constituted by F3, F4, and F5 and of these the latter two have been shown to be particularly sensitive to the larynx tube (Sundberg 1974; Takemoto *et al.*, 2006). If male singers' systematic variation of pharynx length is reflected also in the larynx tube length, the difference in the center frequency of the formant cluster could be explained.

Our results failed to provide support for Gutzmann's (1949, 1958) idea that the height of the palatal arch varies systematically with singer voice classification. Rather our findings agree with those of Ackermann (1967) and Marunick and Menaldi (2000) that there is no systematic variation of palate height and singer voice classification.

For the entire group of singers we found a significant correlation between the total VT length and vocal fold length, approximated from the maximum sagittal diameter of the subglottal trachea. This correlation is interesting. As long vocal folds are needed for producing low fundamental frequencies, it suggests that voices with a low pitch range have longer VTs than voices with a higher range. This finding

again seems to explain Cleveland's (1977) finding that for the same vowel basses tended to have the lowest formant frequencies and tenors the highest. It also seems to explain Sundberg's (2001) observation that basses tend to have the lowest and tenors the highest center frequencies of the singer's formant cluster. The female subjects, on the other hand, did not show a significant correlation between total VT length and our estimate of vocal fold length.

Some experts on singing have reported that they have observed a relationship between the classification of singer voices and both body height and bodyweight (see, e.g., Martienssen-Lohmann, 1956; Luchsinger, 1970; Pfau 1973). Fitch and Giedd (1999) analyzed the VT length and body size and weight and found strong correlations in 129 individuals in the age range of 2–25 years. For our entire group and for our male singers we found much lower correlations between VT length and body height, and none for the female singers. A reason for the difference between our results and theirs may be that our material included adults only. With regard to body weight we found a significant correlation with VT length for the entire group only, the reason seemingly being the gender difference in this regard as shown by the ANOVA. The fact that our singers were all rather young may be relevant to this result.

Among teachers of singing a correct classification of a singer's voice is regarded as very important to the successful development of a solo singer. This classification is based mainly on subjective criteria. Hence, a possibility to support this classification by objective data would be valuable. Our results failed to demonstrate any VT measure that was specific to a particular classification. On the other hand our material revealed VT length and particularly pharynx length as factors of relevance to voice classification. This relationship may be obscured in an image material taken during phonation.

V. CONCLUSIONS

Our analysis of x-ray profiles of 132 young singers of different voice classifications has revealed that the total VT length differs systematically between the classifications. The length of the pharyngeal cavity showed a particularly strong relationship with this length. This appears to explain earlier findings on formant frequency differences between voice classifications. Our analyses further showed a relationship between body length and voice classification, sopranos tending to be short and basses to be tall. In addition the VT length showed a significant correlation with the sagittal diameter of the trachea, which, in turn, is strongly correlated with vocal fold length.

ACKNOWLEDGMENTS

The database at the Voice Research Laboratory, University of Music Carl Maria von Weber, Dresden, Germany is administered by Hartmut Zabel. The investigation including analysis of x rays and data acquisition was carried out in the years 1998–2000 by F.R. with support in terms of a Marie Curie scholarship from the European Commission during her stay at the Department of Speech, Music, and Hearing, KTH, Stockholm, Sweden. Eberhard Kuhlisch performed the ANOVA analysis. The manuscript was written by D.M. and J.S. A preliminary version of this investigation was presented at the PanEuropean Voice Conference (PEVOC), Stockholm, Sweden, August 2001.

[1]This peak has previously been referred to as the singer's formant.

Ackermann, R. (**1967**). "Zur frage von gaumenform und stimmqualität (On the question of palate shape and voice quality)," Z. Laryngol. Rhinol. Otol. **46**, 280–284.

Chiba, T., and Kajiyama, M. (**1941**). *The Vowel—Its Nature and Structure* (Tokyo-Kaisekan, Tokyo, Japan).

Cleveland, T. F. (**1977**). "Acoustic properties of voice timbre types and their influence on voice classification," J. Acoust. Soc. Am. **61**, 1622–1629.

Damsté, P. H. (**1968**). "X-ray study of phonation: Applications," Folia Phoniatr. (Basel) **20**, 65–88.

Dmitriev, L., and Kiselev, A. (**1979**). "Relationship between the formant structure of different types of singing voices and the dimensions of supraglottic cavities," Folia Phoniatr. (Basel) **31**, 238–241.

Fant, G. (**1966**). "A note on vocal tract size factors and nonuniform F-pattern scalings," Speech Transm. Lab. Q. Prog. Status Rep. **4**, 22–30.

Fant, G. (**1960**). *Acoustic Theory of Speech Production*, 1st ed. (Mouton, The Hague, Paris).

Fitch, W. T. (**1997**). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," J. Acoust. Soc. Am. **102**, 1213–1222.

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.

Gutzmann, H. (**1949**). "Über das Erkennen der Stimmgattung (On the identification of voice classification)," Folia Phoniatr. (Basel) **2**, 1–5.

Gutzmann, H. (**1958**). "Über die Voraussetzungen eines Erfolgversprechenden Gesangsstudiums (On the requirements for a successful singer education)," Arch. Ohren Nasen Kehlkopfheilkd **173**, 546–550.

Hentschel, F., Heuck, F. H. W., Voigt, K., and Bast, B. R. G. (**1999**). *Schädel-Gehirn-Wirbelsäule-Rückenmark (Skull, Brain, Cervical Vertebrae, Back Markings) Radiologische Skizzen und Tabellen* (Georg Thieme, Stuttgart).

Lieberman, D. E., McCarthy, R. C., Hiiemae, K., and Palmer, J. B. (**2001**). "Ontogeny of postnatal hyoid and larynx descent in humans," Arch. Oral Biol. **46**, 117–128.

Luchsinger, R. (**1970**) in *Handbuch der Stimm-und Sprachheilkunde*, 3rd ed., edited by R. Luchsinger and G. E. Arnold (Springer, Wien), Vol. **1**.

Martienssen-Lohmann, F. (**1956**) *Der Wissende Sänger (The Knowledgeable Singer)* (Atlantis, Zürich).

Marunick, M. T., and Menaldi, C. J. (**2000**). "Maxillary dental arch form related to voice classification: A pilot study," J. Voice **14**, 82–91.

Mürbe, D., Sundberg, J., Iwarsson, J., Pabst, F., and Hofmann, G. (**1999**). "Longitudinal study of solo singer education effects on maximum SPL and level in the singers' formant range," Logoped. Phoniatr. Vocol. **24**, 178–186.

Nordström, P. E. (**1977**). "Female and infant vocal tract simulated from male area functions," J. Phonetics **5**, 81–92.

Peterson, G. E. and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.

Pfau, W. (**1973**) in *Klassifizierung der Menschlichen Stimme (Classification of the Human Voice). Hals-, Nasen-und Ohrenheilkunde*, edited by A. Herrmann and H. Jakobi (Johann Ambrosius Barth, Leipzig), Vol. **22**.

Roers, F., Mürbe, D., and Sundberg, J. (**2009**). "Predicted singers' vocal fold lengths and voice classification—Study of x-ray-morphological measures," J. Voice; available at http://www.sciencedirect.com/science?_ob=ArticleListURL&_method=list&_ArticleListID=810766609&_sort=d&view=c&_acct=C000034958&_version=1&_urlVersion=0&_userid=4478132&md5=949eab8b4b9e2dca2614d4b6cd4f407c (last viewed 10/21/2008).

Rudolph, S. (**1969**). "Vergleiche zwischen Gaumenform und Stimmtimbre bei Gesangsstudenten (Comparisons between palate shape and voice quality in singer students)," diss., Medical University, Dresden, Germany.

Sulter, A.M., Miller, D.G., Wolf, R.F., Schutte, H.K., Wit, H.P., and Mooyaart, E.L. (**1992**). "On the relation between the dimensions and resonance characteristics of the vocal tract: A study with MRI," Magn. Reson.

Imaging **10**, 365–373.
Sundberg, J. (**2001**). "Level and center frequency of the singer's formant," J. Voice **15**, 176–186.
Sundberg, J. (**1974**). "Articulatory interpretation of the 'singing formant,'" J. Acoust. Soc. Am. **55**, 838–844.
Takemoto, H., Adachi, S., Kitamura, T., Mokhtari, P., and Honda, K. (**2006**). "Acoustic roles of the laryngeal cavity in vocal tract resonance," J. Acoust. Soc. Am. **120**, 2228–238.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).
Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (**2005**). "Development of vocal tract length during early childhood: A magnetic resonance imaging study," J. Acoust. Soc. Am. **117**, 338–350.

Quantifying bat call detection performance of humans and machines

Mark D. Skowronski[a] and M. Brock Fenton[b]
Department of Biology, University of Western Ontario, London, Ontario N6A 5B7, Canada

(Received 14 April 2008; revised 2 October 2008; accepted 21 October 2008)

Methods for detecting echolocation calls in field recordings of bats vary in performance and influence the effective range of a recording system. In experiments using synthetic calls from five species, human detection accuracy was $89.7 \pm 0.6\%$, compared to $76.3 \pm 0.8\%$ for a model-based detector, $72.2 \pm 0.8\%$ for an energy-based detector, and $98.4 \pm 0.2\%$ for an optimal linear detector. The energy-based detector was 11 times faster than the model-based detector and 110 times faster than humans. Human accuracy was positively correlated with test duration ($R^2=0.43$, $P<0.05$), meaning that higher accuracy was achieved at the expense of slower performance. Species was a significant factor determining accuracy for all detectors ($P<0.001$) because of call bandwidth: Narrowband calls concentrated energy in a narrower frequency band and were easier to detect. For a hypothetical recording system, range at 90% human detection accuracy varied from 10 to 35 m among species, while range dropped by approximately 20% using the automated detectors. The optimal detector outperformed humans by 5 dB and the automated methods by 9 dB. The results quantify the tradeoff between detector speed and accuracy and are useful for designing field studies of bats. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3025913]

PACS number(s): 43.80.Ev, 43.60.Uv [MCH] Pages: 513–521

I. INTRODUCTION

Detecting calls in recordings of echolocating bats is fundamental to monitoring bat activity and distribution. Detecting bat calls is required to further analyze individual calls (i.e., global feature extraction) or to simply count calls in a recording. Traditionally, call detection and analysis was performed by hand.[1–5] Calls may be detected by viewing a signal in the time domain (oscillogram) or the time-frequency domain (spectrogram) or by listening to a recording using techniques that move calls into the audible frequency range for humans.[6] Listening to calls in real time while recording is useful for producing manually triggered recordings.[7,8] Spectrograms are more useful than oscillograms for manual detection of calls because they (1) display more information about a call, (2) provide a better signal-to-noise ratio (SNR), and (3) handle overlapping calls better.

Several automated methods have been employed to detect calls. BatSound (Pettersson Elektronik AB, Uppsala, Sweden) uses a threshold for spectrogram pixel value for detection.[9] Masters *et al.*[10] also used spectrogram analysis to autodetect calls; a call was a group of maximum-energy spectral values at least 12 dB above the noise level or within 30 dB of a local spectral peak, whichever was greater. Parsons and Jones[11] used a Hilbert transform to compute the amplitude envelope of a signal; an envelope sequence was normalized between 0 and 1, and all groups of values above a threshold constituted a call. By normalizing the envelope, the threshold scaled relative to the peak call value (nearly independent of the noise level). They used a threshold of 0.003, which is 50 dB below a call's peak power. Such a threshold was adequate for their analysis because they only analyzed the call with highest amplitude from each recorded pass. Obrist *et al.*[12] used a "simple integrating detector algorithm" to automatically detect calls although no further details of the algorithm were given. Bayefsky-Anand *et al.*[13] used a rule-based detector to automatically detect calls from field recordings of *Tadarida teniotis*.

Most automated methods rely solely on signal power to separate calls from background noise. A limitation of this approach is that *any* signal of sufficient power to exceed the various thresholds will be detected as a call (e.g., insects, birds, rustling leaves, and rain). Such false positives are typically removed from analysis by human screening, but this approach requires significant expert intervention, which is time consuming and subject to bias. Skowronski and Harris[14] employed statistical models, inspired by automatic speech recognition research, to increase the quantity of information extracted from a recording for call detection. The models used spectral peak energy and also frequency and local call shape information (temporal derivatives) to separate calls from background noise. A model-based detector was eight times more accurate than a simple broadband energy detector and two times more accurate than a spectral peak detector. The same features used for detection were also useful for species classification, demonstrating increased utility for model-based detection. The models were trained from hand-labeled calls, but Skowronski and Fenton[15] showed that training could be fully automated by using an energy-based detector to replace hand-labeled calls for model initialization.

The relative detection performance of humans and automated methods has not been previously measured. A reasonable hypothesis is that humans are more accurate but slower than automated methods. Quantifying performance would

[a]Electronic mail: mskowro2@uwo.ca
[b]Electronic mail: bfenton@uwo.ca

TABLE I. Call parameters for synthetic calls. Values are ranges for uniform distributions. Turning point parameters, t_0 and g_0, are normalized time and frequency, respectively.

Species	F_{start} (kHz)	F_{end} (kHz)	Duration (ms)	t_0	g_0
Eptesicus fuscus	70–80	26–30	7–11	0.35–0.45	0.24–0.34
Lasiurus cinereus	35–45	20–24	12–18	0.16–0.26	0.33–0.43
Myotis lucifugus	85–95	35–40	5–7	0.32–0.42	0.28–0.38
M. septentrionalis	110–120	39–49	3–4	0.43–0.53	0.25–0.35
Pipistrellus subflavus	85–95	39–41	9–13	0.16–0.20	0.18–0.22

help bat researchers make better decisions regarding the design and scope of experiments using acoustic methods and understand and assess the tradeoffs involved with automated detection. We measured detection performance for humans using a two-choice experiment and compared the accuracy and speed of humans to two automated methods: an energy-based detector and a model-based detector. All experiments used synthetic echolocation calls embedded in white noise, which allowed precise control of SNR and provided an absolute ground truth for all experiments. The use of synthetic calls also allowed us to measure optimal detector performance using a matched-filter detector for comparison with the other detectors. We characterized performance versus bat species and SNR for humans and machines as well as detector threshold for machine methods. We converted detection accuracy into effective detection range using a hypothetical recording system, which included energy loss due to spherical spreading and atmospheric absorption to quantify detector accuracy in practical terms.

II. METHODS

A. Human detection experiment

The purpose of the detection experiment was to measure the ability of human subjects to detect echolocation calls embedded in noise. Synthetic signals mimicking the calls of five common North American species, embedded in additive white noise, were used in the experiments (Table I, Fig. 1). Calls were synthesized using a piecewise exponential frequency modulation function[15]

$$x(n) = A(n)\sin(2\pi\phi(n)),$$

$$s(n) = \frac{1}{\sqrt{1/K\Sigma_{j=0}^{K-1}x^2(j)}}10^{\mathrm{SNR}/20}x(n), \quad (1)$$

where $A(n)$ is a Hamming window amplitude modulation function of length K samples, n is discrete time, and $\phi(n)$ is the phase of a frequency modulation function. SNR is the desired signal-to-noise ratio of the call $s(n)$ embedded in additive white noise $w(n)$ of zero mean and unity variance

$$\mathrm{SNR} = 10\log\left(\frac{1/K\Sigma_{n=0}^{K-1}s^2(n)}{1/K\Sigma_{n=0}^{K-1}w^2(n)}\right). \quad (2)$$

The scale factor for $s(n)$ in Eq. (1) ensures that all calls from all species at a given SNR have the same signal power, regardless of call length, which is synonymous with having the same sound pressure level (SPL).[16]

The test was administered through a graphical user interface (GUI) designed in MATLAB R2007B (The Mathworks, Natick, MA). The GUI contained a figure axis used to display spectrograms and two colored buttons: a green button labeled "CALL" and a red button labeled "NO CALL" (Fig. 2). The test began by displaying instructions for the test to the test subject. The instructions stated that the subject would be presented with spectrograms of synthetic recordings, which contained either no bat call or a single bat call from one of five Ontario species (all subjects were familiar with the echolocation calls of the eight native Ontario species, although they did not know which five species were represented in the test). If the subject detected a call in a spectrogram, the subject was instructed to press the CALL button; otherwise, the subject was to press the NO CALL button. The instructions stated that each answer was final and that the next spectrogram would be presented after each button

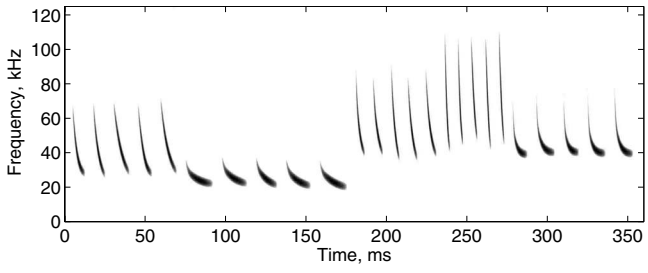


FIG. 1. Example synthetic calls from parameters in Table I. Each group of calls, left to right, represents calls of *Eptesicus fuscus, Lasiurus cinereus, Myotis lucifugus, M. septentrionalis*, and *Pipistrellus subflavus*. The spectrogram was created using 1 ms Blackman windows, 1024-point zero-padded FFTs, and 5000 overlapping windows per second (same parameters used in experiments with human experts).

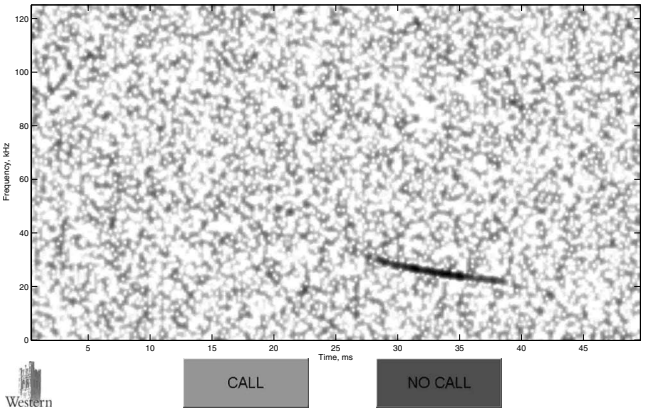


FIG. 2. Test GUI used to administer detection experiment to human experts.

push. The instructions also stated that calls at various SNRs would be presented and that a counter in the GUI would indicate the subject's progress during the test. Subjects were offered payment to take the test, and subjects were informed before the test that the test would take approximately 30 min to complete. No other instructions were given to the subjects.

Each test consisted of 1000 spectrograms: 500 with no call and 500 with a single call. Calls from each of the five species were added to white noise at ten different SNRs (−16 to −7 dB in 1 dB increments), and each combination of species and SNR was repeated ten times. The order of presentation was randomized for each subject. For each of 1000 trials, a 50 ms window of white noise (zero mean, unity variance, and 250 kHz sampling rate) was created. If the trial contained a call, a synthetic call from the appropriate species and SNR was constructed using Eq. (1) and added to the noise at a random location inside the 50 ms window.

A spectrogram of the 50 ms window was constructed using 250-sample Blackman windows (1 ms), with 80% overlap between adjacent windows (5000 windows/s), and 1024-point zero-padded fast Fourier transforms (FFTs). The magnitude of each spectrogram pixel value was log-compressed, the median pixel value was determined, and all pixels less than a threshold of 3 dB below the median were truncated to the threshold (this procedure ensured that small pixel values did not skew the range of color map values used to display each spectrogram). The size of each spectrogram was 513 rows × 246 columns. The test was administered on a computer with a 19 in. flat-panel display with a resolution of 864 rows × 1152 columns of pixels, and the size of the figure axis in the GUI was 536 rows × 973 columns of pixels, ensuring that each spectrogram pixel was represented by at least one display pixel. Spectrograms were displayed using a white-yellow-red-black color map (the "hot" color map in MATLAB) with white representing the lowest pixel value and black representing the highest.

We performed an analysis of variance (ANOVA) to determine the effects of species, SNR, and individual on detection accuracy. The dependent variable was binary: 1 for a correct response (hits and true negatives) and 0 for an incorrect response (misses and false positives) to each trial. The independent variables for species (5 levels) and SNR (10 levels) were treated as fixed factors, while individual (12 levels) was treated as a random factor.

B. Energy-based detector

The energy-based detector operated by high-pass filtering each 50 ms window at a cutoff frequency of 18 kHz and calculating the energy in 1250-sample rectangular windows (5 ms). Adjacent windows overlapped by 50% (400 windows/s); thus, each 50 ms window produced 19 energy values, which were scaled such that the median value was 0 dB. The maximum-energy value for each 50 ms window was compared to a threshold, and the detector determined that a call was detected if the maximum energy exceeded the threshold. Different thresholds controlled the tradeoff between high sensitivity (high hit rate) and high specificity (low false-positive rate) for the energy-based detector.

C. Model-based detector

The model-based detector employed a Gaussian mixture model (GMM).[14] The following frame-based features were extracted for model input: peak energy E (dB), frequency of most energy (FME) (kHz), first-and second-order temporal derivatives of E and FME, and smoothness features of E and FME. Temporal derivatives were estimated using linear regression (±3 frames), and smoothness features were calculated using the log-compressed residual error from linear regression.[15] Features were extracted from spectrograms constructed using 100-sample Blackman windows (0.4 ms) with 50% overlap between adjacent windows (5000 windows/s) and 1024-point zero-padded FFTs. Each spectrogram was scaled such that the median value was 0 dB.

The experiment with human experts used 1 ms windows because they produced the visually most distinct calls. Shorter windows produced blurry spectrograms due to wider spectral main-lobe width of the window, and longer windows blurred call frequency characteristics and obscured call endpoints. Feature extraction for the model-based detector was not affected by the wider main-lobe width of the 0.4 ms window, and smaller windows of frequency-modulated calls were more pure-tone-like, which better agreed with the underlying assumptions of spectral analysis.

Two GMMs were trained as part of the detector: one model for calls and a second model for the background noise. The training data consisted of 100 calls from each species, each one embedded in a 50 ms window of white noise at 0 dB SNR. Features extracted from within the endpoints of each call were used to train the call GMM; the same number of frames was randomly selected from all frames outside the call endpoints for each 50 ms window and used to train the background GMM. The GMMs were trained with full covariance matrices and five mixtures per model using ten training epochs. During testing, features from a 50 ms window were extracted and input into each GMM, producing a log likelihood. The maximum difference in log likelihoods between the call and background GMMs over all frames in the 50 ms window was compared to a likelihood threshold, and the detector determined that a call was detected if the maximum log likelihood difference exceeded the threshold. As with the energy-based detector, different thresholds controlled the tradeoff between sensitivity and specificity for the model-based detector.

Both automated detectors were tested using the same 1000-trial experiment administered to humans. Each detector performed the experiment in multiple trials to match the number of individuals who participated in the human detection experiment, and a different set of random calls and noise was generated for each trial. The energy-based detector used a set of ten energy thresholds (0.1–1.0 dB in 0.1 dB increments), while the model-based detector used a set of ten likelihood thresholds (−10 to 12.5 dB in 2.5 dB increments).

Lower thresholds produced more hits but more false positives, while higher thresholds produced fewer false positives but more misses. Results from the detectors were analyzed using an ANOVA to determine the effects of species, SNR, and threshold on detection accuracy. The dependent variable was the same one used to analyze human detection performance. The independent variables for species (five levels), SNR (ten levels), and threshold (ten levels) were all treated as fixed factors.

D. Optimal detector

A linear filter $h(n)$ of length K is used to detect a signal $s(n)$ embedded in white noise $w(n)$ by filtering

$$y(n) = \sum_{k=0}^{K-1} h(k)(s(n-k) + w(n-k)) = y_s(n) + y_w(n), \quad (3)$$

where $s(n)+w(n)$ is the noisy signal and $y(n)$ is the filter output. The output is decomposed into $y_s(n)$, the output due to $s(n)$, and $y_w(n)$, the output due to $w(n)$. If the filter output exceeds a threshold, the detector outputs that a call was detected; otherwise, no call was detected. The optimal filter maximizes the output SNR $|y_s|^2/E\{|y_w|^2\}$. The solution is the time-reversed signal of $s(n)$, called a *matched filter*.[17] For white noise $w(n)$,

$$h(k) = \alpha s(K-1-k), \quad k \in [0, K-1], \quad (4)$$

where α is an arbitrary scale factor. For our experiments, we want $y_w(n)$ to have the same variance for all species so that the same threshold may be applied to the output regardless of which species call, if any, was embedded in the noise. Setting the variance of $y_w(n)$ to unity,

$$E\{|y_w|^2\} = 1,$$

$$E\left\{ \left| \sum_{k=0}^{K-1} \alpha s(k) w(k) \right|^2 \right\} = 1,$$

$$\alpha^2 \left[\sigma^2 \sum_{k=0}^{K-1} s^2(k) \right] = 1.$$

For $w(n)$, $\sigma^2 = 1$, so

$$\alpha = \sqrt{\frac{1}{\sum_{k=0}^{K-1} s^2(k)}}. \quad (5)$$

We tested a matched-filter detector using the same procedure used to test the other detectors. The lower end of the range of SNR tested was decreased to −25 dB to accommodate the superior noise robustness of the optimal linear detector. We varied the filter output threshold from 2 to 10 in 0.1unit steps (81 thresholds total).

III. RESULTS

A. Human detection

Twelve human subjects, all familiar with spectrogram analysis of bat calls, took the test. The first 100 trials of each experiment were excluded from analysis of the results to reduce start-up effects associated with individuals becoming familiar with the test. All main factors were significant at $P<0.001$ (SNR: $F=147.1$, $d.f.=9$; species: $F=303.4$, $d.f.=4$; individual: $F=3.25$, $d.f.=11$) as was the interaction of SNR and species ($F=36$, $d.f.=21.9$). The significance of the individual factor means that performance varied across the population from which the participants were drawn (i.e., bat researchers). However, none of the interactions that included the individual factor were significant (SNR × individual: $F=0.79$, $d.f.=99$, $P=0.94$; species × individual: $F=1.25$, $d.f.=44$, $P=0.13$; SNR × species × individual; $F=1.05$, $d.f.=396$, $P=0.22$), indicating that individuals with higher accuracy were equally more accurate over all species and SNRs tested.

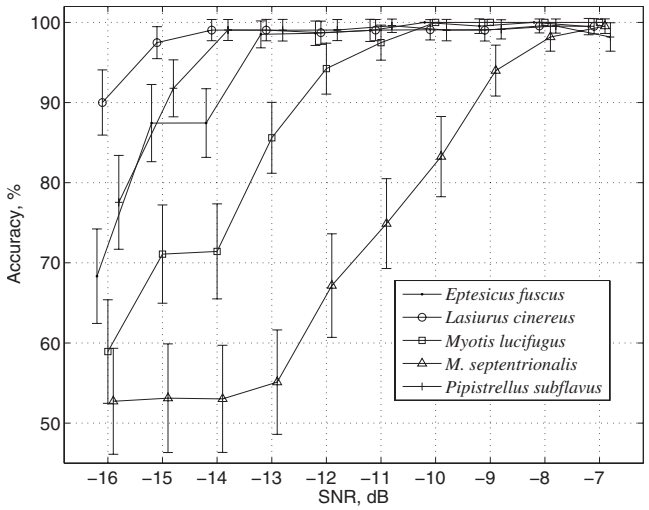


FIG. 3. Accuracy vs SNR for human experts from detection experiment. Results are means ±95% confidence intervals. Results are offset horizontally for clarity.

The significance of the species factor indicates that calls from some species were easier to detect than calls from other species. This result stems from the design of the experiment because the range of tested SNR values was not large enough for all species to exhibit the full range of detectability from random chance to near-perfect performance (Fig. 3).

The significance of the interaction between species and SNR indicates that at a fixed SNR, calls from some species were easier to detect than calls from other species. The long narrowband calls of *Lasiurus cinereus* were the easiest to detect against a white noise background, while the short broadband calls of *Myotis septentrionalis* were the most difficult to detect. The disparity of detectability among species for a given SNR reflects the fact that calls at the same signal power from different species have significantly different spectral peak levels (Table II). Narrowband signals have signal power condensed into a narrow range of frequency, resulting in larger spectral peaks compared to broadband signals. The larger spectral peaks extend farther above the white noise background and are easier to detect by visual inspection. The display settings also, to a lesser degree, affected human performance because long narrowband calls occupied significantly more display area (more display pixels) than short broadband calls.

TABLE II. Spectral peak values from spectrograms used in human detection experiment. Peak level is mean ± standard deviation over 100 spectrograms of each signal source. All signals scaled to unity power.

Source	Spectral peak (dB)
Eptesicus fuscus	40.8 ± 0.2
Lasiurus cinereus	41.6 ± 0.02
Myotis lucifugus	39.9 ± 0.4
M. septentrionalis	36.0 ± 0.5
Pipistrellus subflavus	41.5 ± 0.03
White noise	29.3 ± 0.6

B. Energy-based detector

All factors and interactions were significant at $P < 0.001$ (SNR: $F = 614.7$, $d.f. = 9$; species: $F = 399.7$, $d.f. = 4$; threshold: $F = 204.0$, $d.f. = 9$; SNR × species: $F = 25.8$, $d.f. = 36$; SNR × threshold: $F = 20.8$, $d.f. = 81$; species × threshold: $F = 11.4$, $d.f. = 36$; SNR × species × threshold: $F = 3.35$, $d.f. = 324$). The threshold factor replaced the individual factor from the human detection experiment; however, the significance of the threshold factor and all related interaction terms indicates that different thresholds produced significantly different detection accuracies when either species or SNR or both were held fixed. Maximum accuracy was achieved with a threshold of 0.5 dB. Thresholds below 0.5 dB produced more hits but also more false positives, resulting in overall lower accuracy. Similarly, thresholds above 0.5 dB produced more true negatives but also more misses. Calls from *Myotis septentrionalis* were the most difficult to detect for the energy-based detector (Fig. 4). Calls from *M. lucifugus* were the next difficult to detect (same for humans), while calls from the other three species were equally detectable.

Because all calls from all species at a given SNR were normalized to have equal power [Eq. (1)], longer calls contained more energy. For calls shorter than the 5 ms analysis window, all call energies contributed to the energy calculation, and shorter calls contributed less energy. For calls longer than the analysis window, only 5 ms of the call contributed to the energy calculation. Therefore, longer calls contributed nearly the same energy to the energy calculation, which explains why detection performance for the three species with longer calls was nearly identical. If a shorter analysis window was used, say 1 ms, then calls from all species at a given SNR would contribute nearly the same energy to the energy calculation and would be equally detectable. As long as an analysis window is "filled up" by call energy because the window is shorter than total call duration, the SNR [calculated using the entire call in Eq. (2)] does not depend on analysis window length. Therefore, we conclude that the performance curves in Fig. 4 for the three species with the longest calls are the best performance curves attainable by the energy-based detector *for any analysis window length*.

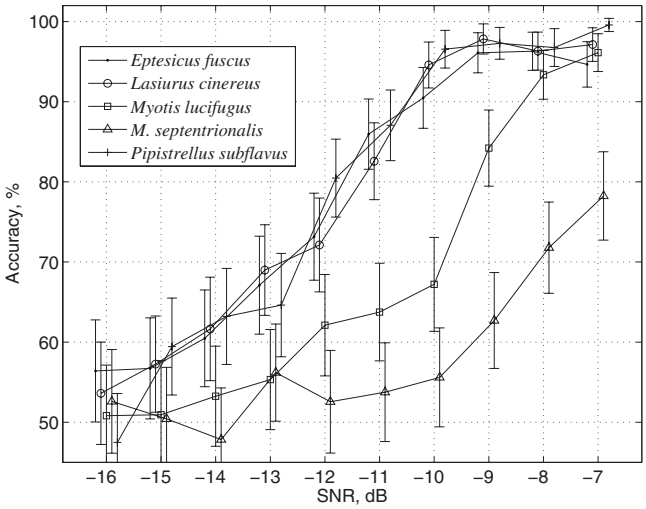


FIG. 4. Accuracy vs SNR for energy-based detector using an energy threshold of 0.5 dB. Results are means ±95% confidence intervals. Results are offset horizontally for clarity.

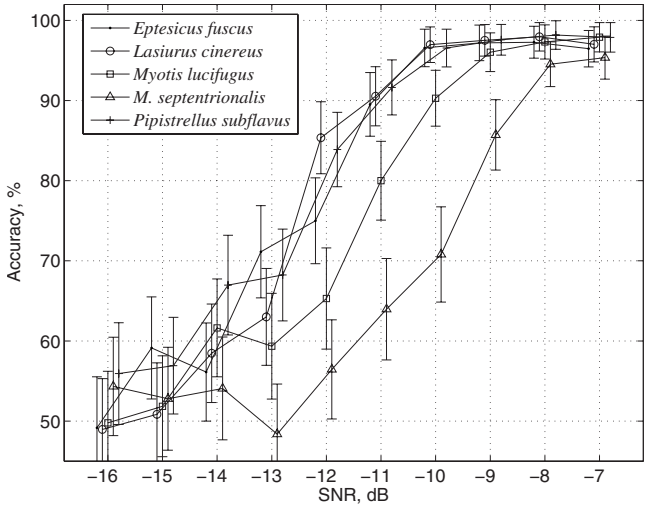


FIG. 5. Accuracy vs SNR for model-based detector using a likelihood threshold of 0 dB. Results are means ±95% confidence intervals. Results are offset horizontally for clarity.

C. Model-based detector

As with the energy-based detector, all factors and interactions were significant at $P < 0.001$ (SNR: $F = 1380$, $d.f. = 9$; species: $F = 267$, $d.f. = 4$; threshold: $F = 525$, $d.f. = 9$; SNR × species: $F = 21.5$, $d.f. = 36$; SNR × threshold: $F = 41.2$, $d.f. = 81$; species × threshold: $F = 6.2$, $d.f. = 36$; SNR × species × threshold: $F = 1.64$, $d.f. = 324$). Maximum accuracy was achieved at a threshold of 0 dB, which is not surprising because trials with and without calls had equal *a priori* probabilities in the experiment. Maximum accuracy for the model-based detector (76.3 ± 0.8%) was slightly higher than that for the energy-based detector (72.2 ± 0.8%) but significantly lower than that for humans (89.7 ± 0.6%). Similar to the energy-based detector, *Myotis septentrionalis* and *M. lucifugus* were harder to detect, while the other three species were nearly equally difficult to detect (Fig. 5).

D. Optimal detector

Maximum accuracy over the same SNR range as the other detectors was 98.4 ± 0.2% and was achieved at a threshold of 4.8. We extended the lower end of the SNR

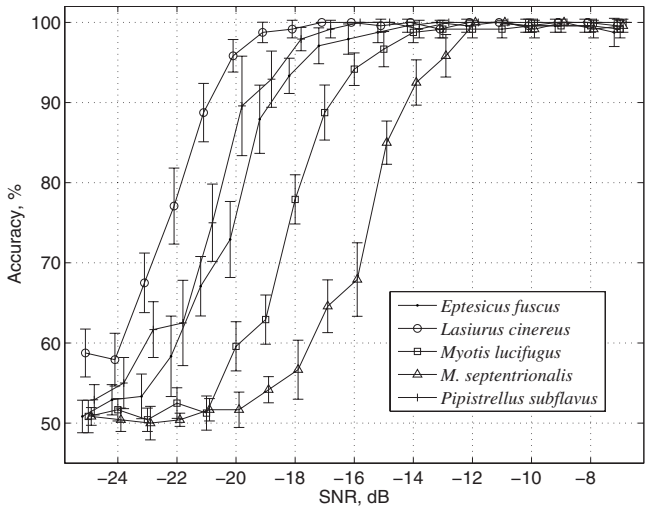


FIG. 6. Accuracy vs SNR for matched-filter detector using a likelihood threshold of 4.8. Results are means ±95% confidence intervals. Results are offset horizontally for clarity.

range to −25 dB to show the full range of accuracy for the matched-filter detector (Fig. 6). Performance varied among species, with calls of *Lasiurus cinereus* being the easiest to detect and calls of *Myotis lucifugus* being the most difficult to detect. Figure 7 shows the output of the matched-filter detector compared to other signal representations of a noisy call repeated at several SNRs. The noisy call of *M. lucifugus* at −10 dB SNR is not visible in the time domain and faintly visible in spectrogram form, but the output of the matched filter spikes well above the noise.

To explain the performance of the optimal filter, consider the matched-filter output SNR$=|y_s|^2/E\{|y_w|^2\}$. We set $E\{|y_w|^2\}=\sigma^2=1$ to solve for the matched-filter scale factor α in Eq. (5). From Eq. (3),

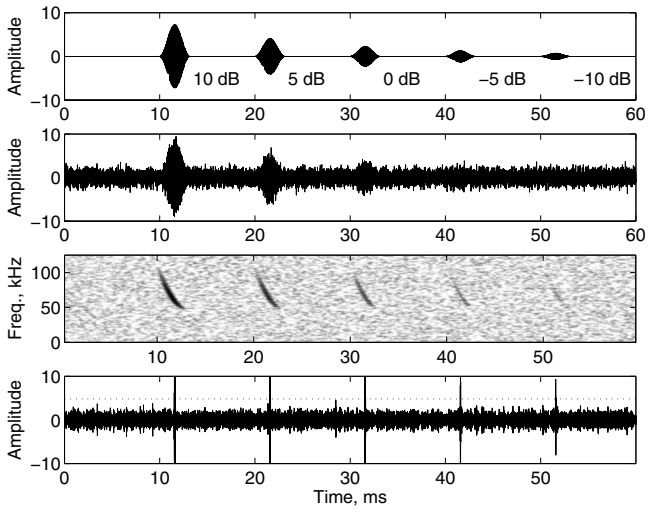


FIG. 7. Examples of signal representations in noise. Top pane: Noise-free time-domain signal of a call of *Myotis lucifugus* repeated at various SNRs. Second pane: White noise of zero mean and unity variance added to the noise-free signal from the top pane. Third pane: Spectrogram of the signal in second pane (spectrogram parameters same as those used in human detection experiment). Bottom pane: Matched-filter output. The horizontal dotted line marks a threshold of 4.8. Note that the variance of the noise signals in the second and bottom panes are both unity.

TABLE III. SNR (dB) at 90% accuracy. More negative values denote better performance. Results for humans, the energy-based detector, the model-based detector, and the optimal detector were tabulated using the results in Figs. 3–6, respectively. * denotes extrapolated value.

Species	Human	Energy	Model	Optimal
Eptesicus fuscus	−14.0	−10.1	−10.9	−18.6
Lasiurus cinereus	−16.1	−10.4	−11.1	−20.8
Myotis lucifugus	−12.5	−8.4	−10.0	−16.8
M. septentrionalis	−9.3	−5.0*	−8.5	−14.3
Pipistrellus subflavus	−14.9	−10.7	−11.2	−19.9

$$y_s(n)=\sum_{k=0}^{K-1} h(k)s(n-k).$$

When the filter and signal overlap,

$$|y_s|^2=\left|\sum_{k=0}^{K-1}\alpha s(k)s(k)\right|^2$$

$$=\left|\frac{1}{\sqrt{\Sigma_{j=0}^{K-1}s^2(j)}}\sum_{k=0}^{K-1}s^2(k)\right|^2=\sum_{k=0}^{K-1}s^2(k).$$

Combining with Eq. (1),

$$|y_s|^2=\sum_{k=0}^{K-1}\left(\frac{1}{\sqrt{1/K\Sigma_{j=0}^{K-1}x^2(j)}}10^{\mathrm{SNR}/20}x(k)\right)^2$$

$$=\frac{K}{\Sigma_{j=0}^{K-1}x^2(j)}10^{\mathrm{SNR}/10}\sum_{k=0}^{K-1}x^2(k)=K10^{\mathrm{SNR}/10}. \quad (6)$$

Thus, $|y_s|^2$ scales with call length K, and a doubling of K allows for a decrease in SNR by 3 dB for the same $|y_s|^2$.

E. Performance summary

Table III summarizes the performance of the four detectors for each species. The SNR at which 90% accuracy was achieved was determined for each detector using linear interpolation of mean values from the plots of accuracy versus SNR. The optimal linear detector outperformed human experts by 4–5 dB, which outperformed the model-based detector by 1–5 dB depending on species. The model-based detector outperformed the broadband energy detector by 0.5–3.5 dB depending on species.

F. Speed

We measured the response time for each trial in the human detection experiment as well as the computation time for both automated methods. While computation time is heavily dependent on available computer resources and algorithm efficiency, our findings report the current speed advantage for existing computer resources that the automated methods have over humans. We consider our algorithms to be rather highly efficient (i.e., extensive use of matrix computation and compiled MEX functions in MATLAB), although MATLAB, being an interpreted computer language, is considered slow compared to compiled languages such as C++.

Humans answered the last 900 trials of the experiment in 16.5 min on average with a range 12.1–25.7 min. The median time to respond to each trial was 1.17 s with a range 0.92–1.59 s, and linear regression between accuracy and median time showed significant correlation (R^2=0.43, P <0.05). The energy-based detector took approximately 10 s to detect calls in 1000 50 ms windows, which is about 5 times faster than real time and about 110 times faster than humans. About 90% of the energy-based detector computation time was spent on filtering. The high-pass filter improved the SNR for the energy-based detector and is essential for real recordings, which are typically dominated by low-frequency acoustic noise. The model-based detector took approximately 110 s to detect calls in 1000 windows, about 11 times slower than the energy-based detector and 10 times faster than humans. Thus, the energy-based detector was the fastest yet least accurate detector; the model-based detector was slightly more accurate yet an order of magnitude slower, and the performance of humans was significantly more accurate yet another order of magnitude slower. Humans and the model-based detector have the additional advantage over the energy-based detector of being in position to further analyze the detected calls (e.g., determine species).

G. Effective detection range

We applied our results to estimate the effective detection range of the species under test when recorded in the field to quantify detection accuracy in more practical terms. Calls attenuate as they propagate away from bats because of (1) spherical spreading and (2) atmospheric absorption. Acoustic energy density decreases 6 dB for each distance doubling due to spherical spreading (frequency independent). The atmosphere absorbs acoustic energy roughly linearly in frequency from 0.9 dB/m at 30 kHz to 3.3 dB/m at 100 kHz at standard temperature, humidity, and pressure.[18] The limiting factor in detecting calls at range is noise, including background acoustic noise, electronic noise, and quantization noise from sampling. From our experiment results, we found that detection accuracy varied with SNR as well as call characteristics (species) and detection method. By combining our results with absorption curves and known species call intensities, we created a useful tool for estimating effective detection range.

Figure 8 shows the attenuation curves for our test species versus range. The curves include intensity of search-phase calls in dB SPL re 20 μPa at 10 cm: (1) *Eptesicus fuscus*, 120 dB;[19] (2) *Lasiurus cinereus*, 110 dB; (3) *Myotis lucifugus*, 103 dB;[20] (4) *M. septentrionalis*, 102 dB;[20] and (5) *Pipistrellus subflavus*, 110 dB. The curves were estimated by constructing 100 synthetic calls for each species at the appropriate SPL at 10 cm and calculating SPL at range by scaling the calls to account for spherical spreading and filtering the calls to account for atmospheric absorption. SPL at range was determined by calculating signal power after filtering and scaling the original call, the so-called "rms" version of SPL, which is consistent with the ANSI standard definition.[16] Our attenuation curves include the assumption

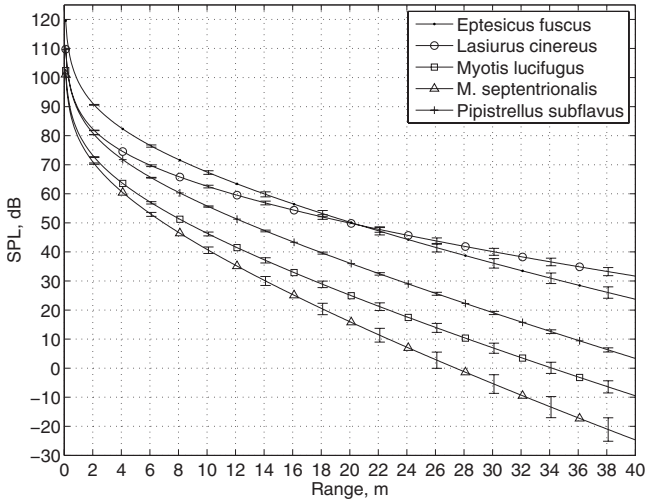


FIG. 8. Attenuation curves vs range for synthetic calls. Attenuation includes atmospheric absorption and spherical spreading. Acoustic intensity at 10 cm for each species was taken from literature. Error bars show standard deviation over 100 random calls for each species.

that the bat and microphone were omnidirectional with a flat transfer function.

Range varied widely across species. For example, assume that a recording system had a noise level equivalent to 50 dB SPL. Calls from *Lasiurus cinereus* were detectable by humans with at least 90% accuracy down to 34 dB SPL (−16 dB SNR; Table III) at a range of up to 35 m. Using the energy-based detector (−10 dB SNR), range dropped to 30 m. Calls from *Myotis septentrionalis* were detectable by humans down to 41 dB SPL (−9 dB SNR) at a range of up to 10 m. Using the energy-based detector (−5 dB SNR), range dropped to 8 m. These measurements assume that the spectral tilt toward low frequency created by the atmospheric absorption filter does not affect human detection performance. Our experiments did not use atmospheric absorption filters; we set call SNR without spectral tilt (independent of range).

IV. DISCUSSION

Detecting calls by hand is labor intensive and tedious. Such difficulties are a bottleneck to current acoustic-based bat research given the availability of high-quality commercial recording systems, which easily generate an overwhelming amount of data for analysis. Researchers cope by either reducing the amount of data collected or discarding data based on a "usefulness" criterion (e.g., the five strongest calls in a pass per minute of recording). Reducing the amount of data collected limits the scope of field experiments and places a burden on the design of such experiments, especially for studies involving difficult-to-record or rare species. Commercial recording systems are able to record continuously all night long (limited by battery capacity) over an entire field season (limited by data storage capacity) so that activity can be monitored essentially uninterrupted for months at a time. Limiting the amount of data collecting for analysis reasons does not fully utilize the capabilities of the recording system. The usefulness criterion, on the other hand, selects a subset of all detectable calls according to their

perceived value for analysis. The problem with this approach is that the perceived value of a call or pass of calls may not match the true value, which introduces the bias of the researcher into the analysis. Put another way, conclusions based on the subset of useful calls may not generalize to the whole population of detectable calls.

Automated methods offer an alternative to hand analysis that overcomes many of the limitations of human analysis. Automated methods are tireless, fast, and apply equal rigor in call detection from the first minute of recorded data to the last. In a classification experiment, Jennings *et al.*[21] showed that neural networks classified individual bat calls to genus and species as accurately as human experts. Our experiments showed that the difference in detection accuracy between humans and automated methods translated into approximately 20% reduction in effective detection range, but our experiments did not include several real-world issues involving field recordings. Nonstationary noise sources, echoes, and overlapping calls all contribute to automated detection errors and lead to a shorter effective detection range. Model-based detection can reduce detection errors compared to an energy-based detector because the detection models can be trained to discriminate between calls and echoes and other noise sources.[22] Our results showed a slight increase in accuracy for the model-based detector compared to the energy-based detector using only a white noise source, while a previous study using real field recordings showed a wider gap in accuracy.[14]

One reason humans were significantly more accurate for longer calls while only slightly more accurate for shorter calls compared to the model-based detector is that the model-based detector only employed local (frame-based) information while humans could consider local and global features of each call. For short calls, the neighborhood of local features was a significant portion of the entire call but a smaller portion for longer calls. The model-based detector could be improved by incorporating more global information into the model. This may be achieved by using more than ± 3 frames for calculating temporal derivatives. The drawback to using more frames is that derivative estimates at the ends of calls are more affected by frames of background noise. The gap in accuracy between the energy-based and model-based detectors may be considered conservative given that the noise source used in our experiments contained no burst noise or nonstationary noise. Such noise sources are more likely to produce false positives for the energy-based detector compared to the model-based detector because the latter uses more than just spectral energy (e.g., frequency, slope, and smoothness) to differentiate between calls and noise.

Several factors affect effective detection range for an echolocating bat. The most significant is directionality of the bat and the microphone. Bats are not omnidirectional, as assumed in our experiments, but have frequency-dependent directional patterns that focus acoustic energy in regions in front of them. Ghose and Moss[23] measured the beam width of several hunting *Eptesicus fuscus* and found that SPL dropped by 3 dB at an angle of 35° from the center line of the main lobe of intensity. Zhuang and Müller[24] showed that SPL varied significantly with frequency and position around the head of a simulated *Rhinolophus rouxi*, forming deep spectral notches and side lobes that may be 30 dB below the peak main-lobe intensity (personal communication). Direction-dependent attenuation of SPL results in a vertical shift downwards for the curves in Fig. 8. From the example above, if a call from *Lasiurus cinereus* was attenuated by 25 dB due to directionality, then the effective detection range would drop from 35 to 12 m. The main lobe of intensity tends to decrease in width as frequency increases, making high-frequency bats much more difficult to detect. Microphones also have directionality, and main-lobe width also tends to decrease as frequency increases. For example, the CM16 condenser microphone from Avisoft at 30° off axis drops in sensitivity by 5 dB at 40 kHz, 14 dB at 60 kHz, and 20 dB at 80 kHz.[25] The drop in sensitivity due to microphone directionality may be minimized by manually aiming the microphone at a moving bat while recording.[3]

Our experiments measured the performance of humans and automated methods to detect calls, a subset of which may be considered useful calls depending on the information needed from the acoustic data. Relevant call information includes species, frequency characteristics, and harmonic amplitudes. For any detected call, FME may be estimated reliably since the FME has the highest SNR for any part of a call (assuming flat stationary noise). In our experiments, FME may be used to estimate species although overlapping frequency ranges limit classification accuracy (Table I). Note that FME may be estimated by humans detecting calls from spectrograms and by the model-based detector but not by the energy-based detector since no frequency information is used by that detector. Skowronski and Fenton[22] quantified call feature estimation accuracy versus SPL and showed that FME and minimum call frequency were most accurately estimated from detected calls, and estimate errors steadily increased as detected SPL decreased. Error for maximum frequency estimates was higher due to increased atmospheric absorption.

The results of the matched-filter detector are an upper limit of performance for linear methods of detection, and they put the performances of the other detectors in perspective. The 4–5 dB gain above human performance translates into an increase in effective detection range of 2 m for *M. septentrionalis* and 5 m for *L. cinereus*. The matched filter is not a realizable filter because the method requires knowledge of the target noise-free call, which is not available in practice. However, near-optimal performance is attainable if the filter *nearly* matches the target call, and a bank of filters spanning the range of expected target calls would provide near-optimal performance over the range. Degradation from optimal performance depends on the correlation properties of the target signal and the filter bank design,[26] which also affects computational cost of the detector.

V. CONCLUSIONS

In our experiments, humans were more accurate but slower compared to automated methods in detecting synthetic calls embedded in additive white noise. The energy-based detector was fastest (five times faster than real time)

but least accurate ($72.2 \pm 0.8\%$) when averaged over all species and SNRs, while the model-based detector was modestly more accurate ($76.3 \pm 0.8\%$) but slower (two times slower than real time). Humans were significantly more accurate ($89.7 \pm 0.6\%$) than either automated method but much slower (20 times slower than real time). High-bandwidth calls were more difficult to detect compared to low-bandwidth calls because low-bandwidth calls have energy concentrated in a narrower bandwidth, resulting in higher spectral peaks, which extend further above the white noise spectrum compared to high-bandwidth calls. Calls from *Lasiurus cinereus* were the easiest for humans to detect and were detectable with 90% accuracy at −16 dB SNR, while calls from *Myotis septentrionalis* were the most difficult for humans to detect and were detectable with 90% accuracy at −9 dB SNR. The automated methods required between 0.8 and 5 dB higher SNR for comparable accuracy with humans depending on species. A matched-filter detector outperformed humans by 4–5 dB, indicating the upper bound in performance by linear automated detection algorithms. Variation in human performance was partially explained by median time per test trial ($R^2 = 0.43$, $P < 0.05$), which demonstrates that a tradeoff exists between human accuracy and speed.

Our results may be used to estimate maximum effective detection range for the species under test, assuming omnidirectional bats and microphone, for a given detection method and recording equipment noise level. Automated methods decreased effective detection range by approximately 20% for all species compared to human detection. Directionality attenuation and a threshold for useful detected calls would decrease the effective detection range depending on application. Knowing the effective detection range is important for designing and performing acoustic field measurements and for interpreting the results from call analysis. Range varied across the five species in our experiments by a factor of 3 for low-noise recording environments and by a higher factor for noisier environments, which translates into a range of effective detection *volume* of at least a factor of 27. Thus, a recording system would cover a much larger volume of space for species with easily detected calls compared to species whose calls are more difficult to detect, which impacts the interpretation of call analysis in terms of activity level and species abundance.

ACKNOWLEDGMENTS

The current work was funded by the International Fellowship Research Program from the National Science Foundation. The authors would like to thank the members of the Fenton Laboratory for participating in the human detection experiments.

[1]D. R. Griffin, *Listening in the Dark: The Acoustic Orientation of Bats and Men* (Comstock, Ithaca, NY, 1958).
[2]M. B. Fenton and G. P. Bell, "Recognition of species of insectivorous bats by their echolocation calls," J. Mammal. **62**, 233–243 (1981).
[3]M. K. Obrist, "Flexible bat echolocation: The influence of individual, habitat and conspecifics on sonar signal design," Behav. Ecol. Sociobiol. **36**, 207–219 (1995).
[4]S. Parsons, C. W. Thorpe, and S. M. Dawson, "Echolocation calls of the long-tailed bat: A quantitative analysis of types of calls," J. Mammal. **78**, 964–976 (1997).
[5]M. J. O'Farrell, B. W. Miller, and W. L. Gannon, "Qualitative identification of free-flying bats using the Anabat detector," J. Mammal. **80**, 11–23 (1999).
[6]S. Parsons, A. M. Boonman, and M. K. Obrist, "Advantages and disadvantages of techniques for transforming and analyzing chiropteran echolocation calls," J. Mammal. **81**, 927–938 (2000).
[7]N. Vaughan, G. Jones, and S. Harris, "Habitat use by bats (chiroptera) assessed by means of a broad-band acoustic method," J. Appl. Ecol. **34**, 716–730 (1997).
[8]C. Korine and E. K. V. Kalko, "Fruit detection and discrimination by small fruit-eating bats (Phyllostomidae): Echolocation call design and olfaction," Behav. Ecol. Sociobiol. **59**, 12–23 (2005).
[9]L. Pettersson, "Bat specific signal analysis software—BatSound," in *Bat Echolocation Research: Tools, Techniques and Analysis*, edited by R. M. Brigham, E. K. V. Kalko, G. Jones, S. Parsons, and H. J. G. A. Limpens (Bat Conservation International, Austin, TX, 2004), pp.130–133.
[10]W. M. Masters, S. C. Jacobs, and J. A. Simmons, "The structure of echolocation sounds used by the big brown bat *Eptesicus fuscus*: Some consequences for echo processing," J. Acoust. Soc. Am. **89**, 1402–1413 (1991).
[11]S. Parsons and G. Jones, "Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks," J. Exp. Biol. **203**, 2641–2656 (2000).
[12]M. K. Obrist, R. Boesch, and P. F. Fluckiger, "Variability in echolocation call design of 26 Swiss bat species: Consequences, limits and options for automated field identification with a synergetic pattern recognition approach," Mammalia **68**, 307–322 (2004).
[13]S. Bayefsky-Anand, M. D. Skowronski, M. B. Fenton, C. Korine, and M. W. Holderied, "Variations in the echolocation calls of the European free-tailed bat," J. Zool. **275**, 115–123 (2008).
[14]M. D. Skowronski and J. G. Harris, "Acoustic microchiroptera detection and classification using machine learning: Lessons learned from automatic speech recognition," J. Acoust. Soc. Am. **119**, 1817–1833 (2006).
[15]M. D. Skowronski and M. B. Fenton, "Model-based detection of synthetic bat echolocation calls using an energy threshold detector for initialization," J. Acoust. Soc. Am. **123**, 2643–2650 (2008).
[16]*American National Standard Acoustical Terminology* (American National Standards Institute, New York, NY, 1994), S1.1-1994.
[17]S. Haykin, *Communication Systems*, 4th ed. (Wiley, New York, NY, 2000).
[18]H. E. Bass, L. C. Sutherland, and A. J. Zuckerwar, "Atmospheric absorption of sound: Update," J. Acoust. Soc. Am. **88**, 2019–2021 (1990).
[19]M. E. Jensen and L. A. Miller, "Source levels of bat biosonar signals measured in the field using microphone arrays," in *Proceedings of the 25th Gottingen Neurobiology Conference* (Georg Thieme-Verlag, Stuttgart, 1997), Vol. **2**, p. 361.
[20]L. A. Miller and A. E. Treat, "Field recording of echolocation and social signals from the gleaning bat *Myotis septentrionalis*,"Bioacoustics **5**, 67–87 (1993).
[21]N. Jennings, S. Parsons, and M. J. O. Pocock, "Human vs. machine: Identification of bat species from their echolocation calls by humans and by artificial neural networks," Can. J. Zool. **86**, 317–377 (2008).
[22]M. D. Skowronski and M. B. Fenton, "Model-based automated detection of echolocation calls using the links detector," J. Acoust. Soc. Am. **124**, 328–336 (2008).
[23]K. 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).
[24]Q. Zhuang and R. Muller, "Numerical study of the effect of the noseleaf on biosonar beamforming in a horseshoe bat," Phys. Rev. E **76**, 051902 (2007).
[25]Avisoft Bioacoustics, CM16 datasheet, URL: http://www.avisoft.com/usg/cm16_cmpa.htm (last viewed 10/2/2008).
[26]R. Hecht-Nielsen, "Nearest matched filter classification of spatiotemporal patterns," Appl. Opt. **26**, 1892–1899 (1987).

Bone sonometry: Reducing phase aberration to improve estimates of broadband ultrasonic attenuation

Adam Q. Bauer, Christian C. Anderson, Mark R. Holland, and James G. Miller
Washington University, Physics, Saint Louis, Missouri 63130

(Received 3 September 2008; revised 29 October 2008; accepted 3 November 2008)

Previous studies suggest that phase cancellation at the receiving transducer can result in the overestimation of the frequency dependent ultrasonic attenuation of bone, a quantity that has been shown to correlate with bone mineral density and ultimately with osteoporotic fracture risk. Evidence supporting this interpretation is provided by phase insensitive processing of the data, which appear to reduce the apparent overestimates of attenuation. The present study was designed to clarify the components underlying phase aberration artifacts in such through-transmission measurements by conducting systematic studies of the simplest possible test objects capable of introducing phase aberration. Experimental results are presented for a Lexan phantom over the frequency range 300–700 kHz and a Plexiglas phantom over the 3–7 MHz range. Both phantoms were flat and parallel plates featuring a step discontinuity milled into one of their initially flat sides. The through-transmitted signals were received by a 0.6 mm diameter membrane hydrophone that was raster scanned over a grid coaxial with the transmitting transducer. Signals received by the pseudoarray were processed offline to emulate phase sensitive and phase insensitive receivers with different aperture diameters. The data processed phase sensitively were focused to demonstrate the results of planar, geometrical, and correlation-based aberration correction methods. Results are presented illustrating the relative roles of interference in the ultrasonic field and phase cancellation at the receiving transducer in producing phase aberration artifacts. It was found that artifacts due to phase cancellation or interference can only be minimized with phase insensitive summation techniques by choosing an appropriately large receiving aperture. Data also suggest the potentially confounding role of time-and frequency-domain artifacts on ultrasonic measurements and illustrate the advantages of two-dimensional receiving arrays in determining the slope of attenuation (nBUA) for the clinical assessment of osteoporosis.
 [DOI: 10.1121/1.3035841]

PACS number(s): 43.80.Ev, 43.80.Jz, 43.80.Qf, 43.80.Vj [FD] Pages: 522–529

I. INTRODUCTION

Phase cancellation at a phase-sensitive (piezoelectric) receiver has been known to result in artifacts in measurements of broadband ultrasonic attenuation (BUA) in a number of studies of cancellous bone.[1–7] (Normalized broadband attenuation or nBUA is the slope of a line fit to the measured attenuation coefficient plotted as a function of frequency, a quantity sometimes referred to as the slope of attenuation.) Phase cancellation effects result in the overestimation of BUA, a quantity that has been shown to correlate with bone mineral density[1,8–14] and therefore osteoporotic fracture risk.[15–22]

Measurements of bone are complicated by many factors associated with their inhomogeneous character and irregular shapes, making it difficult to sort out potential physical mechanisms underlying phase aberration artifacts. *In vitro* studies of calcaneus bone samples by Wear[6] and Strelitizki *et al.*[3] and studies in human subjects by Wear[4] and Petley *et al.*[2] appear to indicate that phase insensitive processing of array data yields improved estimates of the true ultrasonic attenuation. Studies by Langton *et al.*[7,23] and by Xia *et al.*[5] suggest that phase aberration artifacts in cancellous BUA measurements are exacerbated by the surrounding cortical bone layer. Techniques by Ermert *et al.*,[24] Defontaine *et al.*,[25] and Eddin *et al.*[26] incorporating the use of phased arrays, designed to adaptively correct for phase aberration in through-transmission and backscatter measurements of cancellous bone, support the hypothesis that phase aberration is a source of potential measurement artifacts.

The present study was carried out on simple phantoms whose intrinsic attenuation coefficients could be measured independently. This approach provides a starting point for differentiating between contributions to phase aberration artifacts from two sources, phase cancellation at the receiving transducer and interference effects in the ultrasonic field.

Interference in the ultrasonic field results from a spatial redistribution of energy without energy loss. That is, for any local reduction in the average amplitude of the pressure field (due to partial or total destructive interference) there will always be another region of higher-than-average amplitude of the pressure field. Energy is only ostensibly lost to interference; it may have simply been redistributed, potentially missing the receiving transducer.

In contrast, phase cancellation occurring at the surface of a phase sensitive receiver is an irreversible loss of energy, but is solely an instrumental effect. Local compressions and expansions produce surface charge distributions of opposite sign in the piezoelectric receiving transducer's conductive plating.[27] The associated currents in these electrodes result in

an irrecoverable loss of energy as a result of current flow among locally produced positive and negative electrical signals.

The purpose of the present study is to illustrate how the size of the receiving aperture influences estimates of apparent attenuation as a result of diffraction and interference occurring in the field and phase cancellation occurring at the surface of a piezoelectric receiver. That is, artifacts due to phase cancellation or interference can only be minimized with phase insensitive summation techniques by appropriate choice of receiving aperture. In the present paper, through-transmission measurements of the apparent attenuation properties are conducted over two bandwidths, 300–700 kHz and 3–7 MHz, on the simplest possible test objects capable of introducing phase aberration. The two phantoms studied were flat and parallel Lexan and Plexiglas plates featuring a step discontinuity milled into one of their initially flat sides. Data were acquired over a two-dimensional receiving pseudoarray and processed phase sensitively and phase insensitively to emulate receiving apertures with different diameters. The phase sensitive data were further processed to represent a planar, spherical, and correlation-focused receiving phase sensitive transducer.

II. METHODS

A. Sample preparation

The two phase aberrating phantoms used in this investigation were constructed from single sheets of poly(methylmethacrylate) (Plexiglas) and polycarbonate resin thermoplastic (Lexan) initially machined into flat and parallel pieces. The piece used for the data collected over the bandwidth centered at 5 MHz was 50.8 mm (2 in.) long, 50.8 mm (2 in.) wide, and 12.7 mm (0.5 in.) thick. The piece used for data collected over a bandwidth centered at 500 kHz was 114.3 mm (4.5 in.) long, 114.3 mm (4.5 in.) wide, and 19.1 mm (0.625 in.) thick. Step discontinuities were then milled into the surfaces of both plates, resulting in each plate having two distinct thicknesses. For each plate the thinner of these two sides will be referred to as the thin side and the thicker of the two sides will be referred to as the thick side. The step milled into the Plexiglas sample for the 5 MHz data was approximately 0.32 mm (0.0125 in.) and the step milled into the Lexan sample for the 500 kHz data was approximately 4.6 mm (0.181 in.). These step sizes were chosen so that phase cancellation artifacts occurring at the surface of a phase sensitive receiver would arise at approximately the midbandwidth of each frequency range, 500 kHz and 5 MHz.[28,29]

B. Experimental methods

1. Data centered about 500 kHz

Baseline attenuation measurements of Lexan over a bandwidth of 300–700 kHz were made using a matched pair of planar, 28.575 mm (1.125 in.) diam, single-element transducers nominally centered at 500 kHz (Panametrics v391, Waltham, MA). The two transducers were separated by 127 mm (5 in.), approximately twice the near field distance of either transducer. The transmitting transducer was excited by a single-cycle 500 kHz tone burst from a pulse/function generator (HP 8116, Palo Alto, CA) passed through a 50 dB gain radio frequency (rf) amplifier (ENI 240L, Rochester, NY) and a Diplexer (Ritec, Warwick, RI). The received signal was sent to the 50 Ω input of an 8-bit digitizing oscilloscope (Techtronix 5052b, Beaverton, OR). The resulting time-domain signals were digitized at a rate of 125 MS/s (800 ns/pt) and temporally averaged 256 times.

The attenuation coefficient of the Lexan sample was measured at 25 individual sites over the flat and parallel Lexan plate (over a 5 by 5 site grid in the center of the sample) in through transmission mode to average over slight thickness variations occurring in the sample. These 25 measurements were then averaged to yield the true (unaberatted) attenuation coefficient over this bandwidth (300–700 kHz).

Apparent attenuation measurements of the stepped Lexan plate were performed using a through-transmission setup with the same 28.575 mm planar single-element transducer on transmit and a polyvinylidene fluoride (PVDF) membrane hydrophone (Sonic Industries, model 804, now Sonora Medical Systems, Longmont, CO) on receive. The stepped plastic plate was positioned so that the transmitted field was approximately "split in half" by the step, i.e., approximately 50% of the field insonified the thick part of the plate, and approximately 50% of the field insonified the thin part of the plate. This location has been found to maximize phase cancellation artifacts occurring at the surface of a phase sensitive receiver when transmitting through Plexiglas,[29] the subject of the current investigation. The membrane hydrophone scanned the transverse receive plane in steps of 0.4 mm in a raster scan axially aligned with the transmitter from over a 57.6 by 57.6 mm area in the elevational and azimuthal directions. The transmitting transducer was excited by a single cycle 500 kHz tone burst from a pulse/function generator (HP 8116, Palo Alto, CA) passed through a 50 dB gain rf amplifier (ENI 240L, Rochester, NY). The active element of the hydrophone was 0.6 mm in diameter. The transmitting transducer and the receiving membrane hydrophone were separated by 127 mm (5 in.), and the stepped plastic plate was placed approximately 64.5 mm (2.5 in.) away from the transmitter.

2. Data centered about 5 MHz

Baseline measurements of the attenuation coefficient of Plexiglas over a bandwidth of 3–7 MHz were carried out using the through transmission setup described in Bauer *et al.*[28] The attenuation coefficient was measured over 25 sites on the flat and parallel Plexiglas plate before it was milled using separate transmitting and receiving 6.35 mm (0.25 in.) diam, planar, immersion transducers nominally centered at 5 MHz. The transmitting transducer was excited by a pulser/receiver (Panamtrics 5800 Waltham, MA) and the propagation path between the two transducers was approximately 76.2 mm (3 in.).

For the data collected over the stepped region of the plastic plate, the 6.35 mm (0.25 in.) transmitting transducer was excited by the Panametrics 5800 pulser/receiver above, and similar to the Lexan sample above, the incident ultra-

sonic field was approximately split in half by the stepped plastic plate with approximately 50% of the field insonified the thick part of the plate, and approximately 50% of the field insonified the thin part of the plate to maximize phase cancellation artifacts. The PVDF membrane hydrophone described above was scanned over a 12.8 mm by 12.8 mm area in the transverse receive plane in steps of 0.1 mm across the through-transmitted field.

C. Data analysis

For both bandwidths investigated in this study, the amplitude attenuation coefficients of the flat and parallel and the stepped plastic plates were determined from a log-spectral subtraction technique

$$\alpha(\omega) = \frac{20\log(|\tilde{U}_w(\omega)|) - 20\log(|\tilde{U}_s(\omega)|) + 10\log(T_{\mathrm{ws}}^I T_{\mathrm{sw}}^I)}{d}, \tag{1}$$

where $|\tilde{U}_w(\omega)|$ and $|\tilde{U}_s(\omega)|$ are the magnitudes of the measured response of the water (reference) and sample paths, and T_{ws}^I and T_{sw}^I are the intensity transmission coefficients of the water-sample and sample-water interfaces, respectively. The phase velocity and density were experimentally measured for the flat and parallel samples; these values were then used as the input to the formula $[4Z_1Z_2/(Z_1+Z_2)^2]$ to specify the intensity transmission coefficient for all attenuation coefficient calculations. For the stepped plastic plates, the thickness d used in the calculation was the average thickness of the thin and thick sides of each plate. The responses were calculated offline by Fourier analysis of the processed digitized time-domain signals over the bandwidths investigated using the focusing and summation algorithms described below and in the Appendix.

1. Summation of received signals

Phase sensitive analysis. Phase sensitively summing the individual rf waveforms within the two-dimensional array emulates the response of a larger aperture piezoelectric receiver. The signals recorded at each array position are approximately what would be incident on a local region of this large diameter, single element piezoelectric transducer placed at a specific location in the field. For the phase sensitive analysis performed in this study, the rf signals recorded at each position in the array were focused one of three ways (planar, spherical, and correlation as described in the Appendix), summed, and then normalized by the number of rf signals used for the sum. This procedure produces a single rf waveform for the whole array. This waveform, $U_{\mathrm{array}}(t_k)$, can be written as $U_{\mathrm{array}}(t_k)=(1/N)\Sigma_{ij}U_{ij}(t_k)$, where $U_{ij}(t_k)$ is the focused (time-shifted) rf signal at the ijth position in the array and N is the number of positions in the array used in the sum. Dividing the sum by N normalizes all later analysis to the area one of the array elements. For the 500 kHz data, sampled rf signals within a radius of 28.575 mm (1.125 in.) from the center of the pseudoarray were used in all subsequent analyses in order to simulate a 57.15 mm (2.25 in.) diameter receiving transducer. For the 5 MHz data, rf signals within a radius of 6.35 mm (0.25 in.) from the center were used in all subsequent analyses in order to simulate a 12.7 mm (0.5 in.) diameter receiving transducer. Therefore, $N=16\,029$ for the 500 kHz data and $N=12\,645$ for the 5 MHz data. After the signals were focused and phase sensitively summed, a fast Fourier transformer (FFT) was performed on $U_{\mathrm{array}}(t_k)$ to obtain the squared magnitude of the frequency domain responses for the water and sample paths used in the amplitude attenuation coefficient calculation in Eq. (1).

Phase insensitive analysis. Phase insensitive analysis was performed by calculating the squared magnitude of the frequency domain response obtained at each location in the two-dimensional pseudoarray and then summing these power spectra over the pseudoarray for the different aperture diameters described below. The power spectrum recorded at each array position is approximately what would be incident on a small, finite area of a larger power detector, or acoustoelectric receiver.[27] The phase insensitive analysis performed in this study did not include any of the focusing techniques employed in the phase sensitive analysis because when a function is delayed (or advanced) by an amount Δt_{ij}, its Fourier transform is multiplied by a factor $e^{\mp i\omega\Delta t_{ij}}$.[30] Because the magnitude of the frequency domain response was calculated at each location before summation, and the temporal shifts were not of sufficient length to move the acquired temporally localized rf signals outside of the digitized record, the phase insensitive responses would not have been affected by temporal focusing.

The resultant phase insensitive frequency domain response of the array can then be written as

$$|\tilde{U}_{\mathrm{array}}(\omega)|^2 = \frac{1}{N}\sum_{ij}\left(|FFT[U_{ij}(t_k)]|\right)^2. \tag{2}$$

For the 500 kHz data, masks of two different sizes were implemented in the phase insensitive data analysis to simulate two different diameter receiving apertures: 28.58 mm (1.125 in.) and 57.15 mm (2.25 in.). For the 5 MHz data, masks of two sizes were implemented in the phase insensitive data analysis to simulate two different diameter receiving apertures: 6.35 mm (0.25 in.), and 12.7 mm (0.5 in.).

Another method of phase insensitive summation of acquired rf data would be to sum and average the magnitudes of the rf signals at each array point and then square the average, essentially moving the left parenthesis in Eq. (2) to include the summation symbol; however, only the phase insensitive technique described above was performed for this study.

Nominally, phase insensitive analysis is not subject to artifacts due to phase cancellation at the surface of the receiver. The PVDF membrane hydrophone used is a phase sensitive receiver; however, the active element is only 0.6 mm in diameter. The wavelengths (in water) for the 300–700 kHz data are approximately 5–2 mm, respectively, and 0.5–0.2 mm for the 3–7 MHz data. Phase cancellation should be minimal across the active element of this receiver over both bandwidths studied.

2. Transverse plane analysis

In order to illustrate the potential sources of apparent loss from both phase cancellation and interference for the stepped-Lexan phantom, for each position in the pseudoarray both the water-only reference path and the through sample path were processed to provide the magnitude, in-phase, and

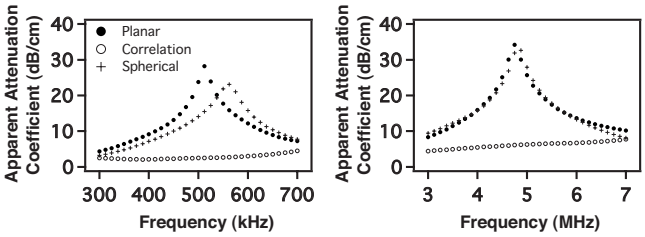


FIG. 1. Phase sensitively processed data collected over stepped plastic plates using three focusing algorithms: spherical, correlation, and planar. The pseudoarray scanned a diameter twice that of the transmitting transducer used for each bandwidth. The data centered at 500 kHz are from a Lexan sample, and the data centered at 5 MHz were collected from a Plexiglas sample.

in-quadrature components for the 500 kHz component of the received signal. Resulting plots of the magnitude, in-phase, and in-quadrature components over the transverse cross section were analyzed by examining line profiles of each component for that frequency. For purposes of illustration, only the 500 kHz component is displayed in Figs. 4 and 5 because the size of the step discontinuity in the Lexan sample was specifically chosen to nominally cancel this particular frequency. The source of this apparent loss should therefore be made most clear at this frequency.

III. RESULTS

Figure 1 displays the apparent attenuation coefficient obtained with phase sensitive planar, spherical, and correlation focusing as a function of frequency for the data collected over the stepped region of each plate. The anticipated large phase cancellation artifact near band center observed with planar focusing is somewhat reduced by spherical focusing, and is significantly reduced with correlation focusing.

The results obtained with correlation focusing and those obtained with phase insensitive processing are compared with the results of measurements of the flat and parallel plate obtained prior to machining the step in Fig. 2. Error bars are too small to be seen and are not shown. For the relatively large aperture employed (twice the transmitting transducer's aperture), both phase insensitive detection and correlation focusing with the maximum intensity through-sample signal yield values for the apparent attenuation coefficient that are quite comparable to the values obtained in a flat and parallel specimen free from the phase aberration-inducing artifact.

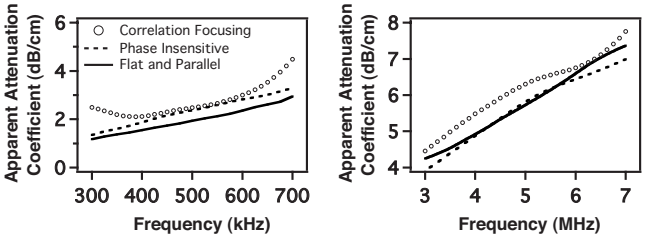


FIG. 2. Phase sensitively processed correlation data and phase insensitively processed planar data collected through the stepped region of each plate. The pseudo-array scanned a diameter twice that of the transmitting transducer used for each bandwidth. Data are plotted along with the results of insonifying a flat and parallel Lexan (500 kHz) and Plexiglas (5 MHz) sample.

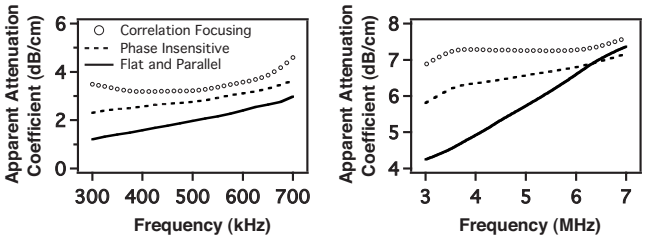


FIG. 3. Phase sensitively processed correlation data and phase insensitively processed planar data collected through the stepped region of each plate for a pseudo-array scanning a diameter equal to that used on transmit for each bandwidth. Data are plotted along with the results of insonifying a flat and parallel Lexan (500 kHz) and Plexiglas (5 MHz) sample.

Figure 3 displays data obtained by using a receiving aperture equal to that of the transmitting transducer (that is, one-half the diameter of that employed in Fig. 2). In contrast with the results shown in Fig. 2, neither phase insensitive processing nor correlation focusing is capable of producing results consistent with those obtained from the gold-standard flat and parallel sample. As discussed in Sec. IV, these overestimates of the attenuation coefficient presumably occur as a result of interference in the ultrasonic field that redistributes the energy in a fashion that cannot be overcome by phase insensitive or correlation processing.

Figures 4 and 5 show the magnitude, in-phase, and in-quadrature components of the transverse, receive-plane signals detected by the entire 57.6 mm by 57.6 mm (2.25 in. by 2.25 in.) area hydrophone scan for both the water path signal and the through-sample path signal for the stepped Lexan phantom at 500 kHz. The range of the color map for each figure is different and scaled relative to the maximum amplitude of the magnitude profile at 0 mm elevation. The maximum red and blue value for each figure is the negative maximum amplitude and positive maximum amplitude value respectively. Line profiles taken across the receiver aperture are denoted by the dashed line at 0 mm elevation for all plots and shown below each magnitude, in-phase, and in-quadrature component.

Losses due to phase cancellation arise from the summation over the aperture of positive and negative values in either the in-phase or the in-quadrature components at a given frequency. For phase sensitive detection, the in-phase and in-quadrature components of the incident pressure field are summed separately across the surface, yielding the overall detected field. This instrumental effect is the source of phase cancellation occurring at the surface of a phase sensitive receiver.

Losses due to interference occurring in the ultrasonic field are attributed to local minima or maxima in the magnitude of the detect field. Regions of positive and negative pressures occurring in the field at a specific spatial location simultaneously will partially cancel, leaving a local pressure field of lower magnitude, while regions of same-signed pressures will additively contribute to the local magnitude of the pressure field. This interference that occurs in the field results in an overall redistribution of the total energy and not a loss in energy; any local region possessing a smaller than average pressure field will always be compensated by another region that contains a higher than average pressure

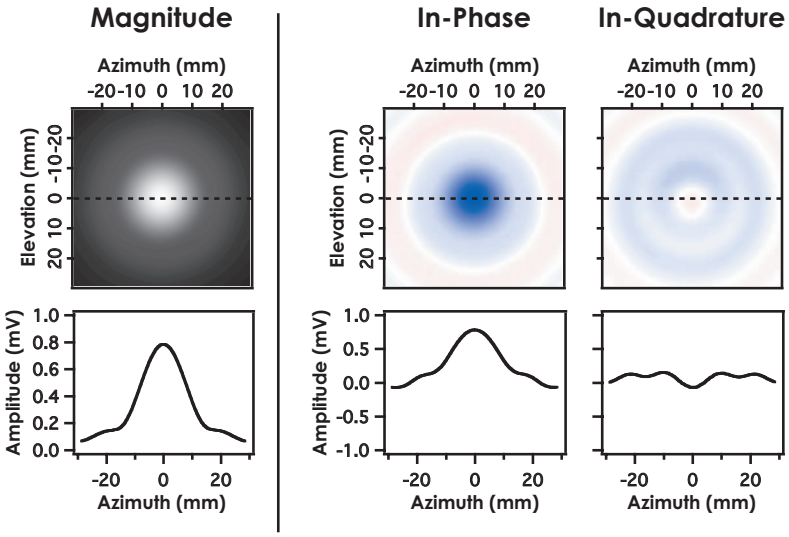


FIG. 4. Magnitude, in-phase, and in-quadrature components of the 500 kHz component of the water reference path. The three line profiles illustrate the symmetry in the pressure field as a function of aperture coordinate for an unaberated field.

field. This can be clearly seen in the magnitude panel of Fig. 5. Again, for purposes of illustration, only the 500 kHz signal is displayed to depict the effects of interference and phase cancellation occurring at the surface of the receiver.

Figure 6 shows the relative impact of aperture size on phase insensitive measurements of the apparent attenuation of the stepped-Lexan phantom compared with the measurement of the true attenuation coefficient of flat-and-parallel Lexan. The four apertures investigated for this figure were 0.6 mm (point-like), 14.28 mm (0.5625 in.), 28.58 mm (1.125 in.), and 57.15 mm (2.25 in.) to simulate four different diameter receiving apertures. Although Figs. 1 and 6 are qualitatively similar, the source of apparent loss is quite different. Figure 1 displays the results of phase sensitively summing rf data over a 2.25-in. aperture, while Fig. 6 shows the results of phase insensitively summing rf signals over different aperture sizes.

For the point-like aperture axially aligned with the transmitter in Fig. 6 apparent losses are largely due to interference effects because little phase cancellation can occur across the face of this 0.6 mm (0.023 in.) aperture. This can be seen by looking at the 0 mm elevational line profile of Fig. 5. The amplitude measured at this center location (0 mm azimuth) is very small. As the size of the aperture is increased, artifacts due to interference occurring in the field are reduced as the local minima in the pressure field are countered by local maxima in the pressure field. Therefore, artifacts due to phase cancellation become increasingly dominant for the larger aperture sizes, as more positive and negative (blue and red) pressures are incident over the aperture, as seen in Fig. 5. Measuring the true attenuation coefficient of the Lexan is only possible by scanning a sufficiently large cross-sectional area to recover all of the redistributed ultrasonic energy in conjunction with a phase insensitive detector to eliminate phase cancellation errors. This same behavior was seen in the 5 MHz data for Plexiglas and is not shown.

IV. DISCUSSION

Studies of trabecular bone are complicated by the inhomogeneous structure and irregular shape. Differences in the speed of sound between the solid structure and the marrow (or water) result in significant aberration of the incident ultrasonic beam. Previous investigators have identified phase cancellation at the receiving transducer as a source of artifact in attenuation measurements such as BUA and have illus-

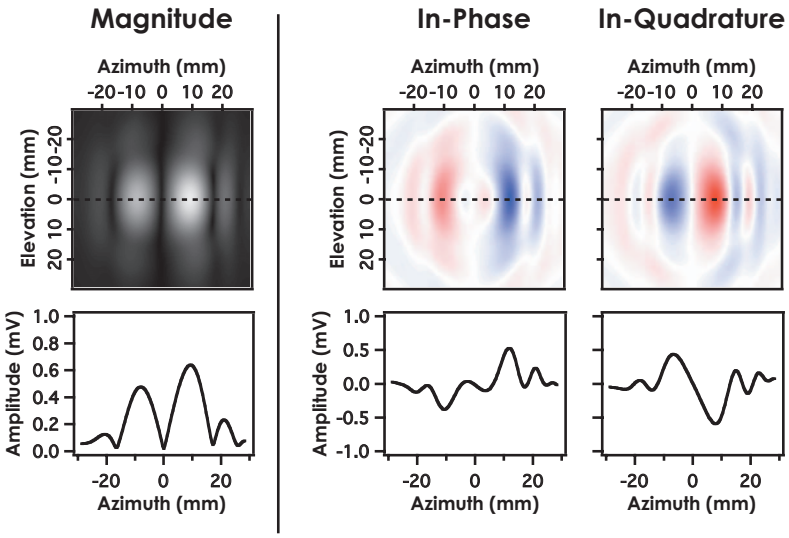


FIG. 5. Magnitude, in-phase, and in-quadrature components of the 500 kHz component in the transverse plane as a result of insonifying a stepped (Lexan) plate. Interference artifacts result from local minima and maxima in the pressure field, while phase cancellation effects result from local positive and negative pressures simultaneously incident on the receiver.

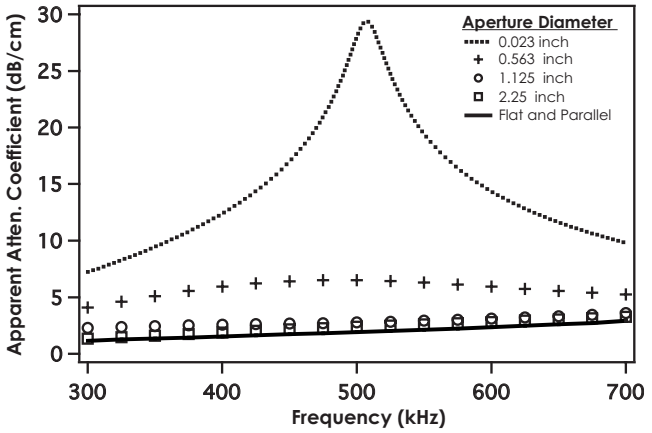


FIG. 6. Phase insensitive measurements of the apparent attenuation coefficient of the stepped Lexan phantom vs frequency for different diameter apertures.

trated the use of phase insensitive and alternative forms of processing to reduce phase aberration effects.[1–7,23–29] In studies of very complicated structures represented by bone, it would be difficult to segregate effects arising from the irreversible loss of energy associated with phase cancellation at the receiving transducer from those arising from lossless redistribution of energy in the ultrasonic field. The present study was designed to permit the segregation of those effects and to illustrate the influence of the size of the receiving aperture on the interplay between these related but distinct phenomena.

Results presented in Figs. 1 and 2 establish that for these specimens phase insensitive detection and phase sensitive correlation processing are capable of minimizing artifacts arising from phase cancellation at the receiver, consistent with results previously reported in studies of bone.[1–7] The data presented in Figs. 3 and 6, however, indicate that such processing is not sufficient to obtain the true attenuation coefficient if the receiver aperture is not of sufficient size. The presumption is that lossless redistribution arising from interference, and not phase cancellation at the receiver, is the source of the overestimation of the attenuation coefficient shown in Figs. 3 and 6. It is also interesting that for this smaller aperture size (Fig. 3), the slope of the measured apparent attenuation coefficient (nBUA) at 5 MHz appears to be underestimated, which counters the presumption that phase cancellation artifacts result in overestimations of nBUA, even though the attenuation coefficient at specific frequencies is still overestimated.

The phase insensitively processed data are in slightly better agreement with the flat and parallel data at 5 MHz than at 500 kHz. This might be due to the difference in the ratio of wavelength to aperture size at each center frequency. The wavelength (in water) to aperture ratio is larger at 5 MHz than at 500 kHz for the two largest apertures used in this experiment (12.7 and 57.15 mm). To compare properly the data at 5 MHz over the 12.7 mm aperture size, one would need a 127 mm (5 in.) diameter receiving aperture for the data centered at 500 kHz. It would be very difficult to employ this large aperture in a clinical setting. Therefore, for specific clinical or experimental investigations, a systematic study may be required to determine the appropriate dimensions for the receiving aperture.

V. SUMMARY

It has been shown that artifacts due to phase cancellation at the surface of a phase sensitive single element receiver can be minimized using pseudoarray data. The different focusing techniques used in this study reduce substantially the overestimation of the apparent attenuation of both Lexan and Plexiglas over two different bandwidths separated by a decade in frequency. A phase sensitive correlation technique yielded measurements comparable to those from a phase insensitive receiver thus avoiding artifacts arising from the processing of a multicomponent signal as if it were a single component signal. Results presented indicate that even phase insensitively processed data are subject to artifacts due to interference occurring in the field if an insufficiently large receiving aperture is chosen.

ACKNOWLEDGMENTS

This work was supported in part by NSF CBET 0717830 "Scholar in Residence at the FDA" and NIHHLBI HL 040302.

APPENDIX: FOCUSING ALGORITHMS

Stored ultrasonic waveforms acquired by the two-dimensional pseudoarray were processed using three focusing techniques: planar focusing, spherical focusing, and correlation focusing. Focusing of the array was achieved by applying the appropriate time shift to the individually acquired rf signals according to their position in the array and type of focusing employed. For the ijth position in the array, the acquired signal is represented by $U_{ij}(t_k)$, where t_k is some timing marker of the captured rf waveform. Then the focused waveform, $\bar{U}_{ij}(t_k)$, can be written $\bar{U}_{ij}(t_k)=U_{ij}(t_k+\Delta t_{ij})$, where Δt_{ij} is the focusing time shift applied to the array element.

1. Planar focusing. For planar focusing of the two-dimensional receiver array, a time shift of zero was applied to each signal in the two-dimensional array, i.e., $\Delta t_{ij}=0$. The data approximately represent what a single-element, planar transducer would measure in the field in place of the membrane hydrophone.

2. Spherical focusing. Spherical focusing was achieved by shifting the rf signals in the two-dimensional pseudoarray by the appropriate time delays according to their positions in the array and the desired focal distance. The time delays were determined by subtracting the time-of-flight of an ultrasonic wave in water from each element in the array to the focal point and the time-of-flight of an ultrasonic wave from the center of the array to the focal point. For the data centered about 500 kHz, the focal point was 63.4 mm (2.5 in.) from the center of the array, and for the data centered around 5 MHz, the focal distance was 38.1 mm (1.5 in.). These focal distances correspond to the distances from the sample to the center of the array in each experiment.

If the focal distance R represents the distance from the center of the array to a point along a line perpendicular to the plane of the array and r' is the distance between the

center of the array and another point on the array, then the distance between any position in the array and the focal point is $r=\sqrt{r'^2+R^2}$. The difference in distance between any array position and the focal point and the center array position and the focal point is then $\Delta r=r-R=\sqrt{r'^2+R^2}-R$. Thus, the time shift Δt_{ij} required to focus any ultrasonic wave on the array to the specified focal point is then $\Delta t_{ij}=\Delta r/c_w=(\sqrt{r'^2+R^2}-R)/c_w$, where c_w is the speed of sound in water.

3. Correlation focusing. The correlation focusing method employed to focus the two-dimensional pseudoarray is based in part on techniques introduced previously by others.[31–33] The correlation time shift applied to the rf waveform at each array position in the pseudoarray was chosen by maximizing the cross correlation between the waveform acquired at a given array position and the waveform at a reference position, that of the position of maximum intensity.

The cross-correlation function, $C_{ij}(s)$ can be written as

$$C_{ij}(s)=\sum_{t_k=0}^{N-I} U'(t_k)U_{ij}(t_k+s),$$

where $U'(t_k)$ is the time domain response at the reference position, $U_{ij}(t_k)$ is the response at the ijth position in the array, t_k is a time index corresponding to the sampled points in the rf waveform, and s is the time shift of an integer number of sample points applied to the waveform. The correlation focusing time shift applied to the waveform recorded in each array position, Δt_{ij}, is then found by maximizing the correlation function, $C_{ij}(s)$.

[1] C. M. Langton, S. B. Palmer, and R. W. Porter, "The measurement of broadband ultrasonic attenuation in cancellous bone," Eng. Med. **13**, 89–91 (1984).

[2] G. W. Petley, P. A. Robins, and J. D. Aindow, "Broadband ultrasonic attenuation: Are current measurement techniques inherently inaccurate?," Br. J. Radiol. **68**, 1212–1214 (1995).

[3] R. Strelitizki, S. C. Metcalfe, P. H. Nicholson, J. A. Evans, and V. Paech, "On the ultrasonic attenuation and its frequency dependence in the os calcis assessed with a multielement receiver," Ultrasound Med. Biol. **25**, 133–141 (1999).

[4] K. A. Wear, "The effect of phase cancellation on estimates of calcaneal broadband ultrasound attenuation *in vivo*," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 1352–1359 (2007).

[5] Y. Xia, W. Lin, and Y. X. Qin, "The influence of cortical end-plate on broadband ultrasound attenuation measurements at the human calcaneus using scanning confocal ultrasound," J. Acoust. Soc. Am. **118**, 1801–1807 (2005).

[6] K. A. Wear, "The effect of phase cancellation on estimates of broadband ultrasound attenuation and backscatter coefficient in human calcaneus *in vitro*," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 384–390 (2008).

[7] C. M. Langton and M. Subhan, "Computer and experimental simulation of a cortical end-plate phase cancellation artefact in the measurement of BUA at the calcaneus," Physiol. Meas **22**, 581–587 (2001).

[8] C. Chappard, P. Laugier, B. Fournier, C. Roux, and G. Berger, "Assessment of the relationship between broadband ultrasound attenuation and bone mineral density at the calcaneus using BUA imaging and DXA," Osteoporosis Int. **7**, 316–322 (1997).

[9] C. M. Langton and D. K. Langton, "Comparison of bone mineral density and quantitative ultrasound of the calcaneus: Site-matched correlation and discrimination of axial BMD status," Br. J. Radiol. **73**, 31–35 (2000).

[10] P. Laugier, P. Droin, A. M. Laval-Jeantet, and G. Berger, "*In vitro* assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and quantitative computed tomography," Bone (N.Y.) **20**, 157–165 (1997).

[11] P. H. Nicholson, R. Muller, G. Lowet, X. G. Cheng, T. Hildebrand, P. Ruegsegger, G. van der Perre, J. Dequeker, and S. Boonen, "Do quantitative ultrasound measurements reflect structure independently of density in human vertebral cancellous bone?," Bone (N.Y.) **23**, 425–431 (1998).

[12] P. Rossman, J. Zagzebski, C. Mesina, J. Sorenson, and R. Mazess, "Comparison of speed of sound and ultrasound attenuation in the os calcis to bone density of the radius, femur and lumbar spine," Clin. Phys. Physiol. Meas. **10**, 353–360 (1989).

[13] M. B. Tavakoli and J. A. Evans, "Dependence of the velocity and attenuation of ultrasound in bone on the mineral content," Phys. Med. Biol. **36**, 1529–1537 (1991).

[14] H. Trebacz and A. Natali, "Ultrasound velocity and attenuation in cancellous bone samples from lumbar vertebra and calcaneus," Osteoporosis Int. **9**, 99–105 (1999).

[15] D. C. Bauer, C. C. Gluer, J. A. Cauley, T. M. Vogt, K. E. Ensrud, H. K. Genant, and D. M. Black, "Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older women. A prospective study. Study of Osteoporotic Fractures Research Group," Arch. Intern Med. **157**, 629–634 (1997).

[16] D. Hans, P. Dargent-Molina, A. M. Schott, J. L. Sebert, C. Cormier, P. O. Kotzki, P. D. Delmas, J. M. Pouilles, G. Breart, and P. J. Meunier, "Ultrasonographic heel measurements to predict hip fracture in elderly women: the EPIDOS prospective study," Lancet **348**, 511–514 (1996).

[17] D. Hans, A. M. Schott, F. Duboeuf, C. Durosier, and P. J. Meunier, "Does follow-up duration influence the ultrasound and DXA prediction of hip fracture? The EPIDOS prospective study," Bone (N.Y.) **35**, 357–363 (2004).

[18] J. Huopio, H. Kroger, R. Honkanen, J. Jurvelin, S. Saarikoski, and E. Alhava, "Calcaneal ultrasound predicts early postmenopausal fractures as well as axial BMD. A prospective study of 422 women," Osteoporosis Int. **15**, 190–195 (2004).

[19] K. T. Khaw, J. Reeve, R. Luben, S. Bingham, A. Welch, N. Wareham, S. Oakes, and N. Day, "Prediction of total and hip fracture risk in men and women by quantitative ultrasound of the calcaneus: EPIC-Norfolk prospective population study," Lancet **363**, 197–202 (2004).

[20] M. A. Krieg, J. Cornuz, C. Ruffieux, G. Van Melle, D. Buche, M. A. Dambacher, D. Hans, F. Hartl, H. J. Hauselmann, M. Kraenzlin, K. Lippuner, M. Neff, P. Pancaldi, R. Rizzoli, F. Tanzi, R. Theiler, A. Tyndall, C. Wimpfheimer, and P. Burckhardt, "Prediction of hip fracture risk by quantitative ultrasound in more than 7000 Swiss women > or = 70 years of age: comparison of three technologically different bone ultrasound devices in the SEMOF study," J. Bone Miner. Res. **21**, 1457–1463 (2006).

[21] P. D. Miller, E. S. Siris, E. Barrett-Connor, K. G. Faulkner, L. E. Wehren, T. A. Abbott, Y. T. Chen, M. L. Berger, A. C. Santora, and L. M. Sherwood, "Prediction of fracture risk in postmenopausal white women with peripheral bone densitometry: Evidence from the National Osteoporosis Risk Assessment," J. Bone Miner. Res. **17**, 2222–2230 (2002).

[22] A. M. Schott, D. Hans, F. Duboeuf, P. Dargent-Molina, T. Hajri, G. Breart, and P. J. Meunier, "Quantitative ultrasound parameters as well as bone mineral density are better predictors of trochanteric than cervical hip fractures in elderly women. Results from the EPIDOS study," Bone (N.Y.) **37**, 858–863 (2005).

[23] C. M. Langton, C. F. Njeh, R. Hodgskinson, and J. D. Currey, "Prediction of mechanical properties of the human calcaneus by broadband ultrasonic attenuation," Bone (N.Y.) **18**, 495–503 (1996).

[24] H. Ermert, O. Keitmann, R. Oppelt, B. Granz, A. Pesavento, M. Vester, B. Tillig, and V. Sander, "A new concept for a real-time ultrasound transmission camera," IEEE Ultrasonics Symposium Proceedings, (2000)., pp. 1611–1614.

[25] M. Defontaine, S. Bonneau, F. Padilla, M. A. Gomez, M. Nasser Eddin, P. Laugier, and F. Patat, "2D arrays device for calcaneus bone transmission: An alternative technological solution using crossed beam forming," Ultrasonics **42**, 745–752 (2004).

[26] M. N. Eddin, M. Defontaine, M. A. Gomez, and F. Patat, "Phase aberration correction: Application to *in vivo* bone images," Proc.-IEEE Ultrason. Symp. **2**, 1357–1360 (2002).

[27] L. J. Busse and J. G. Miller, "Response characteristics of a finite aperture, phase insensitive ultrasonic receiver based upon the acoustoelectric effect," J. Acoust. Soc. Am. **70**, 1370–1376 (1981).

[28] A. Q. Bauer, K. R. Marutyan, M. R. Holland, and J. G. Miller, "Is the Kramers–Kronig relationship between ultrasonic attenuation and dispersion maintained in the presence of apparent losses due to phase cancellation?," J. Acoust. Soc. Am. **122**, 222–228 (2007).

[29] A. Q. Bauer, K. R. Marutyan, M. R. Holland, and J. G. Miller, "Negative dispersion in bone: The role of interference in measurements of the apparent phase velocity of two temporally overlapping signals," J. Acoust. Soc.

Am. **123**, 2407–2414 (2008).
[30]B. Kusse and E. Westwig, *Mathematical Physics* (Wiley, New York, 1998).
[31]S. W. Flax and M. O'Donnell, "Phase-aberration correction using signals from point reflectors and diffuse scatterers: basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 758–767 (1988).
[32]M. O'Donnell and W. E. Engeler, "Correlation-based aberration correction in the presence of inoperable elements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 700–707 (1992).
[33]M. O'Donnell and S. W. Flax, "Phase-aberration correction using signals from point reflectors and diffuse scatterers: Measurements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 768–774 (1988).

Interest of the attenuation coefficient in multiparametric high frequency ultrasound investigation of whole blood coagulation process

Rachel Callé[a] and Camille Plag
Université François Rabelais, INSERM U930, CNRS 2448 FRE, Tours, France

Frédéric Patat and Frédéric Ossant
CHRU, Tours, France and Université François Rabelais, INSERM U930, CNRS 2448 FRE, Tours, France

(Received 31 March 2008; revised 9 October 2008; accepted 20 October 2008)

Previous studies [R. Libgot, F. Ossant, Y. Gruel, P. Lermusiaux, and F. Patat, Proc.-IEEE Utrason. Symp. **4**, 2259–2262 (2005); R. Libgot-Callé, F. Ossant, Y. Gruel, P. Lermusiaux, and F. Patat, Ultrasound Med. Biol. **34**, 252–264 (2008); F. Ossant, R. Libgot, P. Coupé, P. Lermusiaux, and F. Patat, Proc.-IEEE Ultrason. Symp. **2**, 846–849 (2004)] showed the potential of an *in vitro* high frequency ultrasound (beyond 20 MHz) device to describe the blood clotting process. The parameters were simultaneously estimated in double transmission (DT) with the calculation of the velocity of longitudinal waves and in backscattering (BS) modes with the estimation of the integrated BS coefficient and the effective scatterer size. The aim of the present study was to show how the integrated attenuation coefficient (IAC) assessed in DT mode could provide additional information on this process, especially regarding the fibrin polymerization which is an important part of the coagulation process. A characteristic time t_a of the variations in IAC that could be linked to fibrin formation was identified. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3021439]

PACS number(s): 43.80.Ev, 43.80.Qf, 43.80.Vj [FD] Pages: 530–538

I. INTRODUCTION

Due to its clinical interest, several applied research disciplines have focused on the blood clotting process, a very complex phenomenon. Recent studies[1,2] have revealed an important need to develop accurate and global standard *in vitro* coagulation tests using whole blood samples. These new tests should complement existing standard tests performed with plasma samples. Due to the nature of acoustic waves, ultrasound measurement techniques have the major advantages of allowing the use of whole blood samples and providing a description of the clotting process in terms of mechanical properties.

In previous studies,[3–5] our contribution was to propose an *in vitro* high frequency ultrasound (beyond 20 MHz) analysis of this complex process by the simultaneous measurement of the longitudinal wave velocity in double-transmission (DT) mode, and integrated backscatter coefficient (IBC) and effective scatterer size (ESS) in backscatter (BS) mode. We demonstrated that simultaneous measurement of such BS and DT parameters permitted the monitoring of different stages of the blood clotting process, both in terms of mechanical properties (blood gelation) and spatial distribution of scattering structures red blood cells (RBC aggregates).

In this study, we focused on studying the attenuation coefficient in order to provide additional information regarding coagulation stages. Most measurements of the attenuation coefficient have previously been performed using transmission measurement methods. Studies of the evolution of the attenuation coefficient in whole unclotting blood as a function of the frequency, the hematocrit, and the amount of plasma proteins have been reported in the literature.[6–10] Shung *et al.*[11] proposed an ultrasound technique capable of measuring a coagulation time (thrombin time) for the human blood coagulation process, based on the study of the scattered field from coagulating blood. Other studies[12,13] by the same team were performed with 7.5 MHz ultrasound. They simultaneously estimated the BS coefficient, acoustic velocity, and attenuation coefficient over 24 h after initiating coagulation, with 1 h time resolution. They found monotonic variations in attenuation coefficient correlated with those in other acoustic parameters. Other studies of the coagulation process were performed with porcine whole blood over 50 min, including respectively those of Huang *et al.* and Wang *et al.*,[14,15] at 5 and 10 MHz. As a result, very few studies have studied the evolution of the ultrasound attenuation coefficient for frequencies beyond 20 MHz during whole blood coagulation. Wavelengths were much greater than structure size in blood at low frequencies, i.e., RBC aggregates and consequently RBC, and this may have induced differences in the kinetics of acoustic parameters.

However, Calor-Filho and Machado[16,17] studied the plasma coagulation process by measurement of the attenuation coefficient at frequencies between 8 and 22 MHz. They reported a slow decrease in attenuation of less than 0.2 dB/cm during the process. It appears that measurement of the attenuation coefficient in clotting plasma provides less information than in the case of clotting whole blood. More-

[a]Electronic mail: rachel.libgot@univ-tours.fr

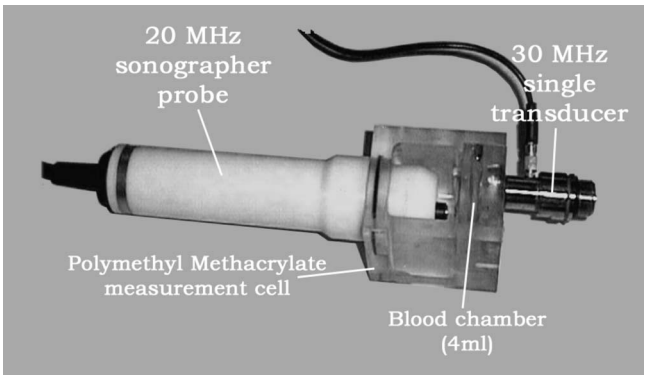


FIG. 1. Photograph of the measurement cell.

over, the isolated study of BS or DT parameters provided only incomplete information and interpretation of variations in these parameters was more difficult.

In another approach, Gennisson *et al.*[18] studied the evolution of velocity and attenuation of shear waves in porcine clotting blood. They also measured kinetics of the elasticity μ and viscosity η of blood clots. They defined t_B as the time from which the state of coagulating blood changed from liquid to solid.

In this study, we measured and analyzed the kinetics of the attenuation coefficient, assessed by DT measurement, in a frequency bandwidth from 10 to 30 MHz. This parameter was compared to the acoustic velocity and to BS parameters for clotting and nonclotting blood in order to identify specific stages of the coagulation process. Assessment of the attenuation coefficient allowed us to explore an intermediate stage between the phase of RBC aggregation (during the first minutes of the process) and the phase of liquid/gel state transition that occurred up to 10 min after the start of the process. Measurement of this parameter could provide valuable information about fibrin polymerization and formation which are crucial phases of the process.

II. MATERIALS AND METHODS

A. Transducer characteristics and measurement cell

The characteristics of transducers and the measurement cell have already been reported in detail.[4] We summarize the main data below.

For DT measurements, we used a 30 MHz focused transducer (PI 50-2 200171, Panametrics) with a broadband (200 MHz) pulser/receiver (model 5900 PR, Panametrics). Each signal was digitized at 2 GHz with 8-bit resolution.

For BS measurements, we used a high frequency echographic device[19] (Dermcup2020™) with a 20 MHz central frequency oscillating single element transducer. This device provided real time images of 5 mm × 6 mm composed of 256 radio frequency (rf) lines sampled at 500 MHz.

The measurement cell was made from polymethylmethacrylate (PMMA extrudé, Weber Métaux et Plastiques, France) material with a 4 ml blood chamber and placed in an incubator regulated at 37 °C ± 0.1 (Memmert® model 200, Schwabach, Germany) (Fig. 1). Blood and water chambers were separated by a thin layer of plastic (15 μm in the case of BS measurement and a 1 mm layer of PMMA material in the case of DT measurement). To avoid liquid evaporation and to maintain a constant temperature in the measurement cell, a lid of molded silicon was placed over the cell during measurements. Measurements were performed at midheight of the blood chamber.

B. Ultrasound DT-mode parameters

Definition of acoustic parameters and their measurement have already been described.[3–5] We summarize the main guidelines below.

1. Acoustic velocity

In DT mode, acoustic velocity V_b was estimated using a classical pulse echo method (Eq. (1) and Fig. 2),

$$V_b = \frac{2L_b}{t_2 - t_1}. \quad (1)$$

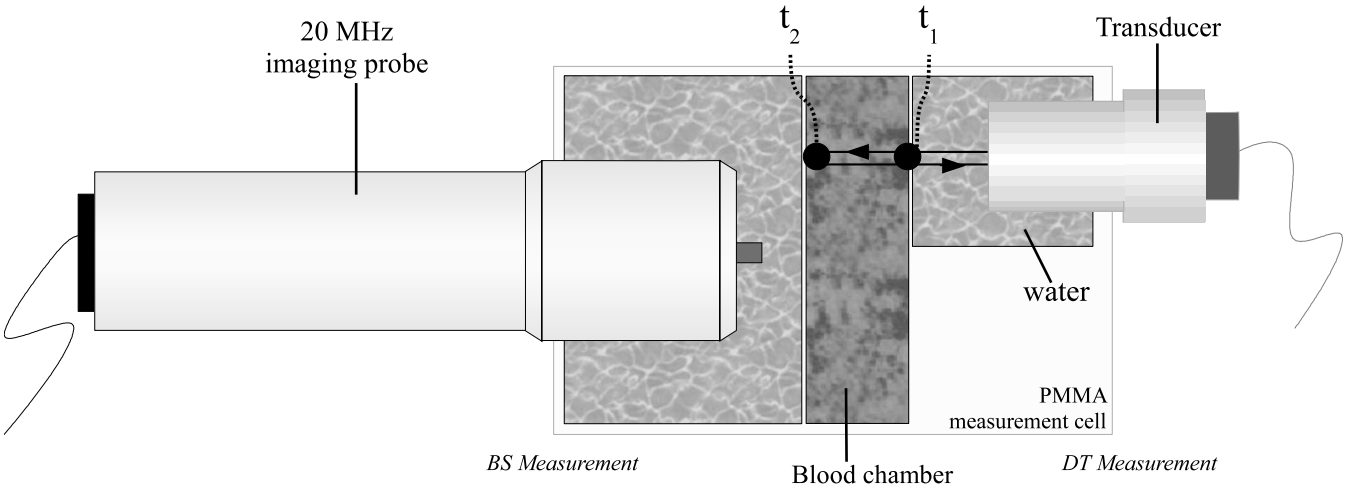


FIG. 2. Experimental device (top view).

The ultrasound pulse was transmitted into the blood sample through the water/blood division and reflected by the side of the blood chamber. We then measured the time interval between ultrasound pulses reflected from the water/blood division (t_1) and the side of the blood chamber (t_2) (Fig. 2). As the time t_1 was constant (for a constant temperature), variations in the difference between the two time intervals were due to changes in velocity in the blood sample along the propagation path (L_b). Time delays were estimated using intercorrelation and interpolation methods.[20] Precision of velocity measurement was estimated to 1 cm/s. L_b, the separating distance between both sides of the blood chamber, was initially determined by measuring the two arrival times (t_1 and t_2) in pure water at 37 °C. The ultrasound velocity in pure water at 37 °C was fixed at 1524.29 m/s.[21]

2. Integrated attenuation coefficient

The attenuation coefficient $\alpha(f)$ was assessed using a substitution method by comparing the power spectrum obtained in blood with that obtained in a reference medium placed in the measurement path (pure water at 37 °C in our case). Power spectra in blood and in the reference medium were deduced from the echo signal of the left side of the blood chamber (at t_2). The attenuation coefficient was deduced according to Eq. (2),

$$\alpha(f) = -\frac{1}{2L_b} 10 \log\left(\frac{S(f)}{S_{\text{ref}}(f)}\right), \tag{2}$$

where $S(f)$ was the power spectrum measured with the blood sample, S_{ref} was the power spectrum of the same reflector obtained with pure water at 37 °C, and $2L_b$ was the length of the acoustic path.

The integrated attenuation coefficient (IAC (dB/cm)) was calculated according to Eq. (3),

$$\text{IAC} = \int_{f_{\min}}^{f_{\max}} \frac{\alpha(f)}{f_{\min} - f_{\max}} df, \tag{3}$$

where IAC was estimated in the 10–30 MHz frequency range.

C. Ultrasound BS-mode parameters

For BS measurement, parameters were estimated in a region of interest (ROI) with 800×296 pixels (i.e., 8 mm^2) positioned 100 μm after the 15 μm layers of plastic separating blood and water chambers. The focal depth of the probe was positioned in the middle of the ROI. In order to reduce the amount of data, BS parameters acquisition was decomposed in two successive times: the first time lasted 4 min meanwhile each rf image was recorded every 5 s, and the second time lasted from 4 to 120 min meanwhile each rf image was recorded every 30 s. They were deduced from the calculation of BS coefficient σ_b, as described by Insana *et al.*,[22] and attenuation and diffraction corrected.

The ROI was subdivided into ten windows, each window representing 160×296 pixels in which the signal was assumed to be stationary, the windows overlapping by 50%. Each parameter was averaged on the ten values estimated on the ten windows. IBC and ESS were estimated in the 10–30 MHz frequency range.

1. IBC

The IBC (/sr/cm) was calculated according to Eq. (4),

$$\text{IBC} = \int_{f_{\min}}^{f_{\max}} \frac{\sigma_b(f,z)}{f_{\max} - f_{\min}} df. \tag{4}$$

2. ESS

Effective scatterer size (μm) was determined from the frequency dependence of the BS coefficient following the method described by Insana,[23] by using a Gaussian correlation model.

D. Blood protocols

Blood samples were obtained from 14 healthy volunteers in compliance with the Ethics Committee of the University Hospital of Tours.

1. Clotting blood experiment

Sixty-four blood samples were collected in 3.5 ml vacuum sodium citrate tubes (3.2%) previously heated to 37 ± 0.1 °C in an incubator. Tubes with whole blood were then kept at 37 °C for 30 min, and coagulation was initiated by adding calcium chloride (100 μl at 0.5 M concentration). Blood was immediately transferred to the blood chamber of the cell, just after stirring (Labquake® model 1419, Barnstead International, Dubuque, IA), and acoustic measurement began. All measurements were performed in static conditions.

In vitro blood coagulation is a rapid phenomenon[24] that lasts between a few seconds and a few minutes, depending on the blood protocol applied. It is well known that most of the useful information about this phenomenon occurs before clot retraction due to fibrin retraction, considered as the end of the coagulation process. It was therefore necessary to adapt the blood protocol to the observation duration of the process. The advantage of choosing sodium citrate as anticoagulant and calcium chloride to recalcify blood samples instead of heparin and thrombin was to increase the duration of coagulation and to be able to monitor BS parameters during the first few minutes with sufficient time resolution.

2. Nonclotting blood experiment

Whole blood samples from healthy volunteers were also collected in 3.5 ml citrated tubes under the same temperature conditions as above. Blood samples were then transferred from the citrated tubes to the measurement chamber and mechanically stirred (Labquake® model 1419, Barnstead International, Dubuque, IA) just before acoustic measurements, also performed at 37 ± 0.1 °C.

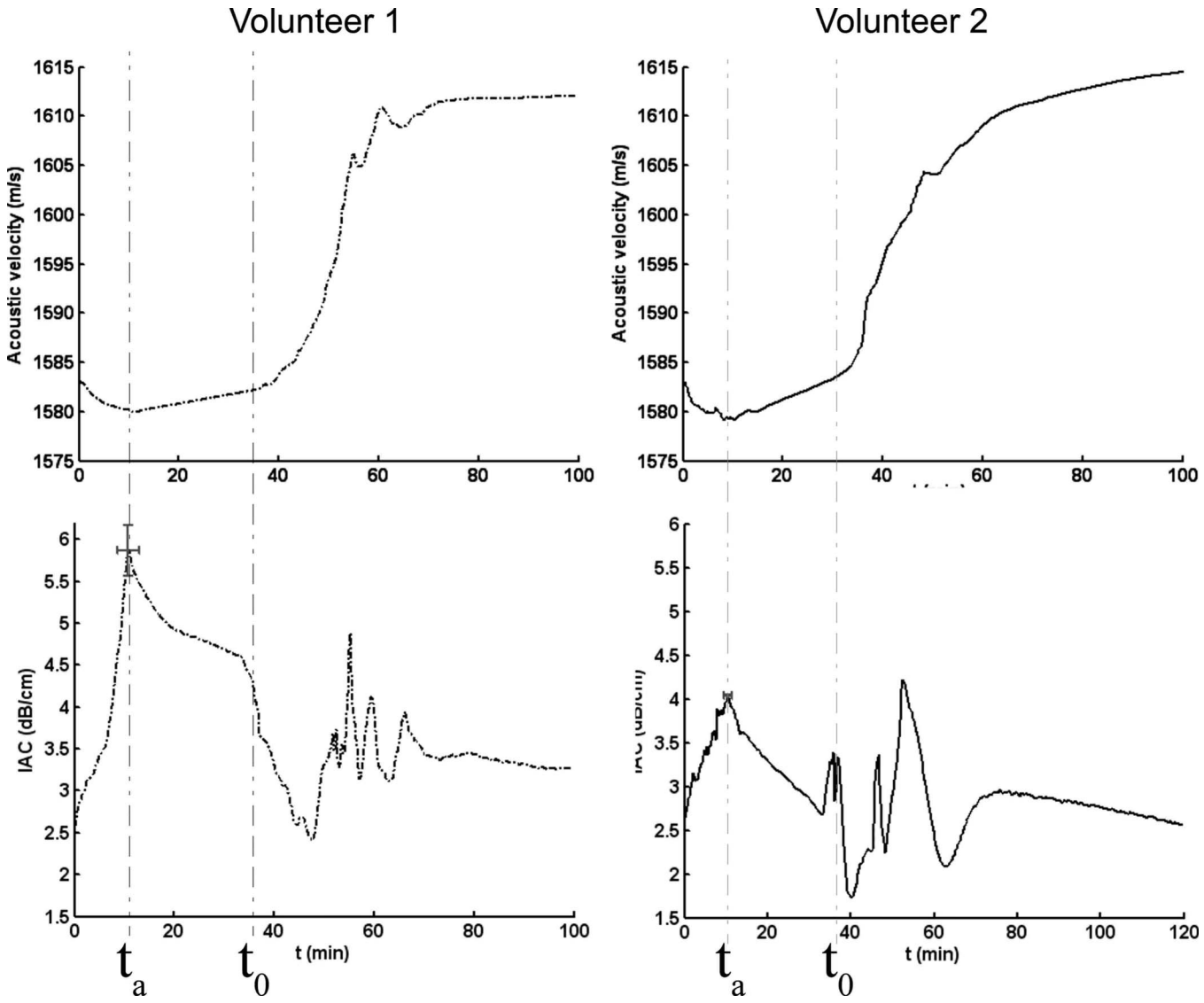


FIG. 3. Velocity and IAC parameters vs time (min) for volunteer 1 (dash-dotted line) and volunteer 2 (solid line) during the coagulation process with standard deviations at t_a.

III. RESULTS

A. Comparison between variations in velocity and IAC

In order to better analyze variations in the IAC parameter, we compared them with the variations in velocity.

1. Coagulation process

Figure 3 presents typical evolutions of the acoustic velocity and IAC parameter during the blood clotting process for two different volunteers.

We showed in previous studies[3–5] that the variations in acoustic velocity were linked to changes in state of the medium and were represented by sigmoid shape curve characterized by three stages. The first stage corresponded to a plateau, the second stage corresponded to several 10 m/s increase in velocity, and the third stage corresponded to another plateau. The beginning of the second stage was noted t_0 (Fig. 3).

The main variations in IAC (Fig. 3) with good reproducibility occurred during the first minutes before t_0 and took place before the main variations in velocity. After t_0, there were noisy variations in IAC. These variations were not reproducible and were presumably associated with clot movements due to retraction. We therefore only explored the kinetics of IAC during the first minutes of the process until $t = t_0$.

Before t_0, variations in IAC were typical: one increase until a maximum value at time t_a, and then one decrease until t_0. The amplitude of these variations was about a few dB/cm.

These variations suggest that IAC values could provide complementary information regarding velocity, particularly to describe the first minutes of the blood clotting process.

2. Comparison between clotting and nonclotting blood

Figure 4 presents the typical evolution of the acoustic velocity and IAC parameter for clotting and nonclotting blood.

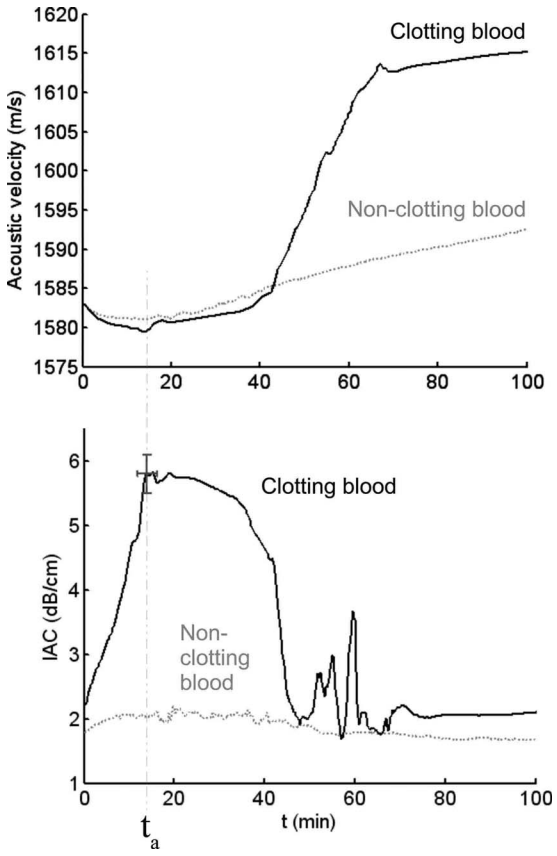


FIG. 4. Velocity and IAC parameters vs time (min) for clotting and nonclotting blood with standard deviations at t_a.

Variations in velocity were different between clotting and nonclotting blood. For clotting blood, as seen above, the mean amplitude of the variations in velocity was about several 10 m/s, including one stage with strong variations and two stages with no or very small variations of around a few m/s. For nonclotting blood, the mean amplitude of the variations in velocity was less than 10 m/s, characterized by regular increases.

Different tendencies of the variations in IAC were also noted for clotting and nonclotting blood. Variations in IAC were about 3.5 dB/cm for clotting blood, a maximum value being reached before 20 min, and less than 0.5 dB/cm for nonclotting blood, during 100 min of observation. A maximum value for IAC reached at time t_a (Fig. 3) did not occur in the case of nonclotting blood.

B. Comparison between BS and DT modes

In order to show the advantage of simultaneous measurement, we compared the variations in every acoustic parameter obtained in DT and BS modes. Figure 5 shows two typical examples of the variations in velocity, IAC, ESS, and IBC, during blood coagulation for two different volunteers.

Four phases were identified, the first phase corresponding to the increase in IBC and ESS up to a maximum value of IBC, the second phase corresponding to where ESS and IAC values increased to reach a maximum at the same instant ($t=t_a$) while IBC values decreased, the third phase corresponding to which ESS values were stable while IAC values decreased, and the fourth phase associated with a strong variation in velocity. Beyond phase 4, all parameters were almost stable.

C. Statistical study of acoustic parameters

1. Intra- and interindividual dispersions

In a previous study,[4] we estimated dispersion of the velocity parameter. We defined three criteria in order to characterize this dispersion: I, the final increase in velocity after 120 min of observation, t_0, corresponding to the time when the velocity reached 5% of I, and the slope S estimated between the times when the velocity reached 10% and 80% of I (Fig. 6). We found that interindividual variations were two times greater than intraindividual variations.

In this study, we estimated dispersions of the IAC parameter. The criterion t_a was chosen as corresponding to the time to reach a maximum IAC value (Fig. 7).

All criterion values were obtained from fourteen volunteers (64 samples). Table I presents criteria obtained from IAC curves in terms of mean and standard deviation values. These values showed that the relative standard deviation of t_a was 1.8 times greater for interindividual than for intraindividual variations.

In practical terms, standard deviations were computed for intra- and interindividual variations respectively, as described in Eq. (5),

$$\mathrm{SD}_{\mathrm{intra}}(X) = \overline{\mathrm{SD}_i},$$

$$\mathrm{SD}_{\mathrm{inter}}(X) = \mathrm{SD}(\overline{X_i}), \tag{5}$$

where X_i and SD_i are the mean and standard deviation of criterion X (t_a, for example) respectively, obtained for volunteer i (64 samples).

2. Correlation between statistical criteria

We measured the correlation coefficient between the two specific times t_a and t_0 (Fig. 8) and found no correlation [$r^2=0.2$ ($p<0.001$)] for 14 volunteers, corresponding to 64 blood samples.

IV. DISCUSSION

Hemostasis is a well known process composed of three stages: primary hemostasis, coagulation, and fibrinolysis. This is a complex chain reaction process involving proteins, called coagulation factors, designed to form a plug, to clot, and to stop bleeding from vascular injury[24–26] when *in vivo*. The biology of these factors has been studied for many years by performing *in vitro* coagulation tests using plasma samples. The use of high frequency ultrasound to characterize *in vitro* blood coagulation could provide information on the mechanics of the process.

The following points emphasize a few comments and some interpretations about the variations in every high frequency acoustic parameter assessed simultaneously. First of all, we compared parameters estimated from DT measurement. Second, we confronted every acoustic parameter synchronously measured. And then, we discussed the statistics of these parameters.

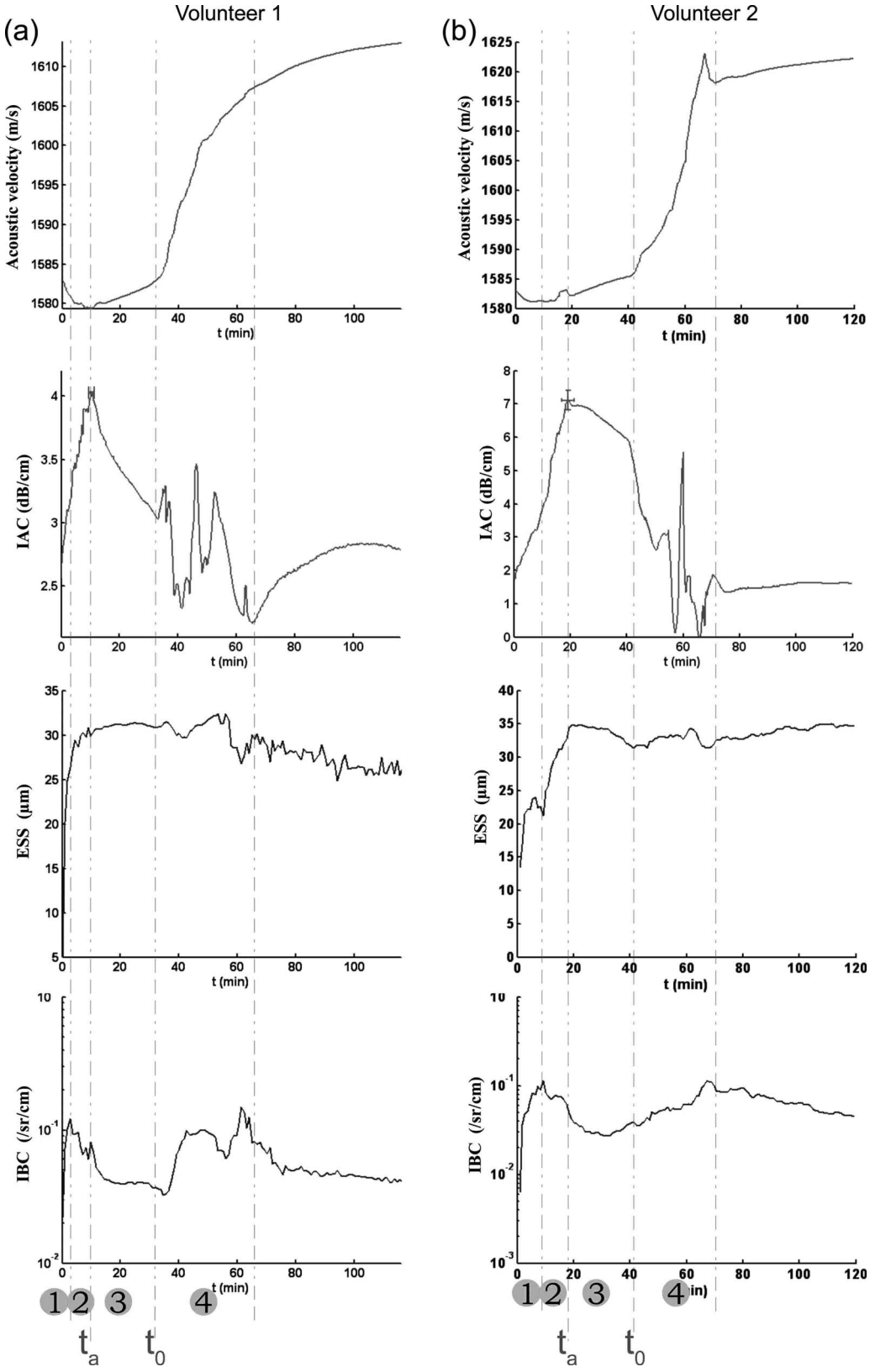


FIG. 5. Comparative study of variations in acoustic parameters estimated in DT (two curves at the top) with standard deviations at t_a and BS (two curves at the bottom) modes during the blood clotting process for two volunteers ((a) and b)).

A. Comparison between variations in velocity and IAC for clotting and nonclotting blood

In a previous study,[4] we focused on acoustic velocity for DT measurement. We concluded that this parameter provided an indirect measurement of the liquid-gel transition phase, as suggested by other authors,[27] these variations being characteristic of the transition phase in different media.[28,29] By comparison, it can be seen in Fig. 4 that the amplitudes of

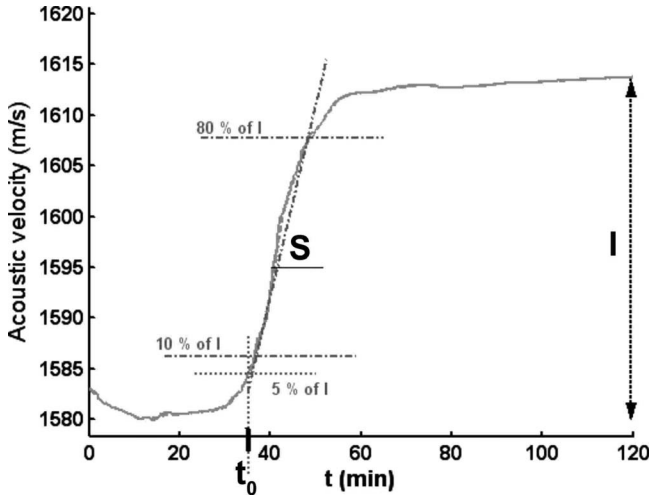


FIG. 6. Criteria of variations in velocity.

the variations in velocity for nonclotting blood were small and showed regular increase from 1580 to 1590 m/s, during the 100 min of the experiment. No irreversible changes in physical structure occurred in this case, and the only mechanical changes that likely influenced velocity arose from the sedimentation of the RBC at the bottom of the measurement cell.[30]

In the case of the IAC parameter (Fig. 4), the kinetics were also different between clotting and nonclotting blood. The IAC parameter showed a maximum value at time t_a, and this characteristic time was measurable for clotting but not for nonclotting blood. The amplitudes of the variations in IAC for both media were also not comparable (3.5 dB/cm for clotting blood versus 0.5 dB for nonclotting blood). Because of its reproducible character, these results suggested that these variations in IAC could be linked to specific transformations of the medium during the coagulation process.

Moreover, variations in IAC obtained on plasma samples by other researchers[16,17] emphasized small and noncomparable variations in IAC during the coagulation process, for frequencies between 8 and 22 MHz. They reported

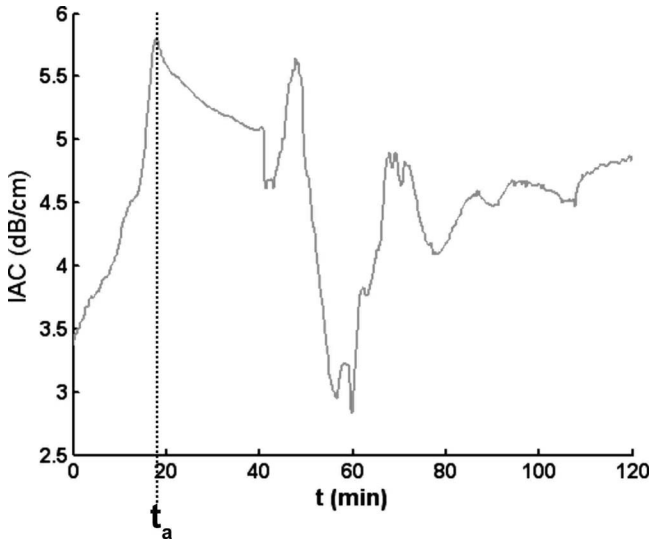


FIG. 7. Criterion of the variations in IAC.

TABLE I. Statistics of criterion for the intra- and interindividual dispersions in IAC parameter.

t_a		
Mean (min)	SD/mean (%)	
	Intra	Inter
17	±18	±32

a slow monotonic decrease in attenuation during the coagulation process which did not allow detection of a specific moment of coagulation. These results showed that our present results were specific to the whole blood clotting process, in particular for the high frequency ultrasound attenuation coefficient. These results emphasized the interest of using ultrasound at frequencies around 20 MHz (size structure closer to the wavelength than in the case of plasma samples) and could be an interesting way to emphasize the significant role of RBC in the coagulation process,[1,2] with whole blood samples.

B. Comparison of all acoustic parameters during the four phases of the coagulation process

During phase 1 (Fig. 5), IBC and ESS increased simultaneously. Several investigations[9,31–36] have already shown that echogenicity increases whereas the shear rate decreases during the very first minutes, and most reports associated this buildup mainly with RBC aggregation.[4]

When we looked at these two parameters during phase 2, the variations were reversed: the simultaneous increase in ESS values and decrease in IBC values could not have been due only to the spatial arrangement of the RBC, but to biochemical transformations of the plasma which changed the acoustic contrast between the RBC and plasma. We therefore associated these specific modifications with the beginning of fibrin polymerization, i.e., the first steps of the transformation of soluble fibrin monomers into insoluble fibrin.[37,38]

Phase 3 was characterized by a decrease in IAC and IBC parameters and stabilization of the ESS parameter. This led to a few hypotheses. As RBC were no longer aggregating and there was stabilization of the speckle of the rf image, we suggest that the structural arrangement was then immobilized

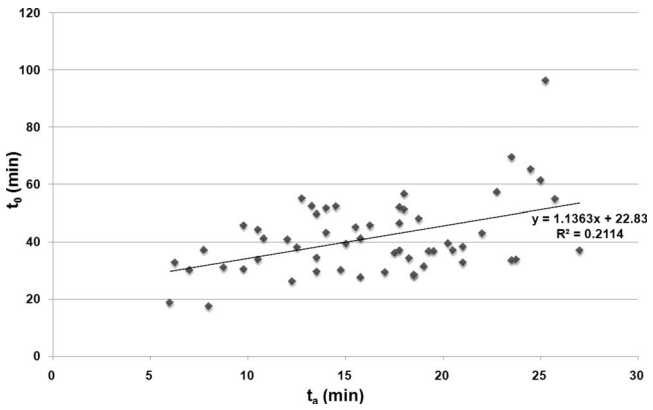


FIG. 8. Correlation between t_0 and t_a criteria.

in the blood, both in terms of size of RBC aggregates and in terms of the relative motion of these aggregates. This could be related to results previously reported by Shung *et al.*[11] where the time interval until the cessation of the fluctuations in the received scattered signal (mode A) was associated with the coagulation time (thrombin time). Phase 3 could therefore be identified as another important stage in fibrin polymerization in view of the decrease in IAC and the stabilization in size of the RBC aggregates and the speckle of the rf image. Moreover, the decrease in IBC confirmed that this polymerization could modify the acoustic contrast between RBC and the surrounding medium. The hypotheses suggested for phases 2 and 3 must be confirmed with further experiments. In this context, the IAC parameter seemed to be a key to describing some important stages in the coagulation process.

Finally, variations in IAC during phase 4 were not relevant and not reproducible, due to strong clot movements during and after serum expulsion. Major variations in velocity showed typical transformations during a liquid/gel transition. The results described above seemed to show that this transition could be initiated earlier (during phases 2 and 3) by biochemical transformations of the plasma. The variations in BS parameters during this phase have already been discussed in a previous study[4] and were not reproducible due to clot motion and inhomogeneity.

C. Statistics of DT parameters

We have previously[4] reported that the relative standard deviation of t_0 was 1.5 times greater for interindividual than for intraindividual variations. We presented here a new criterion linked to the IAC parameter, i.e., time t_a, which was 1.8 times greater for interindividual than for intraindividual variations. It thus appeared that this criterion is potentially able to discriminate between intra- and interindividual variations. These results suggest that they might be at least partly linked to the intrinsic properties of blood coagulation. Moreover, the lack of correlation between criteria t_0 and t_a (Fig. 8) suggested that they were not governed by the same phenomenon and thus would provide complementary information.

V. CONCLUSION

This study confirmed that high frequency acoustic analysis simultaneously using BS and DT modes was a promising approach to studying the whole blood clotting process. In particular, the variations in IAC were associated with fibrin polymerization and complemented the variations in velocity, IBC, and ESS. The important role played by RBC in the whole blood coagulation process was also emphasized by the evolution of the IAC parameter. Moreover, the discriminating potential of the acoustic parameters was evaluated by comparing intra- and interindividual variations in these parameters. Criteria linked to the intrinsic properties of the coagulability of blood were proposed to discriminate between these variations. In conclusion, it appeared that this system provided different specific times of the transformations in blood during coagulation.

ACKNOWLEDGMENTS

The authors thank J.-M. Grégoire and D. Carré (University of Tours, LUSSI, Tours, France) for their contribution to the electronic and mechanical aspects of the experimental setup. Our sincere thanks go to the nurses of the Ultrasound and Nuclear Medicine Departments of the University Hospital of Tours.

[1] http://www.fda.gov/cdrh/meetings/coag.html/ (Last viewed 3/28/2008).
[2] B. Sorensen, P. Johansen, K. Christiansen, M. Woelke, and J. Ingerslev, "Whole blood coagulation thrombelastographic profiles employing minimal tissue factor activation," J. Thromb. Haemostasis **1**, 551–558 (2003).
[3] R. Libgot, F. Ossant, Y. Gruel, P. Lermusiaux, and F. Patat, "High frequency ultrasound characterization of the blood clotting process: Intra- and inter-individual variations," Proc.-IEEE Ultrason. Symp. **4**, 2259–2262 (2005).
[4] R. Libgot-Callé, F. Ossant, Y. Gruel, P. Lermusiaux, and F. Patat, "High frequency ultrasound device to investigate the acoustic properties of whole blood during coagulation," Ultrasound Med. Biol. **34**, 252–264 (2008).
[5] F. Ossant, R. Libgot, P. Coupé, P. Lermusiaux, and F. Patat, "High frequency ultrasound characterization of the coagulation process of whole blood," Proc.-IEEE Ultrason. Symp. **2**, 846–849 (2004).
[6] E. L. Carstensen, K. Li, and H. P. Schwan, "Determination of acoustic properties of blood and its components," J. Acoust. Soc. Am. **25**, 286–289 (1953).
[7] E. L. Carstensen and H. P. Schwan, "Acoustic properties of hemoglobin solutions," J. Acoust. Soc. Am. **31**, 305–311 (1959).
[8] S. Maruvada, K. K. Shung, and S.-H. Whang, "High frequency backscatter and attenuation measurements of porcine erythrocyte suspensions between 30–90 MHz," Ultrasound Med. Biol. **28**, 1081–1088 (2002).
[9] S. H. Wang and K. K. Shung, "*In vivo* measurements of ultrasonic backscatter from blood," Proc.-IEEE Ultrason. Symp. **2**, 1161–1164 (1997).
[10] Y. M. Yuan and K. K. Shung, "Ultrasonic backscatter from flowing whole blood. II: Dependence on frequency and fibrinogen concentration," J. Acoust. Soc. Am. **84**, 1195–1200 (1988).
[11] K. K. Shung, R. A. Sigelmann, and G. Schmer, "Ultrasonic measurement of blood coagulation time," IEEE Trans. Biomed. Eng. **22**, 334–337 (1975).
[12] K. K. Shung, D. Fei, and J. O. Ballard, "Further studies on ultrasonic properties of blood clots," J. Clin. Ultrasound **14**, 269–275 (1986).
[13] K. K. Shung, D. Fei, Y. M. Yuan, and W. C. Reeves, "Ultrasonic characterization of blood during coagulation," J. Clin. Ultrasound **12**, 147–153 (1984).
[14] C. C. Huang, S.-H. Wang, and P.-H. Tsui, "Detection of blood coagulation and clot formation using quantitative ultrasonic parameters," Ultrasound Med. Biol. **31**, 1567–1573 (2005).
[15] S. H. Wang, T.-W. Chung, C. S. Huang, C.-T. Chuang, and P.-S. Tsui, "Detection of the process of blood coagulation and clot formation using quantitative ultrasonic parameters," Proc.-IEEE Ultrason. Symp. **2**, 1653–1656 (2002).
[16] M. M. Calor-Filho and J. C. Machado, "Measurement of the ultrasonic attenuation coefficient of human blood plasma during clotting in the frequency range of 8 to 22 MHz," Ultrasound Med. Biol. **32**, 1055–1064 (2006).
[17] J. C. Machado, A. Lenzi, W. G. Silva, and S. R. Sigelmann, "An ultrasonic method to measure human plasma coagulation time," J. Acoust. Soc. Am. **90**, 1749–1753 (1991).
[18] J.-L. Gennisson, S. Lerouge, and G. Cloutier, "Transient elastography: A new method to assess viscoelastic properties of blood during coagulation," Ultrasound Med. Biol. **32**, 1529–1537 (2006).
[19] M. Berson, L. Vaillant, F. Patat, and L. Pourcelot, "High-resolution real-time ultrasonic scanner," Ultrasound Med. Biol. **18**, 471–478 (1990).
[20] X. Lai and H. Torp, "Interpolation methods for time-delay estimation using cross-correlation method for blood velocity measurement," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 277–290 (1999).
[21] L. E. Kinsley, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
[22] M. F. Insana and T. J. Hall, "Parametric ultrasound imaging from backscatter coefficient measurements: Image formation and interpretation," Ultrason. Imaging **12**, 254–267 (1990).
[23] M. F. Insana, R. F. Wagner, D. G. Brown, and T. J. Hall, "Describing small-scale structure in random media using pulse-echo ultrasound," J.

Acoust. Soc. Am. **87**, 179–192 (1990).
[24]M.-C. Trzeciak and M.-H. Denninger, *L'hémostase en Question (The Hemostasis into Question)* (in French) (Biomérieux, Lyon, 2004).
[25]K. G. Mann, "Biochemistry and physiology of blood coagulation," Thromb. Haemostasis **82**, 165–174 (1999).
[26]S. Butenas and K. G. Mann, "Blood coagulation," Biochemistry (Mosc.) **67**, 3–12 (2002).
[27]A. Voleisis, R. Kazys, L. Mazeika, R. Sliteris, B. Voleisiene, and P. Grybauskas, "Ultrasonic method for the whole blood coagulation analysis," Ultrasonics **40**, 101–107 (2002).
[28]M. Corredig, M. Alexander, and D. G. Dalgleish, "The application of ultrasonic spectroscopy of the study of gelation of milk components," Food Res. Int. **37**, 557–565 (2004).
[29]M. Matsukawa and I. Nagai, "Ultrasonic characterization of a polymerizing epoxy resin with imbalanced stoichiometry," J. Acoust. Soc. Am. **99**, 2110–2115 (1996).
[30]S. M. Razavian, D. Quemada, M. T. Guillemin, Y. Beuzard, and M. Boynard, "Contribution of an ultrasonic interferometry method (Echocell) to the determination of red blood cell sedimentation rate at low hematocrit," J. Phys. III **5**, 903–912 (1995).
[31]M. Boynard, J. C. Lelievre, and R. Guillet, "Aggregation of red blood cells studied by ultrasound backscattering," Biorheology **24**, 451–461 (1987).
[32]G. Cloutier and Z. Qin, "Ultrasound backscattering from non-aggregating and aggregating erythrocytes—A review," Biorheology **34**, 443–470 (1997).
[33]S. Y. Kim, I. F. Miller, B. Sigel, P. M. Consigny, and J. Justin, "Ultrasonic evaluation of erythrocyte aggregation dynamics," Biorheology **26**, 723–736 (1989).
[34]H. Kitamura, B. Sigel, J. Machi, E. J. Feleppa, J. Sokil-Melgar, A. Kalisz, and J. Justin, "Roles of hematocrit and fibrinogen in red cell aggregation determined by ultrasonic scattering properties," Ultrasound Med. Biol. **21**, 827–832 (1995).
[35]C. Mahony, J. Ferguson, and P. L. C. Fischer, "Red cell aggregation and the echogenicity of whole blood," Ultrasound Med. Biol. **18**, 579–586 (1992).
[36]K. K. Shung and J. M. Reid, "Ultrasonic instrumentation for hematology," Ultrason. Imaging **1**, 280–294 (1979).
[37]M. Kaibara, "Rheology of blood coagulation," Biorheology **33**, 101–117 (1996).
[38]J. Wang, R. Feng, and Z.-H. Chen, "Ultrasound investigation of human blood coagulation," Chin. Sci. Bull. **35**, 48–51 (1990).

Phonation behavior of cooperatively foraging spinner dolphins

Kelly J. Benoit-Bird[a)]
College of Oceanic and Atmospheric Sciences, Oregon State University, 104 COAS Admin. Building, Corvallis, Oregon 97330

Whitlow W. L. Au
Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

(Received 2 January 2008; revised 20 June 2008; accepted 23 June 2008)

Groups of spinner dolphins have been shown to cooperatively herd small prey. It was hypothesized that the strong group coordination is maintained by acoustic communication, specifically by frequency-modulated whistles. Observations of groups of spinner dolphins foraging at night within a sound-scattering layer were made with a multibeam echosounder while the rates of dolphin sounds were measured using four hydrophones at 6 m depth intervals. Whistles were only detected when dolphins were not foraging and when animals were surfacing. Differences in click rates were found between depths and between different foraging stages but were relatively low when observations indicated that dolphins were actively feeding despite the consistency of these clicks with echolocation signals. Highest click rates occurred within the scattering layer, during transitions between foraging states. This suggests that clicks may be used directly or indirectly to cue group movement during foraging, potentially by detecting other individuals' positions in the group or serving a direct communicative role which would be contrary to the existing assumption that echolocation and communication are compartmentalized. Communicating via clicks would be beneficial as the signal's characteristics minimize the chance of eavesdropping by competing dolphins and large fish. Our results are unable to support the established paradigm for dolphin acoustic communication and suggest an alternate coordination mechanism in foraging spinner dolphins. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.2967477]

PACS number(s): 43.80.Ka, 43.30.Sf [JAS] Pages: 539–546

I. INTRODUCTION

Spinner dolphins, *Stenella longirostris*, are a pelagic subtropical species that feed on small mesopelagic prey one at a time (Cadenat and Doutre, 1959; Fitch and Brownell, 1968; Norris and Dohl, 1980; Perrin *et al.*, 1999). In Hawaiian waters, their prey migrate both vertically and horizontally and are only accessible to foraging dolphins for a limited amount of time during the night (Benoit-Bird and Au, 2003; Benoit-Bird *et al.*, 2001). Yet, spinner dolphins, as mammals, have relatively high energetic needs. Benoit-Bird (2004) used simple models to predict foraging needs of spinner dolphins in Hawaii and found that to meet its basic energetic needs, the average dolphin would need to consume 3–4 average sized prey/min during the time that the prey is shallow enough for feeding. This work suggests that spinner dolphins are limited by their feeding efficiency rather than by the abundance of food so that the density of prey rather than its total biomass determines feeding success (Benoit-Bird, 2004). Spinner dolphins in Hawaii seem to deal with this problem by actively aggregating their prey (Benoit-Bird and Au, 2003; Benoit-Bird and Au, 2009) They do so in highly stereotyped social behavior, foraging cooperatively at night in groups of 16–28 (Benoit-Bird and Au, 2009). Benoit-Bird and Au (2009) showed that groups of spinner dolphins used precise coordination in strict four-dimensional patterns to increase prey density by up to 200 times using the prey's own avoidance behavior (Fig. 1). Dolphins then took turns feeding within the aggregation they created. As a result, each dolphin had more access to prey than it would if feeding individually, despite the costs of participating in the group, meeting the criteria for cooperation.

The highly coordinated foraging behavior of spinner dolphins leads us to ask how they facilitate these interactions. Vision is of limited utility underwater where light is attenuated rapidly and is probably particularly ineffective for these nocturnally foraging animals. Sound, however, travels well in the ocean and is important in mediating odontocete social interactions (Tyack, 1999). Delphinids produce sounds that can be classified into two categories; (1) tonal frequency-modulated whistles and (2) short broadband clicks (Herman and Tavolga, 1980). Whistles are used primarily in social contexts (Herman and Tavolga, 1980) and may play a key role in maintaining contact between individuals (Janik, 2000b). Broadband clicks can be further broken down into two groups based on both their acoustic characteristics and their hypothesized functions. Clicks can be used by dolphins to perceive their environment: to find prey, avoid obstacles and predators, termed echolocation (Au, 1993). Clicks used for echolocation are typically temporally spaced to allow the sound to travel to the object of interest, return to the dolphin, and be processed (Au, 1993). Clicks are also produced much more quickly in trains termed "burst pulses" (Herman and Tavolga, 1980) which are thought to play a role in social interactions and are often observed in agonistic interactions

[a)]Author to whom correspondence should be addressed. Electronic mail: kbenoit@coas.oregonstate.edu.

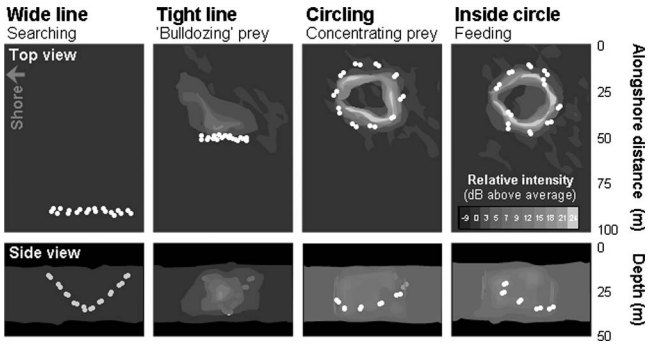


FIG. 1. Reprinted from Benoit-Bird and Au (2009) showing the four phases observed during spinner dolphin foraging from the top and side views. For the top view, scattering was integrated through the entire water column for each m^2. For the side view, scattering was integrated for the 20 m immediately in front of the dolphin group. The relative scattering intensity of prey is indicated by grayscale. Prey density can be related to the scattering intensity with each 3 dB increase in scattering representing a doubling of prey density. The positions of the air passage echoes from dolphins are shown as dots. In the side view, the dolphins behind the integration plane are shown a darker shade.

(Caldwell and Caldwell, 1967; Dawson, 1991; Norris *et al.*, 1994). Clicks may also produce in rapid trains in an echolocation role. Termed "terminal buzzes," these short interclick interval pulsed signals are produced when a dolphin is within a few body lengths of a target and are thought to provide rapid information updates in the terminal phase of food capture (Madsen *et al.*, 2005). We hypothesize that foraging spinner dolphins maintain social cohesion through acoustic communication. We predict that the maintenance of group contact will lead to the production of many whistles within foraging groups and that echolocation clicks will be detected most often when animals are attempting to capture prey.

The relationship between dolphin vocalizations and behavior is difficult to assess because most behavior and acoustic signaling by wild dolphins occur underwater. A variety of approaches have been utilized to associate behavior with acoustic signals in natural settings. For example, tags capable of recording acoustic signals along with dive depth and orientation have elucidated much about beaked whale behavior (see Madsen *et al.*, 2005, for example). Towed hydrophone arrays have been combined with observations of surface behavior to observe spatial relationships between animals and associated behaviors (Lammers *et al.*, 2006). Acoustic recordings have been combined with underwater video observations to associate specific sounds with behaviors (Herzing, 1996; Thomas *et al.*, 2003). Each of these approaches has strengths and limitations. Tagging requires that the animal be capable of carrying the instrument package and is relatively invasive. Both the video and surface observation approaches require that animals be visually observed. Because of the small size of spinner dolphins and the limited success of previous tagging attempts in Hawaii (Norris *et al.*, 1994), tags are not currently a viable approach for studying the use of sound by foraging spinner dolphins. Their nocturnal habit and primarily underwater feeding make any visual approach ineffective as well. We used new sonar techniques to quantitatively track foraging dolphins and their prey simultaneously (Benoit-Bird and Au, 2009) while recording their acoustic signals at four different depths. Our objective was to determine what sounds were being produced by foraging spinner dolphins and how these sounds were associated with various stages in their foraging pattern.

II. METHODS

To measure the distribution of dolphins and their prey, two sites along the leeward coast of Oahu were surveyed with active sonar techniques. Surveys were conducted between 2100 and 0300 h local time during ten nights in May and June of 2003. Sampling covered approximately 11 km surrounding Makua Bay and 10 km surrounding Electric Beach. All sampling was conducted on linear transects running parallel to the shoreline. These transects were located between 1.0 and 2.2 km from the shoreline and a total of 186 km were surveyed.

Surveying alternated between a 200 kHz Kongsberg-Simrad-Mesotech MS 2000 multibeam sonar and a Simrad EK-60 echosounder at 200 kHz. The instruments were selected because 200 kHz is outside the hearing range of this genus of dolphin and thus is likely to have the least impact (Kastelein *et al.*, 2003). The multibeam sonar provided three-dimensional information on the distribution of dolphins and their prey while the vessel was moving. The head of the multibeam sonar was mounted off the starboard side of the vessel, approximately 0.5 m below the water's surface, at a 45° angle to the right of the direction of travel and 30° up from the vertical plane perpendicular to the bottom rather than looking directly downward from the vessel. This permitted us to observe animals from the side, rather than having to drive directly over them, which, based on previous experience, limits the potential impacts of the survey on the dolphins. The multibeam system used a 150 μs long outgoing pulse transmitted at a rate of 5.6/s with a source level of 215 dB re 1 μPa. The system had a resolution of 0.22 m in the direction 30° from vertical resulting in a true depth resolution of 0.19 m. The system had 120, $1.5° \times 20°$ beams that overlapped by 0.25° in the narrower, nearly across-track direction, providing an angular coverage 120° with a 1° resolution in this direction. Data were taken using the external imaging transducer of the sonar, thus forming a Mills Cross to provide the greatest spatial resolution, giving a received beam width of 1.5° in the direction 45° from along track. The sonar system was set to cover a range of 65 m. In the geometry used, the instrument thus had a maximum range of 37.5 m from the vessel, giving a swath width of 113 m at the furthest point in its range and a depth range of 50 m. Further technical details about the sonar can be found in Cochrane *et al.* (2003). As in Benoit-Bird and Au (2009), the multibeam echosounder system was calibrated using a reference sphere following the protocol established by Foote *et al.* (2005).

Data from the multibeam were displayed in real time using the system's native software. The prey were sufficiently dense to be measured as volume scattering rather than as individual targets (Benoit-Bird and Au, 2009) while the dolphins were detected as single targets. The scattering characteristics of prey in terms of individual target strength can be found in Benoit-Bird and Au (2001) and in terms of vol-

ume scattering in several other papers (Benoit-Bird and Au, 2006; Benoit-Bird and Au, 2009; Benoit-Bird *et al.*, 2001). Previous active sonar studies have found that spinner dolphins have a unique set of scattering characteristics that allow them to be separated from other animals in the environment (Benoit-Bird and Au, 2003). The scattering strength of dolphins is much higher than that of their prey because of the difference in size between them (~2 m versus ~10 cm) and the air-filled lungs and sinus cavities of the dolphins. This permitted dolphins to be detected in real time with the multibeam sonar. A total of 323 groups of dolphins were detected with the multibeam sonar. During some of these real-time detections, movement of the vessel was stopped with the dolphins within range of the multibeam and a series of hydrophones was used to record sounds produced by dolphins. Four BM8263 wideband omnidirectional hydrophones (Biomon, Santa Barbara, CA), each consisting of a 1.3 cm diameter PST spherical element with a sensitivity of −155 dB re 1V/μPA, were deployed at 6.4 m depth intervals. The first hydrophone was deployed at a depth of 17 m if the top of the layer was at least 15 m deep or 3 m if the layer was close to the surface. The signals were low pass filtered at 50 kHz before recordings from each of the four hydrophones were made using a Measurement Computing PCI-DAS4020/12 analog to digital converter capable of simultaneous sampling on four channels. This board was operated using custom LABVIEW 6I software and a transportable computer. Each channel was sampled simultaneously for 34 s at a rate of 125 000 samples/s. This provided a Nyquist frequency of 64.5 kHz, precluding full spectral analysis of signals but not affecting temporal identification of clicks. Data collection alternated between recording from the hydrophones and the multibeam. The multibeam sonar collected for approximately 20 s of data between each of the hydrophone recordings. These combined hydrophone and multibeam sonar recordings were made from 56 groups of dolphins with each session lasting between 2 and 16 min (mean=7 min).

A. Multibeam sonar analysis

Details of data analysis can be found in Benoit-Bird and Au (2009). Briefly, raw data from the multibeam sonar were saved to hard disk, beam formed, and the amplitude and range of echoes in each beam were extracted using a MATLAB program. The range versus azimuth data were transformed into a rectangular coordinate system where one axis is along the direction of travel, the second is across the direction of travel, and the third is depth. To obtain three-dimensional information, six successive sonar returns, or about 1 s of data, were combined. All position data were corrected using coupled differential global positioning system (GPS) to eliminate the movement of the vessel. GPS readings were taken with each sonar signal, at a rate of 5.6/s. Using this transformed, averaged, and position corrected data, isoscattering surfaces could then be identified.

Dolphins were identified from the data based on their unique set of acoustic scattering characteristics (Benoit-Bird and Au, 2003), most notably, a very strong echo that is caused by their air filled lungs (Au, 1996) and possibly their nasal passages that is surrounded by remarkably low-amplitude echoes presumably by their blubber which is nearly impedance matched to seawater. Echoes from spinner dolphins identified during daylight surfacings were taken nearly simultaneously with a calibrated 200 kHz echosounder (Simrad EK60) and the SM2000. Echoes from spinner dolphins SM2000 had values between −25.9 and −28.2 while the values measured with the EK60 were within 0.3 dB of −27.1 dB. The SM2000 echoes showed more variance in their amplitude, likely because of the increase in the variability in insonified angle relative to the narrow beam EK60. In both systems, the presumptive lung echo provided an extremely high target strength within this field of ~ −27 dB echoes. The length of animals measured with the multibeam was consistent with the size for adult spinner dolphins in Hawaiian waters. For analysis of dolphin position, the single strongest echo from each animal, likely its lungs, was utilized, creating a standardized method of locating the animal in space and measuring distance between animals.

The sonar system has a resolution of 1.5° in the azimuth and range direction and 9.6 cm in the range direction. To increase the resolution of measurement of animal spacing, information from six successive pings was used to calculate interanimal distances. Changes in spacing in this 1 s time period were assumed to be negligible. The movement of the sonar relative to the animals during this time, however, decreased the minimum interspacing distance that could be measured in all dimensions to approximately 10 cm.

Benoit-Bird and Au (2009) categorized the foraging behavior of these animals using the multibeam sonar into four distinct stages, summarized briefly here [Fig. 1, reprinted from Benoit-Bird and Au (2009)]. First, 8–14 pairs of dolphins swam in a widely space line in which they search for an existing patch within the prey field. Second, dolphins reduced their interpair spacing, a stage termed "tight line," and began to undulate their swimming path up and down. During this phase of foraging, the density of prey began to increase as the prey avoided the dolphins and piled up on itself as snow in front of a plow. During the third phase of foraging, dolphins formed a circle around the increased density prey patch which distributed the prey more evenly over the circle. Finally, individual pairs of dolphins moved inside the circle to actively feed on the most dense regions of the patch while other animals continued to maintain the patch by swimming around it. Two pairs of dolphins moved into the circle at the same time, from opposite sides of the circle for about 10 s of feeding before taking their place at the circle's edge and allowing the pairs behind them into the circle. Each pair of dolphins got an average of 45 s inside the region of most dense prey in each approximately 5 min foraging bout.

We used these categories to characterize the observed dolphins' foraging stage observed with multibeam data taken immediately before and after each recording. If the stage before and after the recording was the same, the recording was labeled using that stage. If, however, the stage changed during the recording, the recording was labeled as the transition between the stages.

TABLE I. Univariate analysis of variance for dolphin click rate (clicks/hydrophone/30 s).

Source	SS	*df*	MS	*F*	*p*
Foraging stage	1 211 232	8	15 140 4	150.7 7	<0.001
Time	1 253	5	251	0.25	0.94
Date	6 800	8	850	0.85	0.56
Depth	10 338	7	1477	1.47	0.17
Hydrophone position	819 930	2	40 996 5	408.2 4	<0.001
Number of dolphins	10 030	6	1 672	1.66	0.13
Error	1 901 000	1893	1 004		
Total	7 117 561	1972			

B. Dolphin signal analysis

The waveforms and spectra from each hydrophone of 34 s recordings from 56 recording sessions were displayed using COOL EDIT software. Echolocation clicks and whistles were confirmed aurally and counted from the spectra. Any signal detected on more than one hydrophone was attributed to the hydrophone that received the highest amplitude whistle. Clicks were only rarely detected on more than one hydrophone and the number of whistles detected was low so no localizations were attempted.

Univariate analyses of variance (ANOVAs) were used to test the effects of the depth of the hydrophones, the position of hydrophones relative to the scattering layer, sampling day, sampling hour, the number of dolphins, and the foraging state of the dolphins on click and whistle rates (Tables I and II). Dunnett's C *post hoc* tests with Bonferonni corrections for multiple comparisons were used to test the effects of the position of hydrophones relative to the scattering layer on measured echolocation rates. Dunnett's C *post hoc* tests with Bonferonni corrections for multiple comparisons were conducted to test differences between foraging states using only the data from hydrophones within the scattering layer.

III. RESULTS

Direct testing of potential effects of the vessel and the sonars was not possible. Indications of changes in dolphin behavior—e.g., increases in audible surfacing rate, bowriding, playing at the surface, and swimming around the boat—were noticed during a few times when the vessel was moving unpredictably for reasons unassociated with sampling. However, these types of behaviors were not detected during normal transect sampling, suggesting limited effects of the vessel or the instruments on the dolphins during sampling.

Univariate ANOVAs were used to test the effects of the depth of the hydrophones, the position of hydrophones relative to the scattering layer, sampling day, sampling hour, the number of dolphins, and the foraging state of the dolphins on click and whistle rates (Tables I and II). For both clicks and whistles, there was a significant difference in their detected rate as a function of foraging stage. Note that there is no significant effect of the depth of the hydrophones on either click or whistle rate but there is a significant effect of the position of the hydrophones relative to the scattering layer on the rate of detected clicks.

Dunnett's C *post hoc* tests with Bonferonni corrections for multiple comparisons were used to test the effects of the position of hydrophones relative to the scattering layer on measured echolocation rates (Fig. 2). Echolocation rates measured by hydrophones within the scattering layer were significantly higher than the rates measured by hydrophones above and below the layer. No significant differences in echolocation rates were observed between hydrophones above and below the scattering layer (Fig. 3).

Dunnett's C *post hoc* tests with Bonferonni corrections for multiple comparisons were conducted to test differences

TABLE II. Univariate analysis of variance for dolphin whistle rate (whistles/hydrophone/30 s).

Source	SS	*df*	MS	*F*	*p*
Foraging stage	37.53	8	4.7	138.02	<0.001
Time	0.01	5	0.0	0.04	0.99
Date	0.04	8	0.0	0.14	0.99
Depth	0.50	7	0.1	2.10	0.04
Hydrophone position	0.16	2	0.1	2.35	0.10
Number of dolphins	2.05	6	0.3	10.07	<0.001
Error	64.34	1 893	0.0		
Total	117.00	1 972			

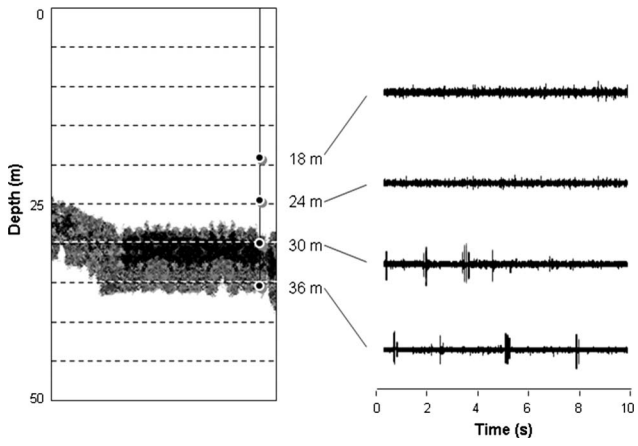


FIG. 2. A sample echogram showing the scattering layer over approximately 2 min while dolphins were circling. Local time was 22:18 h. The depths of deployed hydrophones are shown. Only the two lower hydrophones were within the scattering layer. Waveforms recorded from each of the four hydrophones are shown. The high amplitude pulses from the two deeper hydrophones are consistent with clicks from spinner dolphins.

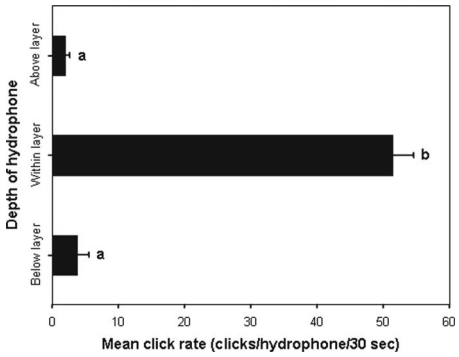


FIG. 3. Mean click rate recorded as a function of hydrophone position. Data that were statistically indistinguishable from each other but statistically distinct from other categories in the analysis are designated by lowercase letters ($p<0.05$).

in click and whistle detections between foraging states using only the data from hydrophones within the scattering layer. Results are shown in Figs. 4 and 5.

The interval between clicks in series was measured to determine if clicks were consistent with echolocation or burst pulses and terminal buzzes. Because we were unable to record long click trains, the interclick interval was only measured when at least one other interval was observable before and after it. These beginning and end intervals were discarded. A histogram of the measured intervals is shown in Fig. 6.

The potential relationship between the spacing of circling dolphins and the spacing of clicks recorded during our observations was examined. The amount of time between pairs of dolphins passing a stationary point was compared with the time between recorded groups or bursts of clicks. Note that the term "click train" is not used here because we cannot be certain that entire trains were indeed recorded. The center of each recorded click series was used in this measurement (Fig. 7). The relationship between the spacing of dolphins and the spacing of clicks was linear with a slope of

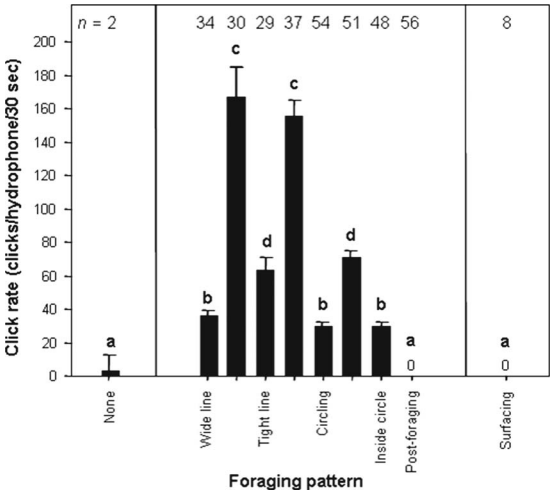


FIG. 4. Click rate as a function of foraging stage. Error bars represent the 95% confidence interval about the mean. Data that were statistically indistinguishable from each other but statistically distinct from other categories in the analysis are designated by lowercase letters ($p<0.05$).

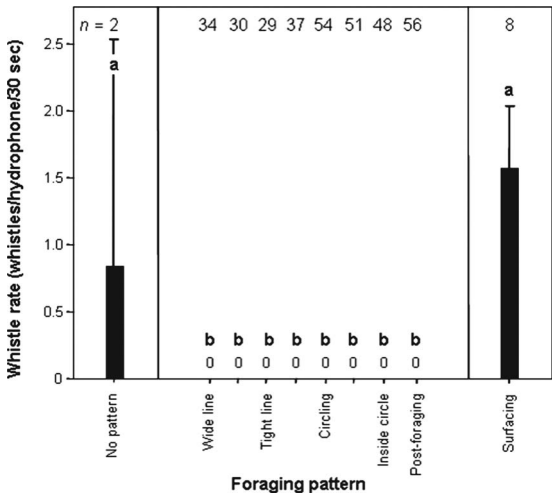


FIG. 5. Whistle rate as a function of foraging stage. Error bars represent the 95% confidence interval about the mean. Data that were statistically indistinguishable from each other but statistically distinct from other categories in the analysis are designated by lowercase letters ($p<0.05$).

approximately 1 (interval between click bursts = 1.09 interval between dolphin pairs—0.19; $R^2=0.56$; F test of slope $p<0.01$).

IV. DISCUSSION

As detailed in Benoit-Bird and Au (2009), the observed behavior of spinner dolphins is highly suggestive of foraging. The highly coordinated behavior of spinner dolphins during presumptive foraging suggests that there must be some method of synchronizing the activity of these cooperative groups. Animals were observed to transition from one distinct behavior to another nearly simultaneously with animals spread out over a distance of about 40 m. Vision is unlikely to be effective at these ranges underwater, especially at night. Based on previous studies of dolphin behavior, we propose that spinner dolphin acoustic methods of maintaining group cohesion while foraging. We hypothesized that whistles, shown to play a role in maintaining contact over these distances (Lammers *et al.* 2006), would be the primary type of acoustic signal used.

However, we recorded very few whistles and none when animals were engaged in behaviors identified with foraging.

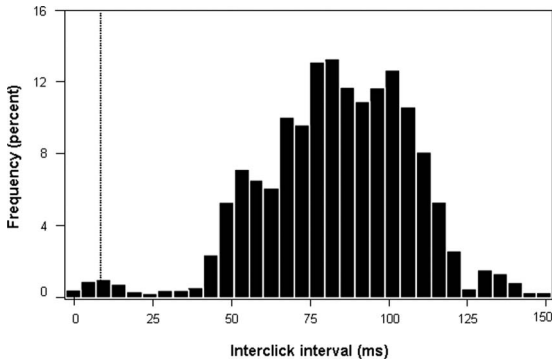


FIG. 6. Histogram of all measured interclick intervals. More than 95% of the clicks are at least 10 ms apart, indicated by the dashed line, consistent with echolocation signals rather than burst pulses.

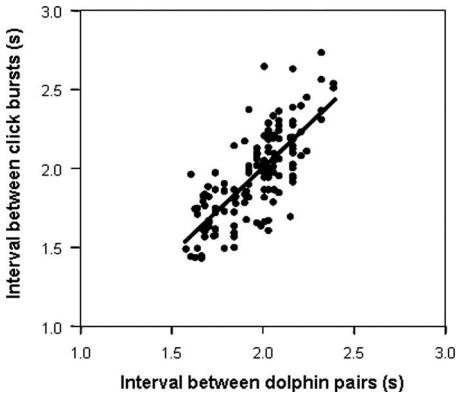


FIG. 7. The relationship between the interval between dolphin pairs swimming in a circle and the interval between detected click bursts. The relationship is linear with a slope of approximately 1.

This result is not likely an artifact of measurement technique or signal propagation. Spinner dolphin whistles are relatively low in frequency, tens of kilohertz (Bazua-Duran and Au, 2002), and are attenuated relatively slowly in seawater. While whistles have mixed directionality (Lammers and Au, 2003), they are omnidirectional at their lower frequencies which were well within our recording bandwidth. All of this suggests that if whistles were being produced by the observed groups, they would have been detected. The hypothesis that whistles are used for group coordination cannot be supported. However, this does not rule out acoustic signaling as a means of maintaining the behavior.

Click signals from spinner dolphins were recorded. Clicks in rapid repetition rate "burst pulsed" click trains have been observed to have a communicative function (Herman and Tavolga, 1980). These clicks are distinctive from clicks typically observed in echolocating animals in that their interclick interval is extremely short. Lammers *et al.* (2003) found that burst pulsed clicks in spinner dolphins typically had an interval less than 10 ms. Less than 2% of the clicks recorded had an interclick interval less than this threshold, suggesting that burst pulses are not a large component of signaling in these foraging groups. These results also suggest that terminal buzzes found which have been observed in other species of dolphins while feeding were either not produced or not recorded.

More than 98% of the clicks were consistent with echolocation based on their interclick interval. We hypothesized that echolocation would be used primarily to target individual prey before capture as spinner dolphins feed on prey one at a time. This would suggest that clicks should be identified most often when animals were inside the circle and actively feeding (Benoit-Bird and Au, 2009). However, the number of clicks detected when animals were inside the circle was amongst the lowest measured. In addition, interclick interval showed modes at 55, 80, and 100 ms. Given an approximately 20 ms processing lag time associated with echolocation clicks (Au, 1993), this suggests a two-way transit time equivalent to targets at approximately 26, 45, and 60 m, respectively. If these are indeed echolocation clicks, they are more consistent with animals echolocating off features at the opposite side of the circle of dolphins (and the patch of prey) than with targets at close range that they are attempting to capture.

Further examination of the timing of click detection suggests another use for the detected clicks. Clicks were not produced most often within the discreet foraging stages but rather during the transitions between stages when animals were changing formation. A large number of clicks were detected when animals were moving from the wide line to the tight line and then again when the tight line began to form a circle. This suggests that clicks may be important in coordinating the group. Dolphins may be using clicks to gain information about the prey field at these critical decision points. Perhaps a threshold rule of prey density or distribution is employed for changing foraging stages and the animals are collectively testing for this threshold. They may also be using echolocation to track the changing positions of other members of the group which could be supported by the relatively long target ranges indicated by the measured interclick intervals. Another possibility is that the clicks themselves are being used to transfer information between individuals, serving as a form of communication rather than a sense. Information could be contained in any number of signal parameters of clicks including their temporal spacing that could serve to direct the movements of the group. Social communication via echolocationlike clicks has been suggested in cetaceans that do not produce whistles (Amundin, 1991; Dawson and Thorpe, 1990; Dawson, 1991) and in sperm whales (Watkins, 1980) supporting the hypothesis that these click signals that evolved for one purpose could be developed for others (Tyack, 1997).

The two possibilities raised here for the use of clicks in foraging spinner dolphins are difficult to test. While we were able to record clicks, we were unable to record single clicks on multiple hydrophones which would have allowed us to localize the signal to its production site and perhaps correlate that with the multibeam sonar observations to identify the individual signaler. This suggests that we likely missed many of the clicks produced. Dolphin clicks are high in frequency, typically with a peak frequency over 100 kHz, and thus attenuate rather quickly in water. In addition, clicks are highly directional with a half power (3 dB) beam angle of about 10° (Au, 1993), meaning that clicks are difficult to detect unless the receiver is within this narrow transmission cone. Further evidence of this can be seen in the depth at which clicks were recorded. Clicks were only recorded when the hydrophones were within the scattering layer which was also the depth range at which dolphins were present, suggesting we were only collecting clicks from animals close to the depth of each hydrophone. In addition, we were likely only recording clicks from animals directing their beam at the hydrophone. This is evidenced by the data shown in Fig. 7 relating the time interval between pairs of dolphins swimming in a circle past a fixed point and the interval between successive bursts of clicks. Individual pairs of dolphins within a foraging circle were probably each clicking in a directly relative to their position in the circle so as one pair passed, the hydro-

phone near that depth would be within the beams of those animals for a short period of time related to their distance and swimming speed.

After passing out of those individuals' beams, another pair slightly deeper in depth would pass similarly, providing a regular pattern to the collection of clicks as observed. Despite these difficulties in detecting clicks, the observed differences in click rate between different foraging stages likely represent important differences in how clicks are being used with these different behaviors, either through changes in click production rate or other changes such as sweeping of clicks over a wider area that might be detected as rate changes.

The potential use of clicks directly for communication has not been investigated. Most studies in dolphin echolocation have taken place under carefully controlled conditions with captive animals and artificial targets. We know very little about how wild dolphins use echolocation to solve tasks such as avoiding obstacles or detecting, selecting, and capturing prey (Tyack, 1999). Relatively recent field observations using arrays (e.g., Lammers *et al.*, 2006) and acoustic recording tags (e.g., Madsen *et al.*, 2005) are beginning to change this. However, concurrent observations of underwater behavior and acoustics of dolphins and prey behavior remain limited. The idea that dolphins could use existing signals for multiple purposes should not be surprising. Clicks have significant advantages for communication over whistles as pointed out specifically for burst pulsed signals by Lammers *et al.* (2006). For the same reasons that we had a difficult time detecting clicks, so would other animals in the habitat. Communication with clicks that have a relatively short active space, or radius of effective communication, of a few hundreds of meters at most (Au, 1993) and can be directed toward a recipient seems highly advantageous over using whistles which have a potential active space of a few tens of kilometers (Janik, 2000a) and cannot be focused when relaying information about a highly valuable resource. The animals in this group may keep noncooperative individuals from finding the highly dense prey patch they created by reducing eavesdropping. The mesopelagic prey layer that spinner dolphins feed upon is also an important food resource for a variety of fish including snappers (Haight *et al.*, 1993), tunas (He *et al.*, 1997), and billfish (Skillman, 1998). Most fish do not hear well at the high frequencies in dolphin clicks though they can likely hear the low-frequency components of whistles (Popper and Fay, 1999). Competing predatory fish would have even more difficulty than conspecifics using eavesdropping on a cooperating group of dolphins communicating via clicks, limiting the potential for the dolphins to lose the dense patch of prey they created. No information is available on the hearing of myctophid fishes, the primary prey found in these dense patches. However, most fish can only hear at frequencies up to a few kilohertz so it is possible that the use of high-frequency communication signals could limit the ability of the prey to detect the dolphins as well.

Group coordination of spinner dolphins during foraging may be maintained by other sensory modalities that cannot be examined in these data. For example, the prey of spinner dolphins is bioluminescent with photophores that emit light when they are disturbed. Perhaps dolphins create a "wake of light" as they swim through the layer of prey. Dolphin-stimulated bioluminescence has been observed in bioluminescent zooplankton (Rohr *et al.*, 1998); however, there is no evidence available for the induction of bioluminescence by predators for the animals that serve as dolphin prey. Perhaps a combination of sensory stimuli play a role in both the decision making processes involved in prey herding and the communication, either intentional or as a by-product of other behavior, that occurs between individuals to maintain group coordination.

Our results suggest that while the acoustic signals that have been shown in other studies to serve a communicative role were noticeably absent, clicks consistent with echolocation were correlated with changes in coordinated groups of foraging spinner dolphins. These data cannot separate the use of these click signals for sensing of the prey field, leading to a decision of all individuals to exhibit a behavior from the direct use of the signals for communication. However, the extremely rapid and tightly synchronized movements observed among up to 28 individuals suggest that the mechanism is highly efficient. The changes in prey density observed were graded and varied over space which would make synchronized decision making via a threshold difficult. Some form of communication would seem necessary to coordinate the group unless the behavior is so highly stereotyped that no variability in the prey field is considered once the herding behavior begins and all individuals simply fulfill their expected roll. This is not supported by the variability in the timing between transitions in the greater data set presented in Benoit-Bird and Au (2009) Communication might be limited to a few simple cues at transitions, similar to a coach's command for a new play in a football game. All players (animals) would know what to do based on a single short command. This communication method would suggest that animals play a consistent part in the foraging from one bout to another and that groups are maintained over some time period longer than a single, 5 min foraging bout. Data are not available to evaluate these predictions. Whatever the mechanism by which clicks are involved in group coordination, the results highlight the potential flexibility of delphinid acoustic signals, contrary to the existing assumption that echolocation and communication are separate and compartmentalized (Tyack, 1997). Our results are unable to support the established paradigm for dolphin acoustic communication and suggest an alternate communication mechanism for foraging spinner dolphins.

ACKNOWLEDGMENTS

The use of the multibeam sonar was generously provided by Jeff Condiotty of Simrad, Lynwood, WA. Marc Lammers wrote the program that recorded dolphin signals and assisted with fieldwork. Dieter Lammers designed and constructed the multibeam transducer mount and served as our vessel's skipper. Michiel Schotten, Aran Mooney, Kristen Taylor, and Stuart Ibsen provided assistance in the field. Christopher Bird repaired the multibeam hardware after a

technical malfunction in the field. Allen, Bob Asplen, and Steve Carlton of Kongsberg Simrad Mesotech graciously provided emergency assistance with the multibeam. This work was funded by Grant No. 0205752 from the National Science Foundation and Grant No. N00014-02-1-0968 from the Office of Naval Research. This work was conducted under US National Marine Fisheries Service Permit No. 1000-1617.

Amundin, M. (**1991**). "Click repetition rate patterns in communicative sounds from the harbour porpoise, *Phocoena phocoena*," University of Stockholm.

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

Au, W. W. L. (**1996**). "Acoustic reflectivity of a dolphin," J. Acoust. Soc. Am. **99**, 3844–3848.

Bazua-Duran, C., and Au, W. W. L. (**2002**). "The whistles of Hawaiian spinner dolphins," J. Acoust. Soc. Am. **112**, 3064–3072.

Benoit-Bird, K. J. (**2004**). "Prey caloric value and predator energy needs: Foraging predictions for wild spinner dolphins," Mar. Biol. (Berlin) **145**, 435–444.

Benoit-Bird, K. J., and Au, W. W. L. (**2001**). "Target strength measurements of animals from the Hawaiian mesopelagic boundary community," J. Acoust. Soc. Am. **110**, 812–819.

Benoit-Bird, K. J., and Au, W. W. L. (**2003**). "Prey dynamics affect foraging by a pelagic predator (*Stenella longirostris*) over a range of spatial and temporal scales," Behav. Ecol. Sociobiol. **53**, 364–373.

Benoit-Bird, K. J., and Au, W. W. L. (**2006**). "Extreme diel horizontal migrations by a tropical nearshore resident micronekton community," Mar. Ecol.: Prog. Ser. **319**, 1–14.

Benoit-Bird, K. J., and Au, W. W. L., (2009), "Cooperative prey herding by the pelagic dolphin, *Stenella longirostris*," J. Acoust. Soc. Am. **125**, 125–137.

Benoit-Bird, K. J., Au, W. W. L., Brainard, R. E., and Lammers, M. O. (**2001**). "Diel horizontal migration of the Hawaiian mesopelagic boundary community observed acoustically," Mar. Ecol.: Prog. Ser. **217**, 1–14.

Cadenat, J., and Doutre, M. (**1959**). "Notes sur les delphinides Ouest-Africains. 5, Sur un *Prodelphinus* a long bec capture au large des cotes du Senegal: *Prodelphinus longirostris*," Bull. Inst. Fondam. Afr. Noire **21**, 787–792.

Caldwell, M. C., and Caldwell, D. K. (**1967**). "Intraspecific transfer of information via the pulsed sound in captive Odontocete Cetaceans," in *Animal Sonar Systems: Biology and Bionics*, edited by Busnel, R. G. (Laboratoire de Physiologie Acoustic, Jouy-en-Josas, France), pp. 879–936.

Cochrane, N. A., Li, Y., and Melvin, G. D. (**2003**). "Quantifications of a multibeam sonar for fisheries assessment applications," J. Acoust. Soc. Am. **114**, 745–758.

Dawson, S. M. (**1991**). "Clicks and communication: The behavioural and social contexts of Hector's dolphin vocalizations," Ethology **88**, 265–276.

Dawson, J., and Thorpe, L. (**1990**). "A quantitative analysis of the sounds of Hector's dolphin," Ethology **86**, 131–145.

Fitch, J. E., and Brownell, R. L., Jr. (**1968**). "Fish otoliths and their importance in interpreting feeding habits," J. Fish. Res. Board Can. **25**, 2561–2574.

Foote, K. G., Chu, D., Hammar, T. R., Baldwin, K. C., Mayer, L. A., Hufnagle, L. C., Jr., and Jech, J. M. (**2005**). "Protocols for calibrating multibeam sonar," J. Acoust. Soc. Am. **117**, 2013–2027.

Haight, W. R., Parrish, J. D., and Hayes, T. A. (**1993**). "Feeding ecology of deepwater lutjanid snappers at Penguin Bank, Hawaii," Trans. Am Fish. Soc. **122**, 328–347.

He, X., Bigelow, K. A., and Boggs, C. H. (**1997**). "Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery," Fish. Res. **31**, 147–158.

Herman, L. M., and Tavolga, W. (**1980**). "The communications systems of cetaceans," in *Cetacean Behavior: Mechanisms and Function*, edited by L. M. Herman (Wiley-Interscience, New York), pp. 149–209.

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

Janik, V. M. (**2000a**). "Source levels and the estimated active space of bottlenose dolphin (*Tursiops truncatus*) whistles in the Moray Firth, Scotland," J. Comp. Physiol., A **186**, 673–680.

Janik, V. M. (**2000b**). "Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*)," Science **289**, 1355–1357.

Kastelein, R. A., Hagedoorm, M., Au, W. W. L., and De Haan, D. (**2003**). "Audiogram of a striped dolphin (*Stenella coeruleoalba*)," J. Acoust. Soc. Am. **113**, 1130–1137.

Lammers, M. O., and Au, W. W. L. (**2003**). "Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): A signal feature to cue direction of movement?" Marine Mammal Sci. **19**, 249–264.

Lammers, M. O., Au, W. W. L., and Herzing, D. L. (**2003**). "The broadband social acoustic signaling behavior of spinner and spotted dolphins," J. Acoust. Soc. Am. **114**, 1629–1639.

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.

Madsen, P. T., Johnson, M. P., Aguilar de Soto, N., Zimmer, W. M. X., and Tyack, P. L. (**2005**). "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)," J. Exp. Biol. **208**, 181–194.

Norris, K. S., and Dohl, T. P. (**1980**). "Behavior of the Hawaiian spinner dolphin, *Stenella longirostris*," Fish. Bull. **77**, 821–849.

Norris, K. S., Würsig, B., Wells, R. S., and Würsig, M. (**1994**). *The Hawaiian Spinner Dolphin* (University of California Press, Berkeley, CA).

Perrin, W. F., Dolar, M. L. L., and Robineau, D. (**1999**). "Spinner dolphins (*Stenella longirostris*) of the western pacific and southeast Asia: Pelagic and shallow-water forms," Marine Mammal Sci. **15**, 1029–1053.

Popper, A. N., and Fay, R. R. (**1999**). "The auditory periphery in fishes," in *Comparative Hearing, Fish and Amphibians*, edited by R. R. Fay and A. N. Popper (Springer, New York), pp. 43–100.

Rohr, J., Latz, M. I., Fallon, S., Nauen, J. C., and Hendricks, E. (**1998**). "Experimental approaches towards interpreting dolphin-stimulated bioluminescence," J. Exp. Biol. **201**, 1447–1460.

Skillman, R. A. (**1998**). "Central Pacific swordfish, *Xiphias gladius*, fishery development, biology, and research," NOAA Technical Report No. NMFS 101–124.

Thomas, R. E., Fristrup, J. M., and Tyack, P. L. (**2003**). "Linking the sounds of dolphins to their locations and behavior using video and multichannel acoustic recordings," J. Acoust. Soc. Am. **112**, 1692–1701.

Tyack, P. L. (**1997**). "Studying how cetaceans use sound to explore their environment," in *Perspectives in Ethology*, edited by D. H. B. Owings, D. Michael, and Nicholas. S. Thompson (Plenum, New York), pp. 251–297.

Tyack, P. L. (**1999**). "Communication and cognition," in *Biology of Marine Mammals*, edited by Reynolds, J. E., III, and Rommel, S. A. (Smithsonian Institution Press, Washington and London), pp. 304–306.

Watkins, W. A. (**1980**). "Acoustics and behavior of sperm whales," in *Animal Sonar Systems*, edited by R. G. Busnel and F. E. Fish (Plenum, New York), pp. 283–290.

Comparison of stationary acoustic monitoring and visual observation of finless porpoises

Satoko Kimura[a)]
Graduate School of Informatics, Kyoto University, 606-8501 Kyoto, Japan

Tomonari Akamatsu
National Research Institute of Fisheries Engineering, Hasaki, Kashima, Ibaraki 314-0421, Japan

Kexiong Wang,[b)] Ding Wang, Songhai Li, and Shouyue Dong
Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, People's Republic of China

Nobuaki Arai
Graduate School of Informatics, Kyoto University, 606-8501 Kyoto, Japan

(Received 8 May 2008; revised 6 October 2008; accepted 13 October 2008)

The detection performance regarding stationary acoustic monitoring of Yangtze finless porpoises *Neophocaena phocaenoides asiaeorientalis* was compared to visual observations. Three stereo acoustic data loggers (A-tag) were placed at different locations near the confluence of Poyang Lake and the Yangtze River, China. The presence and number of porpoises were determined acoustically and visually during each 1-min time bin. On average, porpoises were acoustically detected $81.7 \pm 9.7\%$ of the entire effective observation time, while the presence of animals was confirmed visually $12.7 \pm 11.0\%$ of the entire time. Acoustic monitoring indicated areas of high and low porpoise densities that were consistent with visual observations. The direction of porpoise movement was monitored using stereo beams, which agreed with visual observations at all monitoring locations. Acoustic and visual methods could determine group sizes up to five and ten individuals, respectively. While the acoustic monitoring method had the advantage of high detection probability, it tended to underestimate group size due to the limited resolution of sound source bearing angles. The stationary acoustic monitoring method proved to be a practical and useful alternative to visual observations, especially in areas of low porpoise density for long-term monitoring. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3021302]

PACS number(s): 43.80.Ka, 43.80.Jz, 43.66.Hg [WWA] Pages: 547–553

I. INTRODUCTION

Visual observation of surfaced cetaceans is well established and has been applied widely to species ranging from small odontocetes to large baleen whales. However, data gathered using visual surveys are limited to daytime since this is the only time visual observation is possible. Weather conditions such as fog and glare also have considerable effects on the visibility of animals. Patient long-term visual observation can be very costly, particularly under the very low-density conditions of endangered species.

The Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*), a freshwater porpoise subspecies unique to the Yangtze River in China, is a typical example. In the early 1990s, the population size was estimated at approximately 2700 individuals (Zhang *et al.*, 1993). By 2006, estimates had decreased to as low as 1800 over the porpoise's entire distribution range (Zhao *et al.*, 2008). Recent genetic studies have confirmed that populations of the Yangtze finless porpoise are scattered throughout the habitat area (Zheng *et al.*, 2005), so monitoring them requires a great deal of effort. In addition, the Yangtze finless porpoise is one of the most difficult species to observe visually due to the turbid river's low visibility (less than 1 m) and the porpoise's lack of dorsal fin and rostrum.

Use of acoustic monitoring can avoid some major difficulties related to visual observation. Researchers have recently applied stationary acoustic monitoring methods to observe many species of aquatic mammals in various water systems. These methods are considered to be suitable for long-term automatic monitoring. The underwater sounds produced by aquatic animals can be used to monitor various characteristics of a species, including presence, behavior, and distribution (Nishimura and Conlon, 1994; Janik, 2000; Janik *et al.*, 2000; van Parijs *et al.*, 2002; Au and Benoit-Bird, 2003; Ichikawa *et al.*, 2006; Tsutsumi *et al.*, 2006). For example, researchers have used the T-POD (a passive acoustic porpoise or dolphin detector system) to monitor harbor porpoises and bottlenose dolphins (Thomsen *et al.*, 2005; Philpott *et al.*, 2007). The T-POD system can detect the presence and sensing effort of echolocating animals, indicated by the detection rate of clicks per hour or day (Thomsen *et al.*, 2005; Verfuß *et al.*, 2007), the number of minutes containing clicks (Carstensen *et al.*, 2006), the click characteristics of animals (Philpott *et al.*, 2007), and the length of interclick intervals (Leeney *et al.*, 2007). However, unlike visual observation, the T-POD system is not suitable for counting the

[a)]Electronic mail: sk0130@bre.soc.i.kyoto-u.ac.jp
[b)]Electronic mail: wangk@ihb.ac.cn

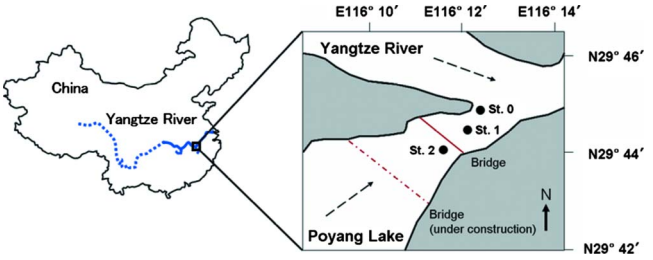


FIG. 1. (Color online) Study area around the junction of the Yangtze River and Poyang Lake, China. The solid line in the left panel indicates the historical habitat of Yangtze finless porpoises (Wei *et al.*, 2002). The dashed arrows in the right panel indicate the direction of the current.

specific number of animals because it is a monaural system. Wang *et al.* (2005) used the stereo acoustic data logger, A-tag (Little Leonardo, Tokyo, Japan), in an oxbow of the Yangtze River that contained an *ex situ* conservation area for finless porpoises. The researches found a weak positive linear correlation between the number of recorded signals and the group size of sighted porpoises. The number of signals may be an indicator of the number of individuals in a group, but because the sound production ratio varies between animals, Wang *et al.* (2005) were not successful in using stationary acoustic data to determine the exact number of porpoises.

Counting the number of finless porpoises from a moving boat is possible using the sound source bearing angle, monitored by a stereo acoustic system (Akamatsu *et al.*, 2008). Acoustic transect observation from a moving platform effectively prevents double-counting of animals, whereas stationary acoustic observations require additional evaluation by comparing with ground truth data such as visually observed number of animals to determine the number of animals. In this study, we used a stereo acoustic monitoring system to conduct a stationary counting of finless porpoises in the channel where Poyang Lake flows into the Yangtze River. We compared the detection performance of a stationary acoustic monitoring system to that of visual observations.

II. MATERIALS AND METHODS

A. Study area

We conducted simultaneous acoustic and visual observations from boats at the confluence of the Yangtze River and Poyang Lake located in the middle reaches of the Yangtze River in South-Central China (Fig. 1). Three stations were used in the study area in April 27–29, 2006 and May 9 and 10, 2007. Data were collected over a summed period of 5 days. Station 0 (29°45′06″ N, 116°12′41″ E) was located at the point where the lake joined the main channel of the Yangtze River. Station 1 (29°44′34″ N, 116°12′10″ E) was located at the mouth of the lake approximately 1300 m upstream from Station 0. Station 2 (29°44′02″ N, 116°11′47″ E) was situated between two bridges and was located approximately 1100 m upstream from Station 1. During observation, boats at each station were fixed using double anchors to minimize drifting. Each boat engine was completely stopped. Water depth was approximately 3 m at all stations.

B. Acoustic data logger

We used stereo acoustic data loggers, A-tag (Little Leonardo Ltd., Tokyo, Japan, in 2006; Marine Micro Technology, Saitama, Japan, in 2007), for the acoustic observations. An A-tag is an event data logger that records sound pressure and the difference in time arrival between two hydrophones. It does not record the waveforms of received sound.

An A-tag consists of a stereo hydrophone, preamplifier with bandpass filter, CPU (PIC18F6620), flash memory (128 Mbytes), and lithium battery cell (CR2). The hydrophones had a sensitivity of MHP-140 (Marine Micro Technology) −201 dB (1 V/μPa) and a resonant frequency of 130 kHz, similar to the dominant frequency of finless porpoise sonar signals. This setting reduced noise outside the sensitive band of the hydrophone at sound reception. Hydrophone sensitivity was calibrated using an acoustical measurement tank (10 m in width, 15 m in length, and 10 m in depth) at the Fisheries Research Agency in Ibaraki, Japan. The ultrasonic sound transmission system used in calibration consisted of a function generator (NF1930A, NF Corp., Tokyo, Japan) and a transducer (B&K8103, Brüel & Kjaer, Naerum, Denmark); the system generated a 10-cycle tone burst for any frequency. A-tags were also confirmed to be able to record sounds made by free-ranging porpoises in an *ex situ* oxbow of the Yangtze River (Akamatsu *et al.*, 2005a).

Each A-tag had two hydrophones, approximately 170 mm apart, which were used to identify the sound source direction. Electronic bandpass filters at the preamplification stage were adjusted to 70–300 kHz (in 2006) or 55–235 kHz (in 2007) to match the frequency band of Yangtze finless porpoise sonar signals, which ranges from 87 to 145 kHz and averages at 125 ± 6.92 kHz (Li *et al.*, 2005a). The acoustic data logger recorded sound pressure at the primary and secondary hydrophones, as well as the difference in sound arrival times between the two hydrophones, every 0.5 ms (2 kHz event sampling frequency). The three data sets and the absolute time were recorded automatically only when the received sound pressure was greater than the trigger level of the primary hydrophone. Otherwise, no data were stored to conserve memory capacity. An A-tag can record information up to 30–40 h, depending on the number of pulses stored.

Peak-to-peak source levels for this species were 163.7–185.6 dB, referred to 1 μPa (Li *et al.*, 2006), and the sound pressure level off the beam axis at 90° reached a maximum of 162 dB (Akamatsu *et al.*, 2005b). Transmitted sound pressure levels can be highly variable, but off-axis signals still reached significant levels during this study, and it was possible to observe them using the data logger. We set the detection threshold level of the data logger at 135.3 dB. Our calibration experiment revealed that each A-tag had a slightly different threshold level, but the threshold level of 135.3 dB was higher than any one A-tag threshold. Our offline analysis used recorded pulses greater than 135.3 dB (5.85 Pa). We allowed a maximum of 50.3 dB propagation loss for detecting signals. Assuming a simple spherical propagation model based on the freshwater values set out by Fisher and Simmons (1977) (absorption coefficient of

0.004 dB/m at 125 kHz), the maximum detection distance of the stereo acoustic data logger was approximately 290 m.

The bearing angle of a sound was calculated using the difference in time arrival between the two hydrophones. The triggering time of both hydrophones was monitored every 271 ns, which was sufficiently fast to determine the onset of a pulse wave. Sounds traveled 0.4 mm in 271 ns, while the baseline (the separation between the two hydrophones) was 170 mm. Therefore, even this short baseline system allowed a fair bearing angle resolution. Signal processing and structure are described in more detail in Akamatsu *et al.* (2005a).

C. Acoustic observations

We used a bamboo rod to fix the acoustic data logger at a 1-m depth from the side of each anchored boat. In 2006, we fixed two A-tags underwater from boats at Stations 1 and 2, and we fixed an additional A-tag at Station 0 in 2007. The stations were spaced more than 1000 m apart, well outside the A-tag detection range of 290 m. This design ensured that the observations at each station were independent. We assumed no simultaneous detection of individual animals. The two hydrophones of each A-tag were set parallel to the current direction to monitor the direction of porpoise movement between the river and the lake. The primary hydrophone of the A-tag was directed upstream of the site (Poyang Lake side), and the secondary hydrophone was directed downstream (the Yangtze River side).

D. Visual observations

During acoustic observations, we conducted simultaneous visual observations from the same anchored boat; four observers each covered a 90° arc from the boats. Observers watched for 1 h and rested for 30 min; eye height was approximately 2 m above the water surface. When porpoises were sighted, the observer recorded the minimum group size, the direction of movement (upstream or downstream), and the distance and bearing angle from the bow of the survey boat. These parameters were the same as those measured by the stereo acoustic data logger, with the exception of distance. To ensure that these results could be compared with those obtained by acoustic detection, we only recorded visual observation data detected within 300 m, similar to the acoustical detection range. The minimum group size was defined as the number of the animals that respired successively within a few seconds because this species has an average shallow dive time of 4.86 ± 4.72 s (Akamatsu *et al.*, 2002). For the purposes of analysis, groups separated by more than 1 min were considered to be different sightings because this species has an average deep dive time of 70.9 ± 22.9 s (Akamatsu *et al.*, 2002). Currents and winds affected the direction of the observation boat; this parameter was identical with the direction of the data logger. The direction of the boat's bow was used as a reference to synchronize data collected through acoustic and visual observations.

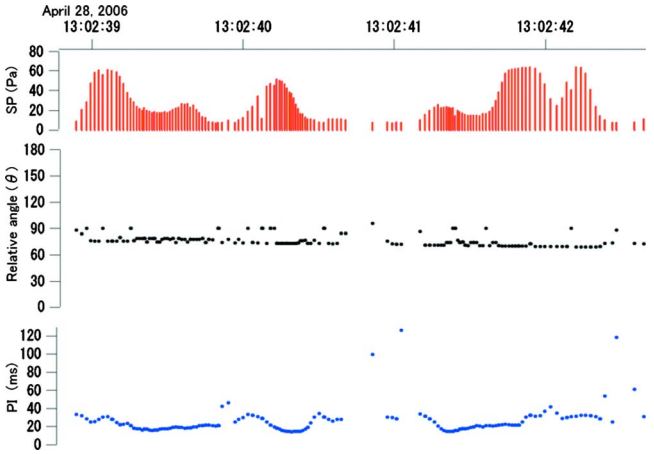


FIG. 2. (Color online) Example of time series data of porpoise sounds recorded by the stereo acoustic data logger. The vertical axes show the received sound pressure (SP), relative bearing angle to the porpoise, and interclick interval (ICI). In the center graph, 180° indicates upstream from the survey boat and 0° means downstream. The porpoise ICIs and sound pressure levels changed smoothly (Akamatsu *et al.*, 2005c).

E. Acoustic signal processing

We eliminated contamination from noise and reflection and calculated the interclick intervals and relative angles of sound sources using a custom-made program developed using IGOR PRO 5.03 (WaveMetrics, Lake Oswego, OR). Relative angles to the sound source were calculated using the difference in time arrival between the two hydrophones.

Sample data shown in Fig. 2 illustrate sound pressure, relative bearing angle, and the interclick interval of porpoise clicks. We were able to track porpoises easily because they phonated frequently. As shown in the figure, interclick intervals and sound pressure levels changed smoothly (Akamatsu *et al.*, 2005c), while background or boat noise caused randomly changing patterns in the interclick interval and sound pressure. We were unable to use frequency information to exclude noise because A-tags do not record waveform. Instead, we used interclick intervals to discriminate signals from noise. We excluded any successive clicks greater than twice or less than half the previous interclick intervals (Akamatsu *et al.*, 1998, 2001).

The multipath propagation in the Yangtze River can cause echolocation signals to have a multipulse structure (Li *et al.*, 2005b). In this shallow freshwater system, reflected signals came just after the direct path signal. Because the animals had a very shallow depth, the surface reflection had an angle similar to the direct path signal; this resulted in the echo's very short delay time. Pulses within 2 ms after a direct path pulse were eliminated during offline signal processing. Since the mean minimum lag time to process returning echoes inside an animal brain is 2.5 ms (Au, 1993), porpoises' sound is considered to be not excluded in this processing.

F. Number and movement direction of animals

The number and movement direction of animals were determined manually from click trains. This species usually produces an interclick interval shorter than 130 ms (Li *et al.*,

2005a; Akamatsu *et al.*, 2007), as shown in Fig. 2. We defined a click train as a series of clicks in which intervals were 130 ms or shorter. We considered some click trains 10 s or less apart and within a similar bearing angle to have been produced by the same individual; these were defined as a single track. The number of independent traces of sound source bearing angle in a 1-min time bin was defined as the observed number of animals in a unit time (or group size). When click trains were separated by intervals longer than 10 s, determining the number of porpoises within a detection area was not possible; one or more porpoises may have phonated. In these cases, we underestimated the number of porpoises. This was used as a conservative criterion to avoid double-counting of animals.

Simultaneous phonation of two individuals swimming close together could be identified through the double different cyclic characteristics of the sound pressure and/or interclick intervals within a single trace, so we counted these as originating from two porpoises. This phenomenon was relatively easy to discriminate from reflections because reflection sound always involves a separation time after the direct path click.

In contrast, when single periodicity in interclick intervals and/or a smoothly changing sound pressure accompanied close parallel traces, we counted only one porpoise. These parallel traces were caused by an error in the trigger point among multiple wavelengths in a click. The trigger point of primary and secondary hydrophones could differ when the sound pressure at the onset of a click is comparable to the detection threshold level. Among finless porpoises, click amplitude rises gradually. Therefore, the second wave highlight next to the first onset wave tends to be triggered by the secondary hydrophone, even if the first onset was triggered by the primary hydrophone. One wavelength ambiguity of the trigger point occurred, resulting in close parallel traces of a single phonating animal.

Animals were counted visually through observation around the survey boat during the same time bin as they were counted using the acoustic method. If porpoises passed near the observation station, they were likely to be observed once within a 1-min time bin, which is close to the average respiration interval of an adult finless porpoise engaged in deep diving activity.

We also compared the visually and acoustically measured movement directions of animals. In the acoustic method, direction was determined by changes in the bearing angles of received sounds. A change in bearing angle from positive to negative indicated that the porpoise moved from the lake side to the river side, and vice versa. The difference in time arrival between the two hydrophones, correlated with the bearing angles, had a minimum resolution 13.6 μs. When the difference in time arrival was considerably greater than 13.6 μs, the swimming direction was determined to be either upstream toward Poyang Lake or downstream toward the Yangtze River. Otherwise, we did not record a swimming direction. When the trace consisted of only one click train, determining the swimming direction was impossible. When the primary hydrophone of the data logger was triggered but the secondary hydrophone received an insufficient sound level, the difference in time arrival was 0 and indicated as a line at 90°, as shown in Fig. 3. We did not use these measurements to count individuals but used them instead to identify simultaneous phonation of multiple individuals.

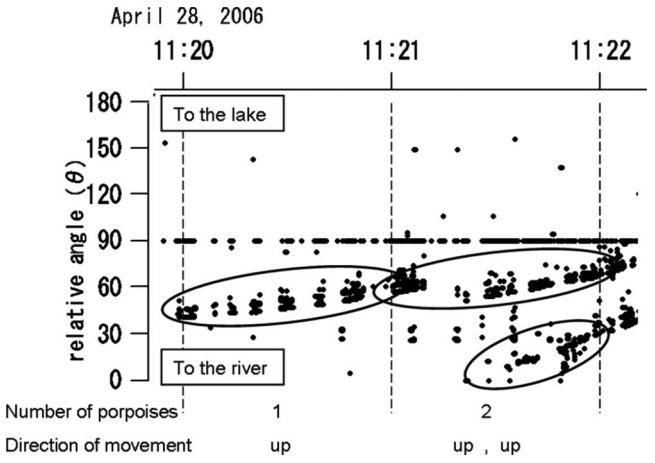


FIG. 3. Number of porpoises counted acoustically in each time bin. This example shows two porpoises. One porpoise continued phonating and swam upstream from the river to the lake. The second porpoise appeared in the observable area at 11:21. The number of animals and movement direction were counted as one upstream at 11:20 and two upstream at 11:21. Within the one track, click trains were not separated by more than 10 s.

III. RESULTS

We obtained 1216 min of effective visually and acoustically measured data at Station 1 and 504 min at Station 2 in 2006, and 464 min at Station 0 in 2007, for an overall total of 2184 min of observations.

In total, 2987 and 591 animals were detected acoustically and visually. At Stations 0, 1, and 2, respectively, animals were detected acoustically in 92.9%, 76.2%, and 76.0% of all time bins, whereas animals were detected visually in 23.5%, 13.1%, and 1.6% of all time bins. On average, porpoises were detected acoustically in 81.7% ± 9.7% and visually in 12.7% ± 11.0% of all observation times; the acoustic detection rate during the total observation time was significantly higher than the visual detection rate (Scheffe's test, $P < 0.01$). As shown in Fig. 4, both methods detected the most porpoises at Station 0 and the least at Station 2 (Scheffe's test, $P < 0.01$). Detection rates differed among observation sites.

We monitored the swimming direction of porpoises using bearing angles and compared the results of acoustic observations of swimming direction with visual observations. At each station, the number of positive swimming direction identifications divided by the total observation time was similar for both observation methods (Fig. 5).

Figure 6 shows the number and size of detected groups by time bin. Note that the ordinate is logarithmic. Over the total observation time (2184 min), both methods determined the same group size for only 458 min. The most detected numbers of animals were zero by the visual method (1881 min) and one by the acoustic one (866 min). The

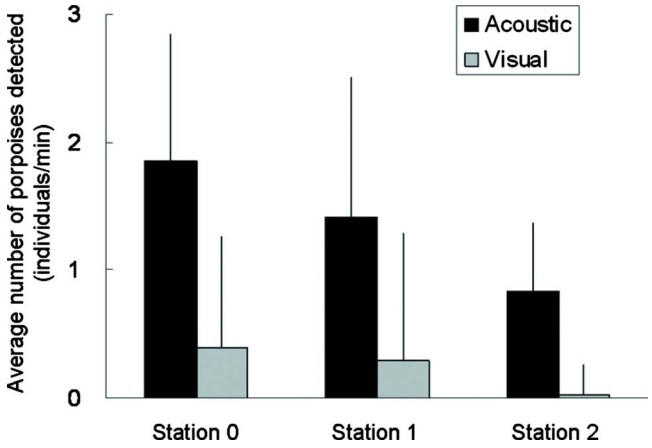


FIG. 4. (Color online) Number of porpoises detected in 1 min at the three stations. The greatest number of porpoises was detected at Station 0 and the fewest at Station 2 (Scheffé's test, $P<0.01$).

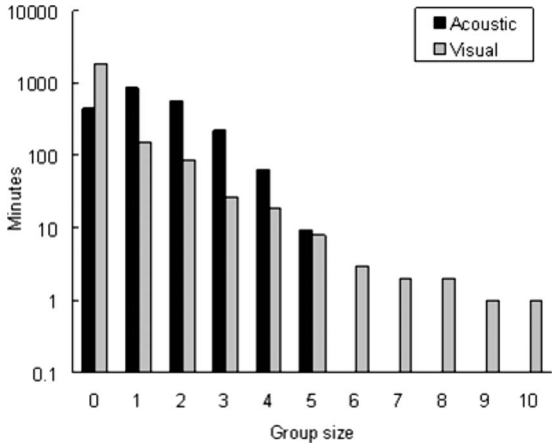


FIG. 6. Comparison of group sizes as determined acoustically and visually. The total of all observations comprised 2184 min. The acoustic data logger could count groups up to five, whereas visual observation could count up to ten individuals at a time. Acoustic detection was greater than visual detection for groups sized 1–5.

acoustic data logger could count group sizes to a maximum of five individuals, whereas visual observation could count group sizes to a maximum of ten individuals.

IV. DISCUSSION

Stationary acoustic monitoring was effective for counting Yangtze finless porpoises that were echolocating; this method yielded a detection rate seven times higher than visual observation (Fig. 4). The results clearly show that the acoustic method was more effective at detecting the presence of animals than the visual method.

We were able to detect porpoises frequently using the acoustic method, while only occasionally using visual observation. This was a result of different visual and acoustic cues from porpoises. Porpoises can be recorded acoustically when they produce sonar phonates within a detection range. Yangtze finless porpoises produce sonar click trains every 5–6 s on average (Akamatsu *et al.*, 2005c, 2007), so their frequent phonation resulted in a high detection rate using the acoustic method. In contrast, we were only able to observe the porpoises visually when they surfaced in the turbid water of the Yangtze River. Among adult Yangtze finless porpoises, long dive duration averages 70.9 s (Akamatsu *et al.*, 2002). The acoustic and visual detection methods used cues of differing intervals. In addition, because finless porpoises are small cetaceans, do not have dorsal fins, and exhibit little aerial behavior, they are easily overlooked even when they are near the surface.

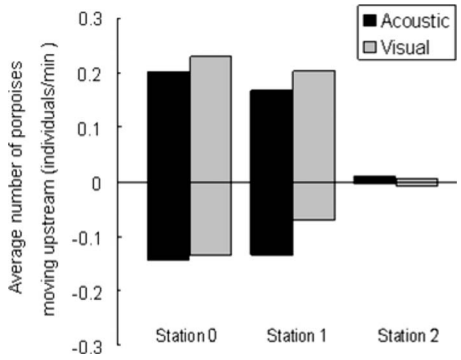


FIG. 5. Average number of moving porpoises per time bin at each station. Positive numbers indicate movement upstream toward Poyang Lake. Negative bars indicate movement downstream toward the Yangtze River. Both methods showed similar trends at all stations.

The stereo acoustic data logger systems revealed high and low porpoise density areas, which agreed with visual observations (Fig. 4). The average number of detected porpoises was highest at Station 0, which was near the confluence of the lake and the river. This finding is consistent with previous research indicating that Yangtze finless porpoises tend to aggregate in that area (Wei *et al.*, 2003).

Acoustic observations were used successfully to detect the movement direction of porpoises underwater (Fig. 5). The stereo system was more powerful than a monaural system because it could separate sound sources to count the number of animals and also identify their swimming direction. This feature is most suitable for long-term monitoring of porpoise migration by using several A-tags.

The number of time bins in which no porpoises were detected acoustically (447 min) was about one-quarter the number in which none were detected visually (1881 min). Therefore, the acoustic method missed fewer porpoises than the visual method. A towed acoustic survey also resulted in a large ratio of individual animals that were missed (Akamatsu *et al.*, 2008). The acoustic method could detect group sizes to a maximum of five individuals but tended to underestimate the size of larger groups (Fig. 6). When the group size was fewer than four individuals, the acoustic method detected porpoises in 1728 1-min bins, approximately four times the number of bins in which porpoises were detected visually (286 min).

The inability of the acoustic method to detect more than five individuals in a 1-min bin was probably due to the limited resolution of the stereo acoustic data logger's bearing angle. The short distance between the two hydrophones (170 mm) was a possible cause for this limitation. A longer baseline should improve the bearing angle resolution. Other

possible causes may have been the small source level of porpoises (including off-axis sounds), alternate phonations of multiple individuals near each other, or eavesdropping to maintain silence. In addition, previous research has suggested that large groups may not vocalize as much as small groups (Götz *et al.*, 2006). Several porpoises swimming together within the detection range phonate alternately, but our passive acoustic system was unable to differentiate them. Yangtze finless porpoises, however, often swim alone or in very small groups and in areas in which visual observations may not be practical, as indicated in Fig. 6. The results indicate that our acoustic monitoring system would have a limited application to species that form larger groups. We used a conservative criterion when counting the number of porpoises to avoid double-counting.

In conclusion, the stationary acoustic monitoring system using stereo acoustic data loggers performed more efficiently than the visual method, especially in areas of low-density echolocating animals. Stationary acoustic observation is suitable for use in areas in which porpoises appear infrequently and form small groups, where visual observation may not be practical. The acoustic system appears to be powerful at monitoring porpoises in a narrow channel such as a river system. In future research, we will monitor porpoise migration using multiple acoustic monitoring systems.

ACKNOWLEDGMENTS

We thank all members of the Laboratory of Conservation Biology of Aquatic Animals at the Institute of Hydrobiology, Chinese Academy of Sciences, the National Research Institute of Fisheries Engineering, the Japan Fisheries Resource Conservation Association, Biosphere Informatics, Graduate School of Informatics, Kyoto University, and the Fisheries Administration Bureau of Hukou Jiangxi Province. This study was partly supported by a Grant-in-aid for Scientific Research (B) (19405005) from the Japanese Research and Development Program for New Bio-Industry Initiatives, the National Natural Science Foundation of China (30730018 and 30570233), the Chinese Academy of Sciences (Present Fund), the Ocean Park Conservation Foundation of Hong Kong, and the Sasagawa Scientific Research Grant from The Japan Science Society (19-740M).

Akamatsu, T., Matsuda, A., Suzuki, S., Wang, D., Wang, K., Suzuki, M., Muramoto, H., Sugiyama, N., and Oota, K. (**2005a**). "New stereo acoustic data logger for tagging on free-ranging dolphins and porpoises," Mar. Technol. Soc. J. **39**, 3–9.

Akamatsu, T., Teilmann, J., Miller, L. A., Tougaard, J., Dietz, R., Wang, D., Wang, K., Siebert, U., and Naito, Y. (**2007**). "Comparison of echolocation behaviour between coastal and riverine porpoises," Deep-Sea Res., Part II **54**, 290–297.

Akamatsu, T., Wang, D., Nakamura, K., and Wang, K. (**1998**). "Echolocation range of captive and free-ranging baiji (*Lipotes vexillifer*), finless porpoise (*Neophocaena phocaenoides*), and bottlenose dolphin (*Tursiops truncates*)," J. Acoust. Soc. Am. **104**, 2511–2516.

Akamatsu, T., Wang, D., and Wang, K. (**2005b**). "Off-axis sonar beam pattern of free-ranging finless porpoises measured by a stereo pulse event data logger," J. Acoust. Soc. Am. **117**, 3325–3330.

Akamatsu, T., Wang, D., Wang, K., Li, S., Zhao, X., Barlow, J., Stewart, B. S., and Richlen, M. (**2008**). "Estimation of the detection probability for Yangtze finless porpoises (*Neophocaena phocaenoides asiaeorientalis*) with a passive acoustic method," J. Acoust. Soc. Am. **123**, 4403–4411.

Akamatsu, T., Wang, D., Wang, K., and Naito, Y. (**2005c**). "Biosonar behavior of free-ranging porpoises," Proc. R. Soc. London, Ser. B **272**, 797–801.

Akamatsu, T., Wang, D., Wang, K., and Wei, Z. (**2001**). "Comparison between visual and passive acoustic detection of finless porpoises in the Yangtze River, China," J. Acoust. Soc. Am. **109**, 1723–1727.

Akamatsu, T., Wang, D., Wang, K., Wei, Z., Zhao, Q., and Naito, Y. (**2002**). "Diving behavior of freshwater finless porpoises (*Neophocaena phocaenoides*) in an oxbow of the Yangtze River, China," ICES J. Mar. Sci. **59**, 438–443.

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

Au, W. W. L., and Benoit-Bird, K. J. (**2003**). "Automatic gain control in the echolocation system of dolphins," Nature (London) **423**, 861–863.

Carstensen, J., Henriksen, O. D., and Teilmann, J. (**2006**). "Impacts of offshore wind farm construction on harbour porpoises: Acoustic monitoring of echolocation activity using porpoise detectors (T-PODs)," Mar. Ecol.: Prog. Ser. **321**, 295–308.

Fisher, F. H., and Simmons, V. P. (**1977**). "Sound absorption in sea water," J. Acoust. Soc. Am. **62**, 558–564.

Götz, T., Verfuß, U. K., and Schnitzler, H. U. (**2006**). "'Eavesdropping' in wild rough-toothed dolphins (*Steno bredanensis*)?," Biol. Lett. **2**, 5–7.

Ichikawa, K., Tsutsumi, C., Arai, N., Akamatsu, T., Shinke, T., Hara, T., and Adulyanukosol, K. (**2006**). "Dugong (*Dugong dugon*) vocalization patterns recorded by automatic underwater sound monitoring systems," J. Acoust. Soc. Am. **119**, 3726–3733.

Janik, V. M. (**2000**). "Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*)," Science **289**, 1355–1357.

Janik, V. M., van Parijs, S. M., and Thompson, P. M. (**2000**). "A two-dimensional acoustic localization system for marine mammals," Marine Mammal Sci. **16**, 437–447.

Leeney, H. R., Berrow, S., MacGrath, D., O'Brien, J., Cosgrove, R., and Godley, J. B. (**2007**). "Effects of pingers on the behavior of bottlenose dolphins," J. Mar. Biol. Assoc. U.K. **87**, 129–133.

Li, S., Wang, K., Wang, D., and Akamatsu, T. (**2005a**). "Echolocation signals of the free-ranging Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*)," J. Acoust. Soc. Am. **117**, 3288–3296.

Li, S., Wang, K., Wang, D., and Akamatsu, T. (**2005b**). "Origin of the double- and multi-pulse structure of echolocation signals in Yangtze finless porpoises (*Neophocaena phocaenoides asiaeorientalis*)," J. Acoust. Soc. Am. **118**, 3934–3940.

Li, S., Wang, D., Wang, K., and Akamatsu, T. (**2006**). "Sonar gain control in echolocating finless porpoises (*Neophocaena phocaenoides*) in an open water (L)," J. Acoust. Soc. Am. **120**, 1803–1806.

Nishimura, C. E., and Conlon, D. M. (**1994**). "IUSS dual use: Monitoring whales and earthquakes using SOSUS," Mar. Technol. Soc. J. **27**, 13–21.

Philpott, E., Englund, A., Ingram, S., and Rogan, E. (**2007**). "Using T-PODs to investigate the echolocation of coastal bottlenose dolphins," J. Mar. Biol. Assoc. U.K. **87**, 11–17.

Thomsen, F., van Elk, N., Brock, V., and Piper, W. (**2005**). "On the performance of automated porpoise-click-detectors in experiments with captive harbor porpoises (*Phocoena phocoena*) (L)," J. Acoust. Soc. Am. **118**, 37–40.

Tsutsumi, C., Ichikawa, K., Arai, N., Akamatsu, T., Shinke, T., Hara, T., and Adulyanukosal, K. (**2006**). "Feeding behavior of wild dugongs monitored by a passive acoustical method," J. Acoust. Soc. Am. **120**, 1356–1360.

van Parijs, S. M., Smith, J., and Corkeron, P. J. (**2002**). "Using calls to estimate the abundance of inshore dolphins: A case study with Pacific humpback dolphins *Sousa chinensis*," J. Appl. Ecol. **39**, 853–864.

Verfuß, U. K., Honnef, C. G., Meding, A., Dähne, M., Mundry, R., and Benke, H. (**2007**). "Geographical and seasonal variation of harbour porpoise (*Phocoena phocoena*) presence in the German Baltic Sea revealed by passive acoustic monitoring," J. Mar. Biol. Assoc. U.K. **87**, 165–176.

Wang, K., Wang, D., Akamatsu, T., Li, S., and Xiao, J. (**2005**). "A passive acoustic monitoring method applied to observation and group size estimation of finless porpoises," J. Acoust. Soc. Am. **118**, 1180–1185.

Wei, Z., Wang, D., Zhang, X., Zhao, Q., Wang, K., and Kuang, X. (**2002**). "Population size, behavior, movement pattern and protection of Yangtze finless porpoise at Balijiang section of the Yangtze River," Resources and Environment in the Yangtze Basin **11**, 427–432.

Wei, Z., Zhang, X., Wang, K., Zhao, Q., Kuang, X., Wang, X., and Wang, D. (**2003**). "Habitat use and preliminary evaluation of the habitat status of the Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*) in the Balijiang section of the Yangtze River, China," Acta Zoologica Sinica

49, 163–170.
Zhang, X., Liu, R., Zhao, Q., Zhang, G., Wei, Z., Wang, X., and Yang, J. (**1993**). "The population of finless porpoise in the middle and lower reaches of Yangtze River," Acta Theriologica Sinica **16**, 490–496.
Zhao, X., Barlow, J., Taylor, B. L., Pitman, L. R., Wang, K., Wei, Z., Brent, S., Stewart, S. B., Turvey, T. S., Akamatsu, T., Reeves, R. R., and Wang, D. (**2008**). "Abundance and conservation status of the Yangtze finless porpoise in the Yangtze River, China," Biol. Conserv. **141**, 3006–3018.
Zheng, J., Xia, J., He, S., and Wang, D. (**2005**). "Population genetic structure of the Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*): Implications for management and conservation," Biochem. Genet. **43**, 307–320.

Behavioral responses of herring (*Clupea harengus*) to 1–2 and 6–7 kHz sonar signals and killer whale feeding sounds

Lise Doksæter,[a] Olav Rune Godø, and Nils Olav Handegard
Institute of Marine Research (IMR), NO-5817 Bergen, Norway

Petter H. Kvadsheim
Maritime Systems Division, Norwegian Defence Research Establishment (FFI), NO-3191 Horten, Norway

Frans-Peter A. Lam
Observation Systems, TNO, Defence, Security and Safety, The Hague, The Netherlands

Carl Donovan
The Centre for Research into Environmental and Evolutionary Modelling (CREEM), University of St. Andrews, St. Andrews, Scotland

Patrick J. O. Miller
Sea Mammal Research Unit (SMRU),University of St. Andrews, St. Andrews, Scotland

(Received 30 April 2008; revised 8 September 2008; accepted 13 October 2008)

Military antisubmarine sonars produce intense sounds within the hearing range of most clupeid fish. The behavioral reactions of overwintering herring (*Clupea harengus*) to sonar signals of two different frequency ranges (1–2 and 6–7 kHz), and to playback of killer whale feeding sounds, were tested in controlled exposure experiments in Vestfjorden, Norway, November 2006. The behavior of free ranging herring was monitored by two upward-looking echosounders. A vessel towing an operational naval sonar source approached and passed over one of them in a block design setup. No significant escape reactions, either vertically or horizontally, were detected in response to sonar transmissions. Killer whale feeding sounds induced vertical and horizontal movements of herring. The results indicate that neither transmission of 1–2 kHz nor 6–7 kHz have significant negative influence on herring on the received sound pressure level tested (127–197 and 139–209 dB_{rms} re 1 μPa, respectively). Military sonars of such frequencies and source levels may thus be operated in areas of overwintering herring without substantially affecting herring behavior or herring fishery. The avoidance during playback of killer whale sounds demonstrates the nature of an avoidance reaction and the ability of the experimental design to reveal it.
 [DOI: 10.1121/1.3021301]

PACS number(s): 43.80.Nd, 43.50.Rq, 43.50.Sr [WWA] Pages: 554–564

I. INTRODUCTION

The interest in how human generated sound might affect marine organisms has grown considerably over the past decade (Richardson *et al.*, 1995). The main focus has been on marine mammals, although other aquatic animals such as fish and some invertebrates may also be affected (Hofman, 2004; Popper *et al.*, 2004). Fish have sensitive hearing organs (Ladich and Popper, 2004) and use sound for communication and to perceive their acoustic environment (Fay and Popper, 2000; Popper, 2003). A wide range of anthropogenic sound sources are present in the marine environment, the most intense being vessel traffic, seismic airguns, pile driving, and military sonars (Hofman, 2004; Popper *et al.*, 2004; Hastings and Popper, 2005). One of the earliest studies to demonstrate the possibility of hearing injury in fish was Enger (1981), who showed that sensory cells in the ears of cod (*Gadus morhua*) were damaged when exposed to high-intensity sounds. Similar effects have also been demonstrated in cichlids (Hastings *et al.*, 1996), snappers (McCauley *et al.*, 2003), and clupeids (Denton and Gray, 1993). Exposure to intense sound may also lead to temporary loss of hearing [temporal threshold shifts (TTSs)] and has been documented for various species of fish in response to seismic shooting (Popper *et al.*, 2005), military sonars (Popper *et al.*, 2007) and simulated white noise (Smith *et al.*, 2004). However, experiments on physical damage and TTS need to be conducted in enclosed environments, with no possibility for the fish to escape from or avoid the sound. Avoidance reactions in the wild have been observed in response to vessel noise (Olsen *et al.*, 1983; Vabø *et al.*, 2002) and seismic shooting (Engås *et al.*, 1996; Engås and Løkkeborg, 2002). Intense sound may also lead to physiological stress (Smith *et al.*, 2004) or prevent fish from hearing biologically relevant sounds (masking) (Popper, 2003).

How anthropogenic sound affects fish will depend on the species, as hearing thresholds among fish are highly variable. Most teleosts are only able to detect frequencies below 500 Hz, called "hearing generalists" (e.g., Chapman and

[a]Author to whom correspondence should be addressed. Electronic mail: lise.doksaeter@imr.no

Hawkins, 1973; Hawkins and Johnstone, 1978; Mann *et al.*, 1998). Others, the "hearing specialists," are sensitive to sounds over a wider frequency range (e.g., Popper, 1972; Kenyon *et al.*, 1998). Herring (*Clupea harengus*) is such a hearing specialist and is able to detect frequencies up to at least 4000 Hz (Enger, 1967; Mann *et al.*, 2005) due to a gas filled channel that connects the swimbladder to the otolith organs (Blaxter *et al.*, 1979; Denton and Gray, 1979; Popper *et al.*, 2004).

Norwegian Spring Spawning (NSS) herring is by far the largest herring stock in the northeast Atlantic. It is an important stock both in terms of fisheries and as prey for many other species (Hamre, 1990; Holst *et al.*, 2004; Røttingen and Slotte, 2001). A negative anthropogenic impact would thus potentially have large consequences for the fishery as well as the ecosystem. The annual distribution of NSS herring is divided into three main parts separated by more or less well defined migrations (Holst *et al.*, 2002); spawning along the Norwegian coast in February–March (Johannessen *et al.*, 1995; Røttingen and Slotte, 2001), feeding in the Norwegian Sea in April–September (Holst *et al.*, 2004), and overwintering in October–January (Dragesund *et al.*, 1997). Since the mid-1980s, almost the entire stock has been overwintering in Vestfjorden, Northern Norway (Dragesund *et al.*, 1997). This area has also frequently been used for military antisubmarine warfare exercises, which have involved use of active sonars transmitting at 5–8 kHz. Modern long-range active sonar also covers a frequency band below 2 kHz. Even though signals above 5 kHz would hardly be audible to NSS herring, the lower frequency bands of these sonars are well within their hearing range (Enger, 1967). Behavioral effects of sonars on marine organisms have been suggested (Hofman, 2004), but very few studies have been carried out on their effects on fish. An examination and a quantification of herring behavior in response to military sonars are therefore of high importance to establish environmentally safe sonar operation procedures in areas of high herring density.

Killer whales (*Orcinus orca*) prey on herring during the overwintering period (Similä and Ugarte, 1993; Similä, 1997; Nøttestad, 1998). Feeding killer whales use communication calls which could resemble the sonar signals tested in this study in both frequency and frequency modulation (Stranger, 1995; Van Opzeeland *et al.*, 2005; Miller, 2006). This similarity could potentially cause confusion in herring between sonar pings and killer whale calls and thus induce an antipredator response during sonar exposure.

The objectives of this study were to investigate whether sonar transmission of two different frequency bands; 1–2 kHz (F1) and 6–7 kHz (F2) elicited any behavioral responses in NSS herring. Controlled exposure experiments were conducted on herring in Vestfjorden in November 2006 using a sonar source representative of an operational naval sonar system. In addition, as a control experiment, herring were exposed to playbacks of sounds of herring-feeding killer whales. Herring behavior was monitored by two upward-looking, bottom-mounted echosounders, in an experimental design similar to a vessel avoidance experiment on herring (Ona *et al.*, 2007).

II. METHODS

Controlled exposure experiments were conducted between 12 and 30 November 2006, onboard the research vessel R/V H. U. Sverdrup II (Kvadsheim *et al.*, 2007).

A. Experimental design

The behavioral response of herring to sonar signals of two different frequency bands were tested: 1–2 kHz (F1) and 6–7 kHz (F2). Herring behavior was monitored acoustically by a system of two upward-looking, bottom-mounted echosounders (Simrad EK 60, Kongsberg Maritime AS, Horten, Norway) placed 400 m apart in a small fjord inside Vestfjorden (Patel, 2007). The echosounders transmitted a narrow beamed 38 kHz signal at a ping frequency of 1 Hz. The northern echosounder (A) was placed at a depth of approximately 400 m, while the southern (B) is at approximately 500 m. The system was connected onshore by an underwater data transmission and power cable. Data were collected in a cabin onshore by a PC running EK60 software, and echograms were continuously steamed to the internet, enabling real-time monitoring onboard the vessel during the experiments. The vessel passed directly above one of the two echosounders (later referred to as "passed echosounder"), while towing the sonar source, transmitting either F1 signals, F2 signals, or no signal (silent control). The echosounder (A or B) having the higher herring density was passed by the vessel in a straight line. Transmission started 1 nmi (nautical mile) from the position of the echosounder, and continued 1 nmi beyond it (referred to as one "passage"). The exact positions of the echosounders are given as their latitude and longitude, and GPS was used to ensure direct passage. During the experiment, the entire ship was darkened in order to prevent any light stimuli from affecting herring behavior. The 38 kHz echosounder of the ship was set in passive mode to avoid interference with the bottom-mounted echosounders. The vessel kept a constant speed (~8 kn) during the run. Killer whale feeding sounds were presented by lowering an underwater speaker (Lubell Labs model LL916, Columbus OH, www.lubell.com) to a depth of 27 m from a small boat, while the vessel made a silent control passage. The sounds played to the herring were monitored by hydrophone to assure that sounds were faithfully played back by the system and that the sounds were not in any way distorted. During passage, the small boat passed within a distance of 5–10 m of the source ship.

The experiment was conducted in a block design. Each block consisted of three passages of the echosounder, with each passage transmitting either F1 signals, F2 signals, or no transmission (silent control) (Table I). When killer whale feeding sounds were played, this stimulus replaced F2 in the block. The order of the different transmission types was randomized to distinguish between presentation order and sonar frequency. One experiment consisted of three blocks, with 1 h between each block. Experiments were conducted at different times of the day in order to separate exposure effects from natural day/night variations associated with diel vertical migration of herring (described in Huse and Korneliussen, 2000). Sound speed profiles through the water column were

TABLE I. Controlled exposure experiments carried out with herring. Experiments 1, 2, 3, and 6 consisted of three blocks each, and herring were exposed to F1 (1–2 kHz) and F2 (6–7 kHz) frequency sonar signals as well as a control run without transmission. Experiments 4 and 5 consisted of one block each, consisting of playback of killer whale feeding sounds (Orca), F1, and a control run. The order of transmission types within each block was randomized.

Experiment	Block	Date	Start time (UTC)	Stop time (UTC)	Transmission order	Echosounder passed
1	1	Nov. 12, 2006	21:05:00	22:16:23	F1-F2-control	A
1	2	Nov. 12, 2006	22:59:29	0:45:04	F2-control-F1	A
1	3	Nov. 13, 2006	1:24:31	3:07:50	control-F2-F1	A
2	1	Nov. 16, 2006	22:40:38	23:54:50	F2-F1-control	B
2	2	Nov. 17, 2006	0:46:31	2:02:04	control-F2-F1	B
2	3	Nov. 17, 2006	2:42:46	3:53:48	F1-control-F2	B
3	1	Nov. 18, 2006	13:21:54	14:34:40	F2-F1-control	A
3	2	Nov. 18, 2006	14:44:05	15:55:48	control-F2-F1	B
3	3	Nov. 18, 2006	16:44:50	17:50:07	F1-F2-control	A
4	1	Nov. 22, 2006	18:21:51	19:32:24	control-Orca-f1	B
5	1	Nov. 25, 2006	18:50:09	19:57:55	F1-control-Orca	B
6	1	Nov. 29, 2006	16:18:15	17:38:19	F1-control-F2	B
6	2	Nov. 29, 2006	18:28:14	17:38:19	F2-control-F1	B
6	3	Nov. 29, 2006	20:32:14	21:43:21	control-F1-F2	B

recorded after each experiment using an STD/CTD (model SD204, SAIV AS, Bergen, Norway). The profiles and sonar source specifications were used as input into an acoustic model (LYBIN, Royal Norwegian Navy and FFI) to estimate received sound pressure levels at the observation point of the echosounders during the experiments (Fig. 1).

B. Sonar source

Sonar signals were transmitted using a multipurpose towed acoustic source (Socrates, TNO-Defence, Security and Safety, The Hague, NL), a military experimental sonar currently used for the sonar research carried out for the Royal

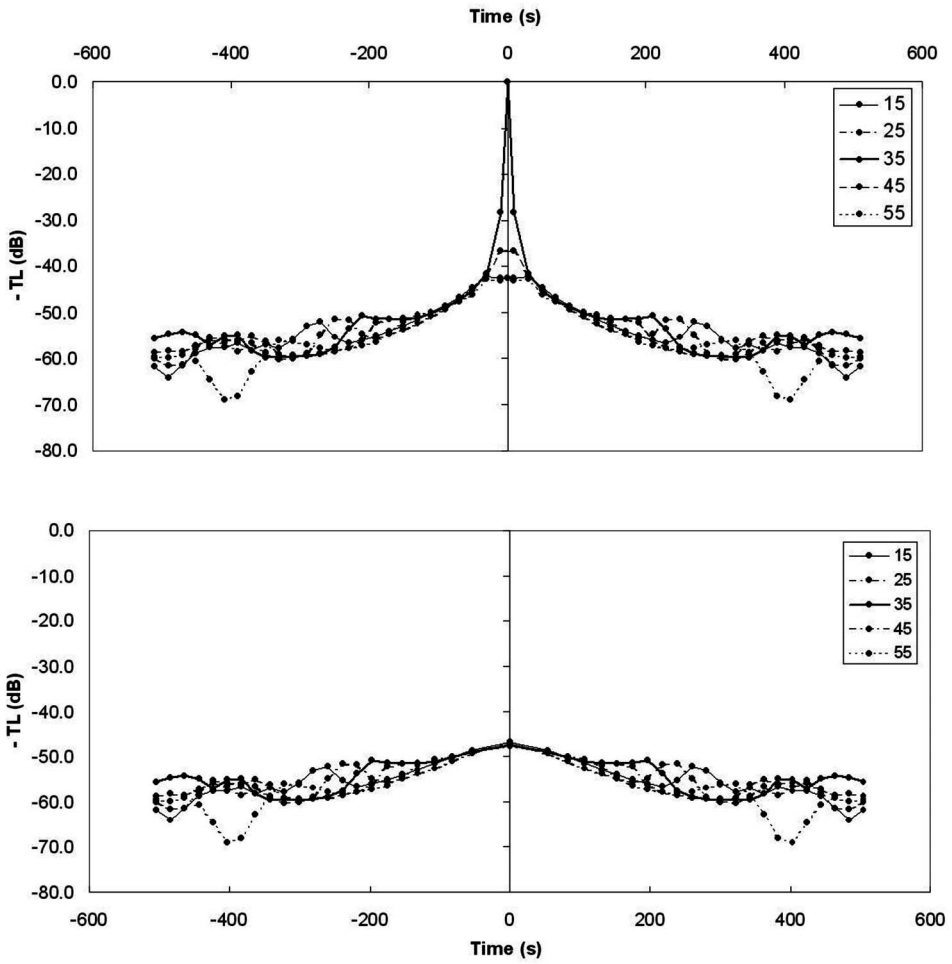


FIG. 1. Typical example of transmission loss (given in figure as TL) from the sonar source to the observation point of the echosounders as a function of time. Transmission started 1 nmi away from the observation point (the echosounder), and the source ship took 10 min (600 s) to pass the echosounder. The herring layer was usually distributed at depths between 10 and 50 m. Data are therefore presented for the individual depths; 15, 25, 35, 45, and 55 m. The source levels were 209 and 197 dB_{rms} (re 1 μPa at 1 m) for F1 and F2 transmissions, respectively, and received levels can be calculated as the difference between source level and TL. Transmission loss were calculated using the acoustic model LYBIN, with input parameters being the measured sound speed profiles and sonar source characteristics. The upper panel shows the received sound pressure levels at the echosounder being passed by the source ship, and the lower panel shows the received levels at the second echosounder, located ~400 m further south.

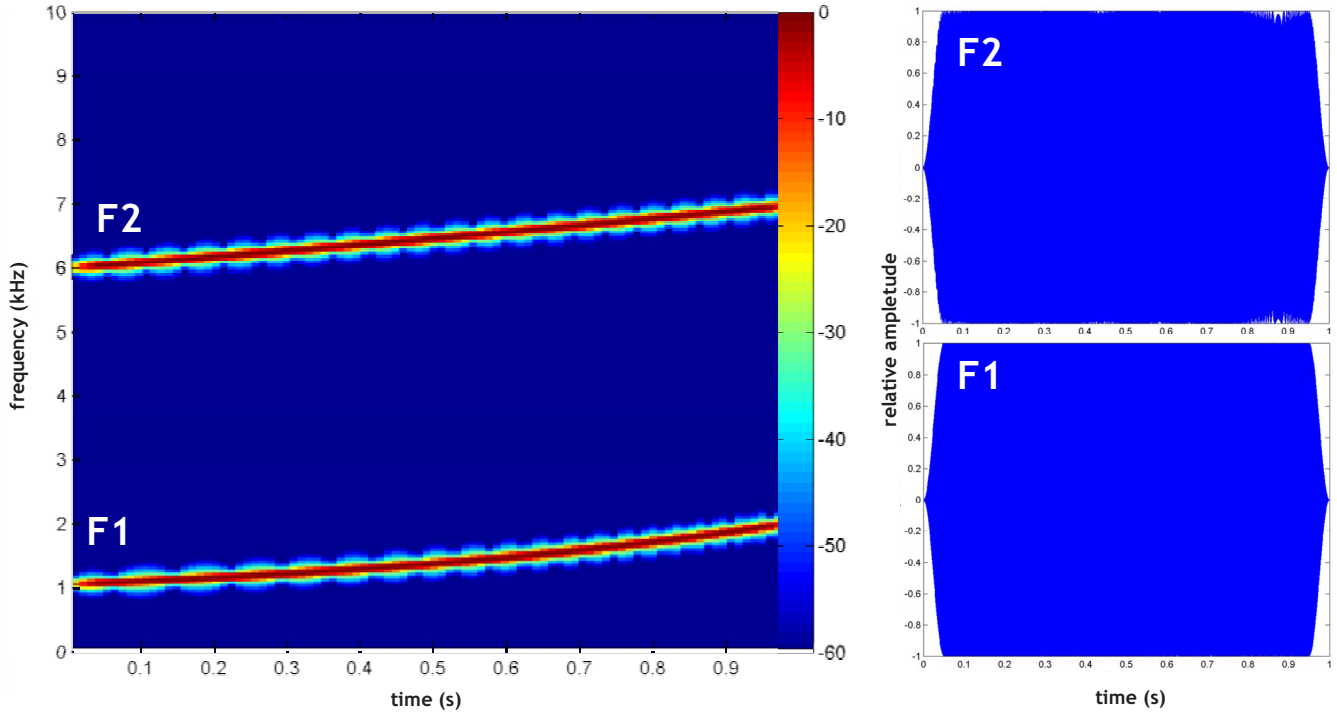


FIG. 2. (Color online) Spectogram and waveform of the transmitted sonar signals: F1 and F2. The left panel shows the spectrogram for F1 (lower curve) and F2 (upper curve), with frequency as a function of time. The scale on the left indicates intensity (dB). The right panel shows the corresponding waveforms, with relative amplitude as a function of time. The transmitted signals were hyperbolic frequency modulated waveforms from 1 to 2 kHz for F1, 6 to 7 kHz for F2, both with duration of 1 s.

Netherlands Navy. Socrates is equipped with two free-flooded ring transducers, one for each of the frequency bands (F1 and F2), installed in a towed body, and the system was operated from within the vessel's laboratory. The depth of the towed source was approximately 35 m in all experiments. Both the F1 and F2 signals were hyperbolic up-sweeps with signal duration of 1.0 s (Fig. 2). Pulse repetition time was 20 s. These signals are commonly used signals in naval sonar operations. The source levels were 209 dB_{rms} and 197 dB_{rms} (re 1 μPa at 1 m) for F1 and F2 signals, respectively. Before transmitting at full power, a ramp-up procedure was performed in order to mitigate potential impacts of sonar transmission on any marine mammals in the area. This procedure consisted of a gradually increasing the source level from 150 to 209 dB_{rms} (re 1 μPa at 1 m) for 10 min before F1 transmission, and from 138 to 197 dB_{rms} (re 1 μPa at 1 m) for 3 min before F2 transmission. Pulse length was 1.0 s and pulse repetition time 10 s during ramp-up. After ramp-up, full power transmission was initiated with 1 s pulses and 20 s pulse-repetition time.

The killer whale feeding sounds played to the herring were recorded using a digital acoustic recording tag attached to killer whales ("Dtag," Johnson and Tyack, 2003). The sound sequence was extracted from a Dtag recording of a killer whale that had been feeding on herring in the same general area a few days earlier. The Dtag contains a 400 Hz one pole high-pass filter and has a flat frequency response up to 45 kHz. Tag recordings also contained surfacing sounds, which were cut out of the record, and low-frequency flow noise due to the tagged animal movements, which were reduced by high-pass filtering at 800 Hz. The Lubell speaker has a response ± 8 dB from 600 Hz to 20 kHz. Therefore, the feedings sounds are only representative of actual killer whale feeding sounds over the frequency band of 800 Hz–20 kHz. During feeding, killer whales produce whistles, pulsed calls, and echolocation clicks, as well as intense sounds such as tail-slaps (Van Opzeeland, 2005; Simon *et al.*, 2007a). The feeding sounds played back included calls, echolocation clicks, and tail-slaps (Fig. 3) produced both by the tagged whale and other nearby whales. Because the feeding group in which the whale was tagged consisted of at least 20 animals, most sounds are likely from other whales than the tagged animal. The frequency content of most calls and whistles predominates above 800 Hz, but some low-frequency components of tail-slap sounds were likely removed due to the high-pass filter of the sound sequence. The source levels of the feeding sounds played from the speaker corresponded to previously described source levels of feeding killer whale calls [150–160 dB_{rms} (re 1 μPa at 1 m)] (Miller, 2006; Simon *et al.*, 2006).

C. Data analysis

One of the two echosounders was passed during each experiment. Herring at the other echosounder, positioned 400 m away, was thus exposed to a lower received sound pressure level. The passed and nonpassed echosounders were therefore compared with respect to the reactions of herring.

The echosounders recorded the acoustic volume backscatter strength by time and depth at a sampling frequency of 1 Hz over a 100 m range that spanned the main herring layer. Volume backscattering strength is defined as s_v

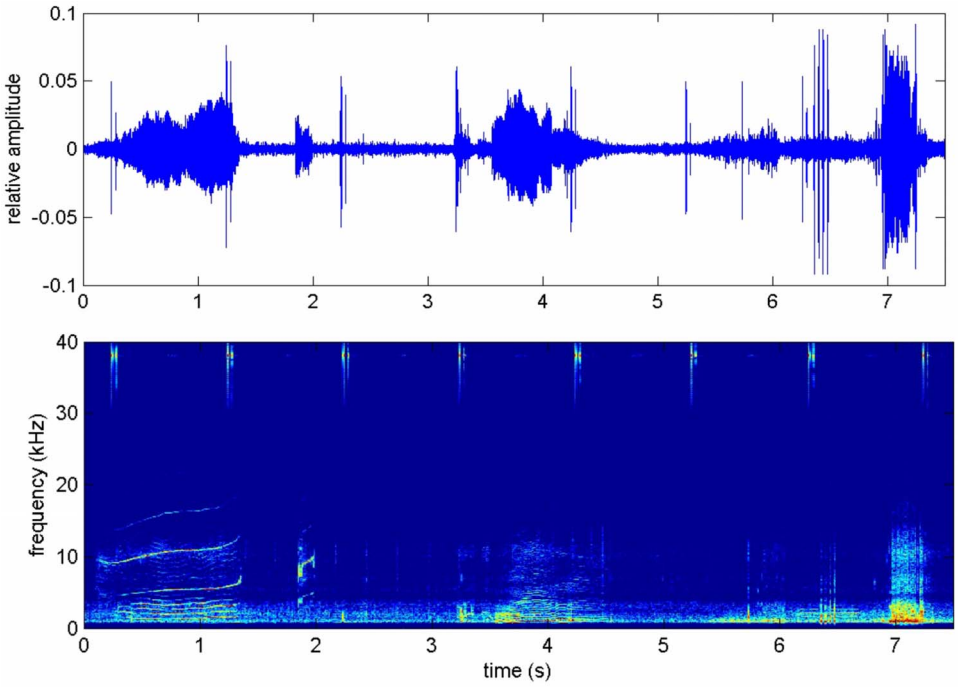


FIG. 3. (Color online) Waveform (top panel) and spectrogram (bottom panel) of a representative segment of the killer whale feeding sounds recorded during playback to herring. Note that the signals from the echosounder used to monitor the behavior of the herring are apparent at 38 kHz. The killer whale sounds include a number of calls, and a tail-slap sound (starting at 7 s). The call around 1 s is a typical example of a call resembling the sonar signals in duration and frequency content (see Fig. 2).

$=\Sigma\sigma_{bs}/V$ (m^{-1}), where V is volume, and σ_{bs} is the backscattering cross sections of individual targets within V (definitions given in MacLennan *et al.*, 2002). Two response variables, depth and s_v, are derived from the data for each passage. One passage is defined as the time interval from when the approaching vessel is 1 nmi away until 1 nmi beyond the bottom-mounted transducer. The depth variable is defined as the s_v weighted median depth throughout the passage, and s_v is defined as the mean volume backscattering s_v over each passing both within the 100 m range. Four different explanatory variables were used to model these response variables: (1) transmission type [F1/F2/control/killer whale playback (orca)], (2) order of the transmission types, (3) experiment, and (4) block number within an experiment. Generalized linear mixed models (GLMMs) (Littell *et al.*, 1996) were fitted to the data. These do not require errors to be independent, and permit a distinction to be made between random and fixed effects, where the errors may be conditional on a set of normally distributed random factors (McCulloch and Searle, 2001). Order and experiment were initially specified as random effects, but due to statistically negligible variance these were fitted as fixed in addition to the initial fixed factors: transmission type and block. In this model, the explanatory variable is linked to the response variable according to

$$g(E[y]) = X\beta + \varepsilon, \tag{1}$$

where y is the response vector (depth or s_v), X is the matrix of the fixed factors, β is the column vector of the fixed effect parameters to be estimated, and ε is the vector of the random errors. The function g is called a link function, a nonlinear function that relates $E[y_i]$ to the linear component $X_i\ \beta$. For the depth response, a normal-error model with an identity link function was used, while for the s_v response a gamma-error model with a log link function. The GLMMs were fitted in SAS® Version 9.1 using the GLIMMIX procedure (SAS Institute Inc., 2003). The model estimates β, and probability limits (p-values) of less than 0.05 were used to indicate whether this represented a significant factor in explaining the response. Multiple comparisons (Tukey's) were performed on those factors found to be significant.

III. RESULTS

Six experiments comprising a total of 14 blocks were performed. Four of the experiments consisted of F1-F2-control blocks, with three blocks in each experiment. The last two experiments each consisted of only one block, and the transmission was F1—killer whale playback (orca)—control (Table I).

Depth and s_v values from the passed and the nonpassed echosounder were compared, but neither were significantly different ($p > 0.05$) in any of the experiments or transmission types. During those passages that produced significant avoidance reactions (killer whale playback passages, see Sec. III B), a reaction was detected on both echosounders, indicating that the produced sound was fully detectable for the herring at this range. Data from both echosounders were therefore included in the analysis.

A. Herring reactions to sonar transmission

No obvious difference between the two types of transmission (F1 and F2) and the control could be seen when inspecting the echograms (Fig. 4). However, there was a local effect around the towed body during passage (Fig. 4), regardless of transmission type.

1. Depth response

Experiment was the only significant factor in explaining the average depth response ($p < 0.001$), with Experiment 4 having a herring layer significantly deeper and Experiment 5

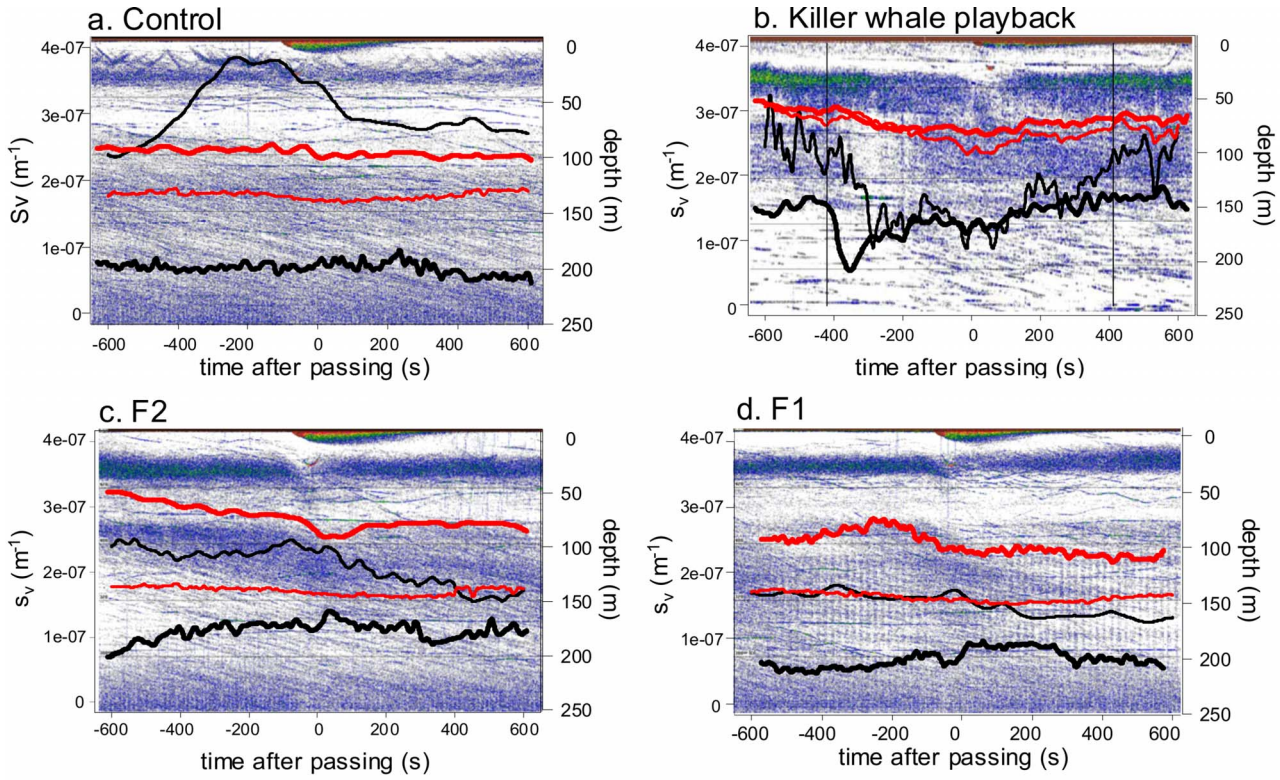


FIG. 4. (Color online) Typical echogram examples. Responses to the sounds were measured as herring density, s_v and herring vertical distribution, *depth*, and are presented as a function of time. Thick lines represent the particular experiment that the echogram is taken from, while thin lines are the average of all passages of this transmission type. The vessel wash from the passing vessel as well as the towed body sonar can be seen as strong distinct echoes around time=0. (a) Control; passage with vessel and sonar source without any transmission. The upper line is s_v experiment, followed by *depth* experiment, *depth* average, and s_v average on the bottom. (b) Playback of killer whale feeding sounds. The vertical lines indicate start and stop of playback. The two upper lines indicate *depth* experiment and *depth* average, respectively, the lower ones indicate average and experiment, respectively. (c) F2 transmission (6–7 kHz). The upper line is *depth* experiment, followed by s_v average, *depth* average, and s_v experiment on the bottom. (d) F1 transmission (1–2 kHz). The upper line is *depth* experiment, followed by *depth* average, s_v average, and s_v experiment on the bottom. No clear differences between sonar transmission (F1/F2) and the control can be detected by inspecting the echograms. A small vertical drop in the herring layer is seen at the point in time when the towed body sonar passes, but this reaction is similar for all types of transmission, and hence probably an avoidance to the source rather than the sound. In response to the passage involving playback of killer whale feeding sounds (b), there is a reduction in density that starts before passage of the source, almost immediately after onset of the sound, indicating that this reaction is to the sound. Echograms (a), (c), and (d) are from November 12, 2006, while (b) is from November 22, 2006.

significantly shallower than the rest [Table II, Fig. 5(a)]. No significant effects were found for the following factors: transmission type ($p=0.247$), block ($p=0.268$), or order of transmission types within a block ($p=0.840$).

TABLE II. Significant Tukey comparisons of the four factors included in the statistical model (experiment, block, transmission type, and order of transmissions), three factors had significant effect in explaining the average horizontal (S_v) and vertical (depth) response of herring; experiment (significant for depth and S_v), block, and transmission (significant for S_v).

Factor	Significant differences, s_v-response	Significant differences, depth-response
Experiment	Exp. 1-Exp. 3	Exp. 1-Exp. 4
	Exp. 2-Exp. 3	Exp. 2-Exp. 4
	Exp. 2-Exp. 4	Exp. 3-Exp. 4
	Exp. 3-Exp. 4	Exp. 5-Exp. 4
	Exp. 3-Exp. 5	
	Exp. 4-Exp. 2	
	Exp. 4-Exp. 5	
Block	Block 1-Block 3	
	Block 2-Block 3	
Transmission	Control-Orca	
	F2-Orca	

2. Density response

Significant factors in explaining the average s_v response were experiment ($p<0.001$) and block ($p=0.0003$). *Post hoc* (Tukey) comparisons showed which of the different experiments and blocks that differed (Table II). Experiment 3 had a significantly higher s_v, and Experiment 4 had significantly lower s_v than the rest [Fig. 5(a)]. Within the experiments, block 3 had a significantly lower s_v than the first two ($p=0.0003$) [Fig. 5(b)]. There was neither significant effect of the order of the transmission types ($p=0.914$) [Fig. 5(d)] nor between the three types of transmission (F1-F2 control) ($p=0.529$) [Fig. 5(c)]; hence the sonar signals (F1 and F2) did not cause any reaction different from that of a passage without any transmission (control).

B. Herring reactions to playback of killer whale feeding sounds

The sonar transmission passages, F1 and F2, were not significantly different from the control passages of no trans-

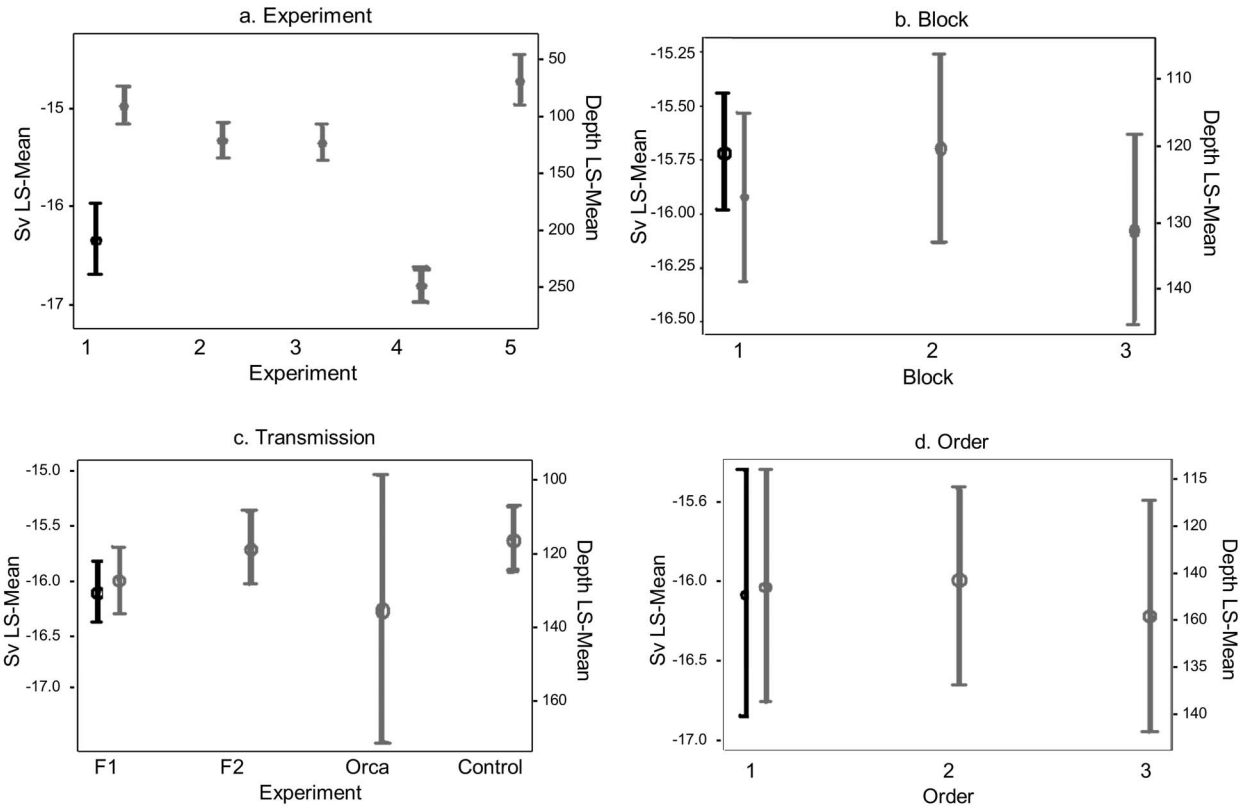


FIG. 5. Estimates and 95% confidence bounds for the predicted average density response, s_v (left bars in black), and vertical response, *depth* (right bars in gray), for the following factors: (a) experiment, (b) block within experiment, (c) transmission type, and (d) order of the types of transmission. (a) There were significant differences between the experiments. Experiment 4 produced significantly lower values than the other days regarding both s_v and depth. Experiment 3 had significantly higher *sv* than the other days. (b) Block 3 had a significantly lower s_v than blocks 1 and 2 within an experiment, but no significant differences were found with respect to depth. (c) There were no significant differences between the two sonar transmission types (F1 and F2) and the control either for s_v or depth. Playback of killer whale feeding sounds (Orca), however, had significantly lower s_v values than F2 and control. (d) There was no significant effect of the order of the types of transmission, nor for s_v nor depth.

mission. The passages involving playback of killer whale feeding sound did, however, produce significantly lower s_v than the control passages ($p=0.016$), indicating a reduced density in herring when exposed to the killer whale sounds. Killer whale playback passages also had clearly lower s_v than those of the F1 and F2 passages, but this was significant only for F2 ($p=0.046$ for F2, $p=0.067$ for F1). Visual inspection of the echograms involving killer whale passages clearly shows a reduction in herring density (s_v) almost immediately after the start of playback [Fig. 4(b)]. The average depth associated with the killer whale playback passages was notably lower than during other transmission types [Fig. 5(c)]. However, this difference was not significant ($p=0.335$). The estimate for killer whale playbacks had far lower precision both for s_v and depth, a natural consequence of having only two experiments of this type, compared to 12 for the other transmission types.

IV. DISCUSSION

This study has documented how overwintering NSS herring react to typical military sonar signals in the frequency band of 1–7 kHz, and has important implications for establishing guidelines for a safe operation of military sonars in areas densely populated by herring.

A. Experimental methods

The present results demonstrate that overwintering herring do not avoid sonar sounds at the tested received levels by neither horizontal nor vertical escape reactions. The reliability of these findings is strengthened by the immediate reduction in density and vertical movement seen during passages involving playback of killer whale feeding sounds. These passages demonstrate the nature of herring avoidance reactions, as well as the capability of the experimental setup to detect and describe such reactions.

We used experimental setups similar to those employed by Vabø *et al.* (2001) and Ona *et al.* (2007) for studying herring reactions to vessel noise. In the present study, it was essential to know the avoidance effects caused by the vessel and the towed sonar source, in order to separate behavioral reactions caused by an emitted signal from that caused by the vessel. Our results shows an intermittent drop in the herring layer at the time the vessel with the towed sonar passed the echosounder (visible effect in echograms in Fig. 4 at around time=0). This reaction lasted less than a few minutes and resembles the response characterized as vessel avoidance by Ona *et al.* (2007). This avoidance was the same for all types of passages, including the silent controls with no sonar transmission. This reaction is therefore likely to be caused by

avoidance to the passing vessel. It might also be an avoidance of the wire towing the sonar, as has been previously described by Handegard and Tjøstheim (2005), or possibly an avoidance of the towed body itself. The observed reaction also occurred within the same time interval as the measured vessel avoidance (within 2 min before vessel passage) (Ona *et al.*, 2007). With the source levels tested, sonar sound was well within the detection range of herring from the onset of transmission (approximately 10 min before vessel passage) (Fig. 1). The reaction to killer whale playback showed an avoidance reaction starting at about the time of sound onset, and a similar reaction should thus be expected for a potential sonar reaction. The statistical analysis was conducted on s_v and depth values averaged over the entire period of full power transmission, totally approximately 20 min. The vessel effect was detected only by the passed echosounder, but there was no significant difference in average s_v and depth between the two echosounders, indicating no confounding effect on the statistical analyses. The experimental setup was therefore considered adequate to separate a reaction to the sonar from that caused by a vessel/wire reaction.

B. Herring reactions to sonar transmission and killer whale playback

There was no significant reduction of herring density (s_v) or vertical position (depth) of the herring layers during runs involving sonar transmission (F1 or F2) compared to the control runs without any transmission. The daytime experiment (Experiment 3) produced significant differences in herring distribution [Table II, Fig. 5(a)], attributable to the observed typical diel variation (Huse and Korneliussen, 2000). Such variations, however, were taken into account in the models. There was also a significant reduction in herring density in the last block of each experiment [Fig. 5(b)]. The results presented here suggest that this is more likely to have been caused by diel variation or an adaptive response to the vessel and towed body than exposure to sonar.

The playback of killer whale feeding sounds induced an immediate dispersal response and downward movement of the herring, clearly visible in the echograms [Fig. 4(b)]. Both the average s_v and depth values were noticeably lower than during the control passages [Fig. 5(c)], although the effect was only significant for s_v, and the power relatively weak. However, the experimental effort was unbalanced with respect to the playback of killer whale sounds, with only approximately 1/6 of the effort used on measuring responses compared to the other types of transmission. It is compelling to speculate in that a study with equal effort put into the killer whale playbacks as to the other transmission types might provide a strong statistical case for differences. More studies of killer whale playback experiments of similar setup should therefore be performed.

In this study, source levels during full power transmission were 197 and 209 dB_{rms} re 1 μPa at 1 m for F1 and F2, respectively. With a maximum transmission loss of −70 dB (see Fig. 1), herring were exposed to a minimum received sound pressure level of 127 dB_{rms} (F1) and 139 dB_{rms} (F2) re 1 μPa. Received level increased as the source ship moved closer to the observation point at the echosounder, as a function of distance/time (Fig. 1). At night, when most of the experiments were carried out, high-density layers of herring were located between 10 and 50 m, and the source was towed at 35 m. At the closest point of drive-by, some herring were thus within a few meters range of the source, and received sound pressure levels will thus approximately equal the source levels (197 and 209 dB_{rms} re 1 μPa at 1 m for F1 and F2, respectively). The precise source level of operational military sonars within the different nations navies are often regarded classified information, but are likely to exceed the level used in the present experiment. We can thus not exclude the possibility of an effect when received levels exceed those tested here. However, the volume of water exposed to such levels is relatively small, and the fish biomass exposed to levels above 209 dB_{rms} re 1 μPa would be too small to have any effect on the population level (Kvadsheim and Sevaldsen, 2005).

Herring in the area are primarily caught by purse seine vessels, with herring catchability being strongly dependent on the diel migration toward the surface at night (described by Huse and Koreliussen, 2000). The present results show that naval sonar does not affect this behavior and is therefore not expected to have any negative influence on the fishing fleet's ability to catch herring. It is also unlikely that the specific conditions in the test location including background noise levels, the presence of shipping or other factors, would decrease the sensitivity of the fish.

Very few studies have examined how military sonars may affect fish. Jørgensen *et al.* (2005) investigated the effect of 1.5–6.5 kHz sonar signals on juvenile fish of different developmental stages and species, including herring. No tissue damage was found and postexposure development was normal, but juvenile herring showed strong behavioral reactions when exposed to sonar signals close to the assumed resonance frequency of the swimbladder. When exposed to levels above 180–190 dB_{rms} re 1 μPa, significant mortality was observed in juvenile herring (Jørgensen *et al.*, 2005), but Kvadsheim and Sevaldsen (2005) showed that this mortality would constitute less than 1% of the daily mortality of juvenile herring. Compared to natural mortality, even large scale military exercises would thus not significantly impact stocks of juvenile herring significantly at a population level. Thus, apparently herring is particularly sensitive to acoustic stimuli when they are exposed to sound corresponding to the resonance frequency band of the swimbladder. This resonance frequency band will depend somewhat on the species morphology but is mainly determined by the size of the fish and the depth (Løvik and Hovem, 1979). Adult herring at 10–50 m depth, which is the depth of the herring layer in most of the present experiments, is expected to have a resonance frequency between 1.0 and 2.5 kHz (Løvik and Hovem, 1979), corresponding to F1. Thus, this study has shown that even when exposed to sonar signals corresponding to swimbladder resonance, adult herring does not appear to react significant to such signals. However, the signal type used was frequency modulated sweeps, which will barely touch on the resonance frequency band for a very short moment. Long duration continuous wave signals may have a

different effect. Popper *et al.* (2007) studied the effect of very low-frequency sonar signals (below 1 kHz) and found a minor auditory temporary threshold shift, but no mortality, nor damage on tissue or sensory cells in the rainbow trout (*Oncorhynchus mykiss*). The present results are supported by those of Popper *et al.* (2007) that fish do not avoid a military sonar transmitting within their hearing range. Slotte *et al.* (2004) studied the behavioral effects of seismic shooting on herring and found no short term effects. A long-term decrease in biomass following a period of seismic shooting was observed, but they pointed out that this might just as well have been caused by feeding migration or natural fluctuations.

The reaction to the killer whale feeding sounds did cause an avoidance reaction by the herring, suggesting the intriguing possibility that fish were able to distinguish the killer whale feeding sounds from the sonar sounds. Reactions by clupeid fish to sounds of odontocete predators are also documented in previous studies (Mann *et al.*, 1998; Wilson and Dill, 2002).

Killer whales are the main predator on overwintering herring in Vestfjorden (Similä and Ugarte, 1993; Similä, 1997; Nøttestad, 1998), and such avoidance reactions as observed here are often seen in response to attacks (Nøttestad, 1998; Nøttestad and Axelsen, 1999), with hearing assumed to be an important cue inducing predator avoidance (Similä and Ugarte, 1993). Predation pressure by killer whales during overwintering is severe (Nøttestad and Axelsen, 1999), and awareness and vigilance are important to be able to escape from a predator. On the other hand, herring do not feed during overwintering, and energy minimization is thus of great importance (Slotte, 1999). Overwintering herring should thus be in a state of high sensitivity to predator calls, but escape reactions, with high energetic costs, should be avoided when unnecessary. Being able to discriminate predator sounds from other similar sounds will offer a great advantage and maximize energy conservation by limiting reactions to real threats. It is, however, not known which sound component in the recording used during the killer whale playback that triggered the escape reactions. In addition to the F1 signals, which are of high similarity in frequency and frequency modulation characteristics as some killer whale feeding calls (Miller, 2006; VanParijs *et al.*, 2004; Moore *et al.*, 1988), the sounds played back also included tail slaps and echolocation clicks. However, the majority of the energy in the echolocation clicks (Simon *et al.* 2007b) is above the hearing capability of herring (Enger, 1967; Mann *et al.*, 2005). Tail slaps (Simon *et al.*, 2005) on the other hand will be highly audible to herring, but potential sounds of frequencies lower than 800 Hz will be cut off due to the high-pass filter of the speaker. Hence, the sounds that the herring reacted to were mainly in a similar frequency range as the sonar signals. In addition to frequency and waveform, fish may be able to evaluate the repetition rate of a sound signal representing an odontocete predator (Astrup and Møhl, 1998). Which characteristic of the sound is played back cannot be determined in this study. Some of the difference in the reaction to the killer whale sounds playbacks may have been due to differences in how they were presented. The killer whale sounds were played back from a speaker located within the herring layer from the start of the playback, while the sonar source was gradually approaching the herring. However, the present results show that the experimental setup used in this study were adequate to reveal a potential escape reaction, thus acting as a negative control for the lack of response to the sonar signals.

Even though the present results demonstrate that overwintering herring show a lack of avoidance when exposed to sonar signals above 1 kHz, herring are known to change their behavior according to their functional, physiological, and motivational states (feeding, spawning, overwintering, and migrating) in terms of catchability (Mohr, 1964; Mohr, 1971) schooling dynamics, swimming speed, and reactions to different stimuli (Nøttestad *et al.*, 1999). Reactions to vessel noise also differed from being strong during the overwintering period (Vabø *et al.*, 2001) to relatively weak during prespawning (Skaret *et al.*, 2006). This may indicate that herring in different stages of their life history phase perhaps also may react differently to a military sonar. Such differences in reactions are controlled by trade-offs between predation risk, spawning success, and feeding, all of which differ between functional states (Nøttestad *et al.*, 1999). This demonstrates the need for more studies of how herring in different life history stages may react to military sonars. Nøttestad *et al.* (1999) found the most pronounced difference in behavior between herring in the nonfeeding state and during feeding after spawning was terminated. The present study concerned nonfeeding, overwintering herring, and a future study should thus focus on postspawned, feeding herring.

V. CONCLUSIONS

The results presented in this study leads to the conclusion that the operation of sonar systems at the tested frequencies and source levels (above 1 kHz and 209 dB_{rms} re 1 μPa at 1 m) will not have any large scale detrimental effects on overwintering herring populations or on the commercial herring fishery. Thus, such sonar systems may be safely operated in areas of overwintering herring, such as in Vestfjorden. More studies of similar character should be performed, involving both sonar transmission and killer whale playback, on herring during parts of their yearly cycle, such as, e.g., during the feeding or spawning period in order to allow the results to be more widely generalized.

ACKNOWLEDGMENTS

We would like to thank the scientific and regular crew onboard the RV HU Sverdrup II during the 3S-2006 field trial. Nina Nordlund and Erik Sevaldsen (FFI) are acknowledged for their assistance in performing oceanographic measurements and acoustic propagation analysis. Frank Benders, Peter Fritz, Adri Gerk, Sander van Ijsselmuide, Joost Kromjongh, Myriam Robert, Timo van der Zwan (TNO), and René Dekeling (Netherlands Defence Materiel Organization) are thanked for their assistance in handling and operating the sonar source during long nights at sea. Ken Foote (Woods Hole Oceanographic Institution) and Michael Ainslie (TNO) are greatly acknowledged for his comments on and correc-

tions to the manuscript. This project was financially supported by the Royal Norwegian Navy and the Norwegian Ministry of Defence and by the Defence Research and Development, The Netherlands.

Astrup, J., and Møhl, B. (**1998**). "Discrimination between high and low repetition rates of ultrasonic pulses by the cod," J. Fish Biol. **52**, 205–208.

Blaxter, J. H. S., Denton, E. J., and Gray, J. A. B. (**1979**). "Herring swimbladder as a gas-reservoir for the acousticolateralis system," J. Mar. Biol. Assoc. U.K. **59**, 1–10.

Chapman, C. J., and Hawkins, A. D. (**1973**). "A field study of hearing in the cod, Gadus morhua," J. Comp. Physiol. **85**, 147–167.

Denton, E. J., and Gray, J. A. B. (**1979**). "The analysis of sound by the sprat ear," Nature (London) **28**, 404–406.

Dragesund, O., Johannessen, A., and Ulltang, Ø. (**1997**). "Variation in migration and abundance of Norwegian spring spawning herring (*Clupea harengus L.*)," Sarsia **82**, 97–105.

Engås, A., and Løkkeborg, S. (**2002**). "Effects of seismic shooting and vessel-generated noise on fish behaviour and catch rates," Bioacoustics **12**, 313–315.

Engås, A., Løkkeborg, S., Ona, E., and Soldal, A. V. (**1996**). "Effects of seismic shooting on local abundance and catch rates of cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*)," Can. J. Fish. Aquat. Sci. **53**, 2238–2249.

Enger, P. S. (**1967**). "Hearing in herring," Comp. Biochem. Physiol. **22**, 527–538.

Enger, P. S. (**1981**). "Frequency discrimination in teleosts—Central or peripheral?," in *Hearing and Sound Communication in Fishes*, edited by W. A. Tavolga, A. N. Popper, and R. R. Fay (Springer, New York), pp. 243–255.

Fay, R. R., and Popper, A. N. (**2000**). "Evolution of hearing in vertebrates: The inner ears and processing," Hear. Res. **149**, 1–10.

Hamre, J. (**1990**). "Life history and exploration of the Norwegian spring-spawning herring," in *Biology and Fisheries of the Norwegian Spring Spawning Herring and Blue Whiting in the Northeast Atlantic*, edited by T. Monstad (Institute of Marine Research, Bergen, Norway), pp. 5–39.

Handegard, N. O., and Tjøstheim, D. (**2005**). "When fish meet a trawling vessel: Examining the behaviour of gadoids using a free-floating buoy and acoustic split-beam tracking," Can. J. Fish. Aquat. Sci. **62**, 2409–2422.

Hastings, M. C., and Popper, A. N. (**2005**). "Effects of Sound on Fish," Technical Report, Jones and Stokes, California Department of Transportation, Sacramento, CA.

Hastings, M. C., Popper, A. N., Finneran, J. J., and Lanford, P. J. (**1996**). "Effects of low-frequency underwater sound on hair cells of the inner ear and lateral line of the teleost fish Astronotus ocellatus," J. Acoust. Soc. Am. **99**, 1759–1762.

Hawkins, A. D., and Johnstone, A. D. F. (**1978**). "The hearing of Atlantic salmon, *Salmo salar*," J. Fish Biol. **13**, 655–673.

Hofman, R. J. (**2004**). "Marine sound pollution: Does it merit concern?," Mar. Technol. Soc. J. **37**, 66–77.

Holst, J. C., Drangesund, O., Hamre, J., Misund, O. A., and Østvedt, O. J. (**2002**). "Fifty years of herring migrations in the Norwegian Sea," ICES Marine Science Symposia, Vol. **215**, pp. 352–360.

Holst, J. C., Røttingen, I., and Melle, W. (**2004**). "The herring," in *The Norwegian Sea Ecosystem*, edited by H. R. Skjoldal (Tapir Academic, Trondheim, Norway), pp. 203–226.

Huse, I., and Korneliussen, R. (**2000**). "Diel variation in acoustic density measurements of overwintering herring (*Clupea harengus L.*)," ICES J. Mar. Sci. **57**, 903–910.

Johannessen, A., Slotte, A., Bergstad, O. A., Drangesund, O., and Røttingen, I. (**1995**). "Reappearance of the Norwegian spring spawning herring (*Clupea harengus L.*) at the spawning grounds of south western Norway," in Biology and Coastal Waters: Proceedings of the Symposium on the Ecology of Fjords and Coastal Waters, Tromsø, Norway 5–9 December 1994, pp. 347–364.

Johnson, M., and Tyack, P. L. (**2003**). "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," Inf. Sci. (N.Y.) **28**, 3–12.

Jørgensen, R., Olsen, K. K., Falk-Pettersen, I.-B., and Kanapthippilai, P. (**2005**). "Investigation of potential effects of low frequency sonar signals on survival, development and behaviour of fish larvae and juveniles," Report, Norwegian College of Fishery Science, Tromso, Norway.

Kenyon, T. N., Ladich, F., and Yan, H. Y. (**1998**). "A comparative study of hearing ability in fishes: The auditory brainstem response approach," J. Comp. Physiol. [A] **182**, 307–318.

Kvadsheim, P. H., and Sevaldsen, E. (**2005**). "The potential impact of 1–8 kHz active sonar on stocks of juvenile fish during sonar exercises," FFI-rapport 2005/01027 http://rapporter.ffi.no/rapporter/2005/01027.pdf (Last viewed April 2008).

Kvadsheim, P., Benders, F., Miller, P., Doksæter, L., Knudsen, F., Tyack, P., Nordlund, N., Lam, F. P., Samarra, F., Kleivane, L., and Godø, O. R. (**2007**). "Herring (sild), killer whales (spekkhogger) and sonar—The 3S-2006 cruise report with preliminary results," FFI-rapport 2007/01189, http://rapporter.ffi.no/rapporter/2007/01189.pdf (Last viewed April, **2008**).

Ladich, F., and Popper, A. N. (**2004**). "Parallel evolution in fish hearing organs," in *Evolution of the Vertebrate Auditory System*, edited by G. A. Manley, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 98–127.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (**1996**). *SAS System for Mixed Models* (SAS Institute Inc., Cary, NC).

Løvik, A., and Hovem, J. M. (**1979**). "An experimental investigation of swimbladder resonance in fishes," J. Acoust. Soc. Am. **66**, 850–854.

MacLennan, D. N., Fernandes, P. G., and Dalen, J. (**2002**). "A consistent approach to definitions and symbols in fisheries acoustics," ICES J. Mar. Sci. **59**, 365–369.

Mann, D. A., Lu, Z., Hastings, M. C., and Popper, A. N. (**1998**). "Detection of ultrasonic tones and simulated dolphin echolocation clicks by a teleost fish, the American shad (*Alosa sapidissima*)," J. Acoust. Soc. Am. **104**, 562–568.

Mann, D. A., Popper, A. N., and Wilson, B. (**2005**). "Pacific herring hearing does not include ultrasound," Biol. Lett. **1**, 158–161.

McCauley, R. D., Fewtrell, J., and Popper, A. N. (**2003**). "High intensity anthropogenic sound damage fish ears," J. Acoust. Soc. Am. **113**, 638–642.

McCulloch, C. E., and Searle, S. R. (**2001**). *Generalized, Linear, and Mixed Models* (Wiley, New York).

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.

Mohr, H. (**1964**). "Reaction of herring to fishing gear revealed by echosounder," in *Modern Fishing Gear of the World 2*, edited by G. von Brandt (Fishing News Books, London, UK), pp. 253–257.

Mohr, H. (**1971**). "Behaviour patterns of different herring stocks in relation to ship and midwater trawl," in *Modern Fishing Gear of the World 3*, edited by H. Kristjonsson (Fishing News Books, London, UK), pp. 368–371.

Moore, S. E., Francine, J. K., Bowles, A. E., and Ford, J. K. B. (**1988**). "Analysis of calls of killer whales, *Orcinus orca*, from Iceland to Norway," Rit Fiskideildar **11**, 225–2250.

Nøttestad, L. (**1998**). "Extensive gas bubble release in Norwegian spring-spawning herring (*Clupea harengus*) during predator avoidance," ICES J. Mar. Sci. **55**, 1133–1140.

Nøttestad, L., and Axelsen, B. E. (**1999**). "Herring school manoeuvres in response to killer whale attacks," Can. J. Zool. **77**, 1540–1547.

Nøttestad, L., Aksland, M., Beltestad, A., Fernö, A., Johannessen, A., and Misund, O. A. (**1999**). "Schooling dynamics of Norwegian spring spawning herring (*Clupea harengus L.*) in a coastal spawning area," Sarsia **80**, 277–284.

Olsen, K., Angell, J., Pettersen, F., and Løvik, A. (**1983**). "Observed fish reactions to a surveying vessel with special reference to herring, cod, capelin and polar cod," FAO Fisheries Report, FAO, Rome, Italy.

Ona, E., Godø, O. R., Handegard, N. O., Hjellvik, V., Patel, R., and Pedersen, G. (**2007**). "Silent vessels are not quiet," J. Acoust. Soc. Am. **121**, 145–150.

Patel, R. (**2007**). "Surveillance of marine resources by use of stationary platforms and Autonomous Underwater Vehicle (AUV)," Ph.D. thesis, Norwegian University of Science and Technology, Trondheim.

Popper, A. N. (**2003**). "Effects of anthropogenic sound on fishes," Fisheries **28**, 24–31.

Popper, A. N. (**1972**). "Pure tone auditory system for the carp (*Cyprinus carpio*)," J. Acoust. Soc. Am. **52**, 1714–1717.

Popper, A. N., Fewtrell, J., Smith, M. E., and McCauley, R. D. (**2004**). "Anthropogenic sound: Effects on the behaviour and physiology of fishes," Mar. Technol. Soc. J. **37**, 35–40.

Popper, A. N., Smith, M. E. Cott, P. A., Hanna, B. W., MacGillivary, A. O., Austin, M. E., and Mann, D. A. (**2005**). "Effects of exposure to seismic airgun use on hearing of three fish species," J. Acoust. Soc. Am. **117**,

3958–3971.
Popper, A. N., Halvorsen, M. B., Kane, A., Miller, D. L., Smith, M. E., Song, J., Stein, P., and Wyscocki, L. E. (**2007**). "The effects of high-intensity, low-frequency active sonar on rainbow trout," J. Acoust. Soc. Am. **122**, 623–635.
Richardson, W. J., Greene, C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).
Røttingen, I., and Slotte, A. (**2001**). "The relevance of a former important spawning area in the present life history and management of Norwegian spring-spawning herring," in *Herring: Expectations for a New Millennium*, edited by F. Funk, J. Blackburn, D. Hay, A. J. Paul, R. Stephenson, R. Toresen, and D. Withell (University of Alaska Sea Grant, Fairbanks, AK), pp. 297–313.
SAS Institute Inc. (**2003**). The SAS System for Windows (Version 9.1), Cary, NC.
Similä, T. (**1997**). "Sonar observations of killer whales (Orcinus orca) feeding on herring schools," Aquat. Mamm. **23**, 119–126.
Similä, T., and Ugarte, F. (**1993**). "Surface and underwater observations of cooperatively feeding killer whales in northern Norway," Can. J. Zool. **71**, 1494–1499.
Simon, M., Wahlberg, M., Ugarte, F., and Miller, L. A. (**2005**). "Acoustic characteristics of underwater tail slaps used by Norwegian and Icelandic killer whales (*Orcinus orca*) to debilitate herring (*Clupea harengus*)," J. Exp. Biol. **208**, 2459–2466.
Simon, M., Ugarte, F., Wahlberg, M., and Miller, L. A. (**2006**). "Icelandic killer whales Orcinus orca use a pulsed call suitable for manipulating the schooling behaviour of herring Clupea harengus," Bioacoustics **16**, 57–74.
Simon, M., McGregor, P. K., and Ugarte, F. (**2007a**). "The relationship between the acoustic behaviour and surface activity of killer whales (*Orcinus orca*) that feed on herring (*Clupea harengus*)," Acta Ethologica **10**, 47–53.
Simon, M., Wahlberg, M., and Miller, L. A. (**2007b**). "Echolocation clicks from killer whales feeding on herring," J. Acoust. Soc. Am. **121**, 749–752.
Skaret, G., Slotte, A., Handegard, N. O., Axelsen, B. E., and Jørgensen, R. (**2006**). "Pre-spawning herring in a protected area showed only moderate reaction to a surveying vessel," Fish. Res. **78**, 359–367.
Slotte, A. (**1999**). "Differential utilization of energy during wintering and spawning migration in Norwegian spring-spawning herring," J. Fish Biol. **54**, 338–355.
Slotte, A., Hansen, K., Dalen, J., and Ona, E. (**2004**). "Acoustic mapping of pelagic fish distribution and abundance in relation to a seismic shooting area off the Norwegian west coast," Fish. Res. **67**, 143–150.
Smith, M. E., Andrew, S. K., and Popper, A. N. (**2004**). "Noise-induced stress response and hearing loss in goldfish (*Carassius auratus*)," J. Exp. Biol. **207**, 427–435.
Strager, H. (**1995**). "Pod specific call repertoires and compound calls of killer whales, Orcinus orca Linneaus, 1758, in waters of northern Norway," Can. J. Zool. **73**, 1037–1044.
Vabø, R., Olsen, K., and Huse, I. (**2002**). "The effect of vessel avoidance of wintering Norwegian spring spawning herring," Fish. Res. **58**, 59–77.
Van Opzeeland, I. C., Corkeron, P. J., Leyssen, T., Similä, T., and Van Parijs, S. (**2005**). "Acoustic behaviour of Norwegian killer whales, *Orcinus orca*, during carousel and seiner foraging on spring-spawning herring," Aquat. Mamm. **31**, 110–119.
Van Parijs, S., Leyssen, T., and Similä, T. (**2004**). "Sound produced by Norwegian killer whales, *Orcinus orca*, during capture," J. Acoust. Soc. Am. **116**, 557–560.
Wilson, B., and Dill, L. M. (**2002**). "Pacific herring respond to simulated odontocete echolocation sounds," Can. J. Fish. Aquat. Sci. **59**, 542–553.

ACOUSTICAL NEWS—USA

Elaine Moran
Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502

Editor's Note: *Readers of this journal are encouraged to submit news items on awards, appointments, and other activities about themselves or their colleagues. Deadline dates for news items and notices are* 2 months *prior to publication.*

Announcement of the 2009 Election

In accordance with the provisions of the bylaws, the following Nominating Committee was appointed to prepare a slate for the election to take place on 8 May 2009: Anthony A. Atchley, Chair; Fredericka Bell-Berti; Courtney B. Burroughs; Brandon Tinianov; Lisa Zurk.

The bylaws of the Society require that the Executive Director publish in the *Journal*, at least 90 days prior to the election date, an announcement of the election and the Nominating Committee's nominations for the offices to be filled. Additional candidates for these offices may be provided by any Member or Fellow in good standing by letter received by the Executive Director not less than 60 days prior to the election date, and the name of any eligible candidate so proposed by 50 Members or Fellows shall be entered on the ballot. Biographical information about the candidates and statements of objectives of the candidates for President-Elect and Vice President-Elect will be mailed with the ballots.

CHARLES E. SCHMID
Executive Director

The Nominating Committee has submitted the following slate:

For President-Elect

George V. Frisk

Donna L. Neff

For Vice President-Elect

James V. Candy

Judy R. Dubno

 0001-4966/2009/125(1)/565/3/$25.00

Juan I. Arvelo

Damian J. Doria

Brenda L. Lonsbury-Martin

James H. Miller

Shrikanth Narayanan

Scott D. Sommerfeldt

USA Meetings Calendar

Listed below is a summary of meetings related to acoustics to be held in the U.S. in the near future. The month/year notation refers to the issue in which a complete meeting announcement appeared.

2009

18–22 May — 157th Meeting of the Acoustical Society of America, Portland, OR [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

24–28 June — 5th International Middle-Ear Mechanics in Research and Otology (MEMRO), Stanford University, Stanford, CA [http://memro2009.stanford.edu].

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

2010

19–23 Apr — 158th Meeting of the Acoustical Society of America, Baltimore, MD [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

15–19 Nov — 2nd Iberoamerican Conference on Acoustics (Joint Meeting of the Acoustical Society of America, Mexican Institute of Acoustics, and Iberoamerican Federation on Acoustics), Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

Cumulative Indexes to the Journal of the Acoustical Society of America

Ordering information: Orders must be paid by check or money order in U.S. funds drawn on a U.S. bank or by Mastercard, Visa, or American Express credit cards. Send orders to Circulation and Fulfillment Division,

American Institute of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2270. Non-U.S. orders add $11 per index.

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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

ACOUSTICAL STANDARDS NEWS

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

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

American National Standards (ANSI Standards) developed by Accredited Standards Committees S1, S2, S3, and S12 in the areas of acoustics, mechanical vibration and shock, bioacoustics, and noise, respectively, are published by the Acoustical Society of America (ASA). In addition to these standards, ASA publishes catalogs of Acoustical Standards, both National and International. To receive copies of the latest Standards Catalogs, please contact Susan B. Blaeser.

Comments are welcomed on all material in Acoustical Standards News.

This Acoustical Standards News section in JASA, as well as the National and International Catalogs of Acoustical Standards, and other information on the Standards Program of the Acoustical Society of America, are available via the ASA home page: http://asa.aip.org.

Standards Meetings Calendar

National Standards Meetings

• ***May 18–22, 2009***

Meetings of the National Standards Committees S1–Acoustics, S2–Mechanical Vibration and Shock, S3–Bioacoustics, S3/SC 1–Animal Bioacoustics, and S12–Noise, and the 10 U.S. TAGs administered by ASA will be held in conjunction with the 157th meeting of the Acoustical Society of America in Portland, Oregon.

International Standards Committee Meetings

• ***March 30–April 3, 2009***

Meetings of **ISO/TC 108/SC 2** will be held in London, UK.

• ***June 15–19, 2009***

Meetings of **ISO/TC 108/SC 5** will be held in Charlottenlund (Denmark).

Thanks to our Volunteers!

The ASA Standards Secretariat would like to thank the following people for volunteering their time and expertise to coordinate comments and formulate recommendations for the U.S. vote on numerous ISO and IEC documents throughout the past year. Their efforts are greatly appreciated and extremely valuable.

William Ahroon
Michael Bahtiarian
Raymond Bankert
Vesta I. Bateman
Philip Battenberg
Kenneth Bever
Bennett M. Brooks
Mahlon D. Burkhard
Kerry Cone
Mattew Craun
Kenneth J. Culverson
Thomas M. Disch
Renguang G. Dong
Bruce E. Douglas
Egons Dunens
Neil H. W. Eklund
Ronald L. Eshleman
David J. Evans
William C. Foiles
Mark Goodman
R. Lee Grason
Robert D. Hellweg
J. J. Henry
Ali T. Herfat
Thomas Jetzer
Arthur F. Kilcullen
Arnold Konheim
Kaleen X. C. Man
J. Adin Mann
Walter Madigosky
Alan H. Marsh
Eric H. Maslen
Macinissa Mezache
Douglas B. Moore
William J. Murphy
Anthony P. Nash
Victor Nedzelnitsky
John Niemkiewicz
Gary Orlove
Larry Pater
Richard J. Peppin
Christine Peterson
John K. Pollard
David A. Preves
Douglas D. Reynolds
Daniel G. Roley
Paul D. Schomer
Christopher J. Struck
David J. Vendittis
Michael J. White
Laura Ann Wilber
K. Allen Woo

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

ACCREDITED STANDARDS COMMITTEE ON ACOUSTICS, S1

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

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

S1 Working Groups

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

 0001-4966/2009/125(1)/568/8/$25.00

S1 Inactive Working Groups

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

S1 Standards on Acoustics

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

ACCREDITED STANDARDS COMMITTEE ON MECHANICAL VIBRATION AND SHOCK, S2

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

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

S2 Working Groups

S2/WG1—S2 Advisory Planning Committee (A.T. Herfat, Chair; R.L. Eshleman, Vice Chair);
S2/WG2—Terminology and Nomenclature in the Field of Mechanical Vibration and Shock and Condition Monitoring and Diagnostics of Machines (D.J. Evans);
S2/WG3—Signal Processing Methods (T.S. Edwards);
S2/WG4—Characterization of the Dynamic Mechanical Properties of Viscoelastic Polymers (W.M. Madigosky, Chair; J. Niemiec, Vice Chair);
S2/WG5—Use and Calibration of Vibration and Shock Measuring Instruments (D.J. Evans, Chair; B.E. Douglas, Vice Chair);
S2/WG6—Vibration and Shock Actuators (G. Booth);
S2/WG7—Acquisition of Mechanical Vibration and Shock Measurement Data (B.E. Douglas);
S2/WG8—Analysis Methods of Structural Dynamics (M. Mezache);
S2/WG9—Training and Accreditation (R. Eshleman, Chair; D. Corelli, Vice Chair);
S2/WG10—Measurement and Evaluation of Machinery for Acceptance and Condition (R.L. Eshleman, Chair; H. Pusey, Vice Chair);
S2/WG10/Panel 1—Balancing (R.L. Eshleman);
S2/WG10/Panel 2—Operational Monitoring and Condition Evaluation (R. Bankert);
S2/WG10/Panel 3—Machinery Testing (R.L. Eshleman);
S2/WG10/Panel 4—Prognosis (A.J. Hess);
S2/WG10/Panel 5—Data Processing, Communication, and Presentation (K. Bever);
S2/WG11—Measurement and Evaluation of Mechanical Vibration of Vehicles (A.F. Kilcullen);
S2/WG12—Measurement and Evaluation of Structures and Structural Systems for Assessment and Condition Monitoring (M. Mezache);
S2/WG13—Shock Test Requirements for Shelf-Mounted and Other Commercial Electronics Systems (B. Lang);
S2/WG39 (S3)—Human Exposure to Mechanical Vibration and Shock (D.D. Reynolds, Chair; R. Dong, Vice Chair).

S2 Inactive Working Group

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

S2 Standards on Mechanical Vibration and Shock

ANSI S2.1-2000/ISO 2041:1990 American National Standard Vibration and Shock—Vocabulary (Nationally Adopted International Standard)
ANSI S2.2-1959 (R 2006) American National Standard Methods for the Calibration of Shock and Vibration Pickups
ANSI S2.4-1976 (R 2004) American National Standard Method for Specifying the Characteristics of Auxiliary Analog Equipment for Shock and Vibration Measurements
ANSI S2.7-1982 (R 2004) American National Standard Balancing Terminology
ANSI S2.8-2007 American National Standard Technical Information Used for Resilient Mounting Applications
ANSI/ASA S2.9-2008 American National Standard Parameters for Specifying Damping Properties of Materials and System Damping
.**ANSI S2.16-1997 (R 2006)** American National Standard Vibratory Noise Measurements and Acceptance Criteria of Shipboard Equipment
ANSI S2.17-1980 (R 2004) American National Standard Techniques of Machinery Vibration Measurement
ANSI S2.19-1999 (R 2004) American National Standard Mechanical Vibration—Balance Quality Requirements of Rigid Rotors, Part 1: Determination of Permissible Residual Unbalance, Including Marine Applications
ANSI S2.20-1983 (R 2006) American National Standard Estimating Air Blast Characteristics for Single Point Explosions in Air, with a Guide to Evaluation of Atmospheric Propagation and Effects
ANSI S2.21-1998 (R 2007) American National Standard Method for Preparation of a Standard Material for Dynamic Mechanical Measurements
ANSI S2.22-1998 (R 2007) American National Standard Resonance Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials
ANSI S2.23-1998 (R 2007) American National Standard Single Cantilever Beam Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials
ANSI S2.24-2001 (R 2006) American National Standard Graphical Presentation of the Complex Modulus of Viscoelastic Materials
ANSI S2.25-2004 American National Standard Guide for the Measurement,

Reporting, and Evaluation of Hull and Superstructure Vibration in Ships
ANSI S2.26-2001 (R 2006) American National Standard Vibration Testing Requirements and Acceptance Criteria for Shipboard Equipment
ANSI S2.27-2002 (R 2007) American National Standard Guidelines for the Measurement and Evaluation of Vibration of Ship Propulsion Machinery
ANSI S2.28-2003 American National Standard Guide for the Measurement and Evaluation of Vibration of Shipboard Machinery
ANSI S2.29-2003 American National Standard Guide for the Measurement and Evaluation of Vibration of Machine Shafts on Shipboard Machinery
ANSI S2.31-1979 (R 2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 1: Basic Definitions and Transducers
ANSI S2.32-1982 (R 2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 2: Measurements Using Single-Point Translational Excitation
ANSI S2.34-1984 (R 2005) American National Standard Guide to the Experimental Determination of Rotational Mobility Properties and the Complete Mobility Matrix
ANSI S2.42-1982 (R 2004) American National Standard Procedures for Balancing Flexible Rotors
ANSI S2.43-1984 (R 2005) American National Standard Criteria for Evaluating Flexible Rotor Balance
ANSI S2.46-1989 (R 2005) American National Standard Characteristics to be Specified for Seismic Transducers
ANSI S2.48-1993 (R 2006) American National Standard Servo-Hydraulic Test Equipment for Generating Vibration—Methods of Describing Characteristics
ANSI S2.60-1987 (R 2005) American National Standard Balancing Machines—Enclosures and Other Safety Measures
ANSI S2.61-1989 (R 2005) American National Standard Guide to the Mechanical Mounting of Accelerometers
ANSI S2.70-2006 (*revision of ANSI S3.34-1986*) American National Standard Guide for the Measurement and Evaluation of Human Exposure to Vibration Transmitted to the Hand
ANSI S2.71-1983 (R 2006) (*reaffirmation and redesignation of ANSI S3.29-1983*) American National Standard Guide to the Evaluation of Human Exposure to Vibration in Buildings
ANSI S2.72/Part 1-2002 (R 2007)/ISO 2631-1:1997 *(redesignation of ANSI S3.18/Part 1-2002/ISO 2631-1:1997)* American National Standard Mechanical vibration and shock -Evaluation of human exposure to whole-body vibration—Part 1: General requirements (Nationally Adopted International Standard)
ANSI S2.72/Part 4-2003 (R 2007)/ISO 2631-4:2001 *(redesignation of ANSI S3.18/Part 4-2003/ISO 2631-4:2001)* American National Standard Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 4: Guidelines for the evaluation of the effects of vibration and rotational motion on passenger and crew comfort in fixed-guideway transport systems (Nationally Adopted International Standard)
ANSI S2.73-2002/ISO 10819:1996 (R 2007) *(redesignation of ANSI S3.40-2002/ISO 10819:1996)* American National Standard Mechanical vibration and shock—Hand-arm vibration—Method for the measurement and evaluation of the vibration transmissibility of gloves at the palm of the hand (Nationally Adopted International Standard)

ACCREDITED STANDARDS COMMITTEE ON BIOACOUSTICS, S3

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

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

S3 Working Groups

S3/Advisory—Advisory Planning Committee to S3 (C.A. Champlin, Chair; D.A. Preves, Vice Chair);
S3/WG35—Audiometers (R.L. Grason);
S3/WG36—Speech Intelligibility (R.S. Schlauch);
S3/WG37—Coupler Calibration of Earphones (C. J. Struck);
S3/WG39—Human Exposure to Mechanical Vibration and Shock—Parallel to ISO/TC 108/SC 4 (D.D. Reynolds, Chair; R. Dong, Vice Chair);
S3/WG43—Method for Calibration of Bone Conduction Vibrators (J.D. Durrant);
S3/WG48—Hearing Aids (D.A. Preves);
S3/WG51—Auditory Magnitudes (R.P. Hellman);
S3/WG56—Criteria for Background Noise for Audiometric Testing (J. Franks);
S3/WG59—Measurement of Speech Levels (M.C. Killion and L.A. Wilber, Co-Chairs);
S3/WG60—Measurement of Acoustic Impedance and Admittance of the Ear (Vacant);
S3/WG62—Impulse Noise with Respect to Hearing Hazard (J.H. Patterson, Chair; R. Hamernik, Vice Chair);
S3/WG67—Manikins (M.D. Burkhard);
S3/WG72—Measurement of Auditory Evoked Potentials (R.F. Burkard);
S3/WG76—Computerized Audiometry (A.J. Miltich);
S3/WG79—Methods for Calculation of the Speech Intelligibility Index (C.V. Pavlovic);
S3/WG80—Probe-tube Measurements of Hearing Aid Performance (W.A. Cole);
S3/WG81—Hearing Assistance Technologies (L. Thibodeau and L.A. Wilber, Co-Chairs);
S3/WG82—Basic Vestibular Function Test Battery (C. Wall);
S3/WG83—Sound Field Audiometry (T.R. Letowski);
S3/WG84—Otoacoustic Emissions (G.R. Long);
S3/WG88—Standard Audible Emergency Evacuation and Other Signals (R. Boyer);
S3/WG89—Spatial Audiometry in Real and Virtual Environments (J. Besing);
S3/WG91—Text-to-Speech Synthesis Systems (C. Bickley and A.K. Syrdal, Co-Chairs).

S3 Liaison Group

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

S3 Inactive Working Groups

S3/WG71—Artificial Mouths (R.L. McKinley);
S3/WG85—Allocation of Noise-induced Hearing Loss (R.A. Dobie).

S3 Standards on Bioacoustics

ANSI S3.1-1999 (R 2008) American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms
ANSI S3.2-1989 (R 1999) American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems
ANSI S3.4-2007 American National Standard Procedure for the Computation of Loudness of Steady Sounds
ANSI S3.5-1997 (R 2007) American National Standard Methods for Calculation of the Speech Intelligibility Index
ANSI S3.6-2004 American National Standard Specification for Audiometers
ANSI S3.7-1995 (R 2008) American National Standard Method for Coupler Calibration of Earphones
ANSI S3.13-1987 (R 2007) American National Standard Mechanical Coupler for Measurement of Bone Vibrators
ANSI S3.20-1995 (R 2008) American National Standard Bioacoustical Terminology
ANSI S3.21-2004 American National Standard Methods for Manual Pure-Tone Threshold Audiometry
ANSI S3.22-2003 American National Standard Specification of Hearing Aid Characteristics
ANSI S3.25-1989 (R 2003) American National Standard for an Occluded Ear Simulator
ANSI S3.35-2004 American National Standard Method of Measurement of Performance Characteristics of Hearing Aids under Simulated Real-Ear Working Conditions
ANSI S3.36-1985 (R 2006) American National Standard Specification for a Manikin for Simulated *in situ* Airborne Acoustic Measurements
ANSI S3.37-1987 (R 2007) American National Standard Preferred Earhook Nozzle Thread for Postauricular Hearing Aids
ANSI S3.39-1987 (R 2007) American National Standard Specifications for

Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance)
ANSI S3.41-1990 (R 2001) American National Standard Audible Emergency Evacuation Signal
ANSI S3.42-1992 (R 2007) American National Standard Testing Hearing Aids with a Broad-Band Noise Signal
ANSI S3.44-1996 (R 2006) American National Standard Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment
ANSI S3.45-1999 American National Standard Procedures for Testing Basic Vestibular Function
ANSI S3.46-1997 (R 2007) American National Standard Methods of Measurement of Real-Ear Performance Characteristics of Hearing Aids

ACCREDITED STANDARDS COMMITTEE ON ANIMAL BIOACOUSTICS, S3/SC 1

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

Scope: Standards, specifications, methods of measurement and test, instrumentation and terminology in the field of psychological and physiological acoustics, including aspects of general acoustics which pertain to biological safety, tolerance and comfort of non-human animals, including both risk to individual animals and to the long-term viability of populations. Animals to be covered may potentially include commercially grown food animals; animals harvested for food in the wild; pets; laboratory animals; exotic species in zoos, oceanaria or aquariums; or free-ranging wild animals.

S3/SC 1 Working Groups

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

ACCREDITED STANDARDS COMMITTEE ON NOISE, S12

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

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

S12 Working Groups

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

S12 Liaison Groups

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

S12 Inactive Working Groups

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

S12 Standards on Noise

ANSI S12.1-1983 (R 2006) American National Standard Guidelines for the Preparation of Standard Procedures to Determine the Noise Emission from Sources
ANSI/ASA S12.2-2008 American National Standard Criteria for Evaluating Room Noise
ANSI S12.3-1985 (R 2006) American National Standard Statistical Methods for Determining and Verifying Stated Noise Emission Values of Machinery and Equipment
ANSI S12.5-2006/ISO 6926:1999 American National Standard Acoustics—Requirements for the Performance and Calibration of Reference Sound Sources Used for the Determination of Sound Power Levels (Nationally Adopted International Standard)
ANSI/ASA S12.6-2008 American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors
ANSI S12.7-1986 (R 2006) American National Standard Methods for Measurements of Impulse Noise
ANSI/ASA S12.8-1998 (R 2008) American National Standard Methods for Determining the Insertion Loss of Outdoor Noise Barriers
ANSI S12.9/Part 1-1988 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 1
ANSI/ASA S12.9/Part 2-1992 (R 2008) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 2: Measurement of Long-Term, Wide-Area Sound
ANSI/ASA S12.9/Part 3-1993 (R 2008) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 3: Short-Term Measurements with an Observer Present
ANSI S12.9/Part 4-2005 American National Standard Quantities and Pro-

cedures for Description and Measurement of Environmental Sound, Part 4: Noise Assessment and Prediction of Long-Term Community Response

ANSI/ASA S12.9/Part 5-2007 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 5: Sound Level Descriptors for Determination of Compatible Land Use

ANSI/ASA S12.9/Part 6-2008 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 6: Methods for Estimation of Awakenings Associated with Outdoor Noise Events Heard in Homes

ANSI/ASA S12.10-2002 (R 2007)/ISO 7779:1999 American National Standard Acoustics—Measurement of airborne noise emitted by information technology and telecommunications equipment (Nationally Adopted International Standard)

ANSI/ASA S12.11/Part 1-2003/ISO 10302:1996 (MOD) (R 2008) American National Standard Acoustics—Measurement of noise and vibration of small air-moving devices—Part 1: Airborne noise emission (Modified Nationally Adopted International Standard)

ANSI/ASA S12.11/Part 2-2003 (R 2008) American National Standard Acoustics—Measurement of Noise and Vibration of Small Air-Moving Devices—Part 2: Structure-Borne Vibration

ANSI/ASA S12.12-1992 (R 2007) American National Standard Engineering Method for the Determination of Sound Power Levels of Noise Sources Using Sound Intensity

ANSI S12.13 TR-2002 ANSI Technical Report Evaluating the Effectiveness of Hearing Conservation Programs through Audiometric Data Base Analysis

ANSI/ASA S12.14-1992 (R 2007) American National Standard Methods for the Field Measurement of the Sound Output of Audible Public Warning Devices Installed at Fixed Locations Outdoors

ANSI/ASA S12.15-1992 (R 2007) American National Standard For Acoustics—Portable Electric Power Tools, Stationary and Fixed Electric Power Tools, and Gardening Appliances—Measurement of Sound Emitted

ANSI/ASA S12.16-1992 (R 2007) American National Standard Guidelines for the Specification of Noise of New Machinery

ANSI S12.17-1996 (R 2006) American National Standard Impulse Sound Propagation for Environmental Noise Assessment

ANSI S12.18-1994 (R 2004) American National Standard Procedures for Outdoor Measurement of Sound Pressure Level

ANSI S12.19-1996 (R 2006) American National Standard Measurement of Occupational Noise Exposure

ANSI S12.23-1989 (R 2006) American National Standard Method for the Designation of Sound Power Emitted by Machinery and Equipment

ANSI S12.42-1995 (R 2004) American National Standard Microphone-in-Real-Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Circumaural Hearing Protection Devices

ANSI/ASA S12.43-1997 (R 2007) American National Standard Methods for Measurement of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions

ANSI/ASA S12.44-1997 (R 2007) American National Standard Methods for Calculation of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions from Sound Power Level

ANSI/ASA S12.50-2002 (R 2007)/ISO 3740:2000 American National Standard Acoustics -Determination of sound power levels of noise sources—Guidelines for the use of basic standards (Nationally Adopted International Standard)

ANSI/ASA S12.51-2002 (R 2007)/ISO 3741:1999 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision method for reverberation rooms (This Standard includes Technical Corrigendum 1–2001.) (Nationally Adopted International Standard)

ANSI S12.53/Part 1-1999 (R 2004)/ISO 3743-1:1994 American National Standard Acoustics—Determination of sound power levels of noise sources—Engineering methods for small, movable sources in reverberant fields—Part 1: Comparison method for hard-walled test rooms (Nationally Adopted International Standard)

ANSI S12.53/Part 2-1999 (R 2004)/ISO 3743-2:1994 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering methods for small, movable sources in reverberant fields—Part 2: Methods for special reverberation test rooms (Nationally Adopted International Standard)

ANSI S12.54-1999 (R 2004)/ISO 3744:1994 American National Standard Acoustics -Determination of sound power levels of noise sources using sound pressure—Engineering method in an essentially free field over a reflecting plane (Nationally Adopted International Standard)

ANSI S12.55-2006/ISO 3745:2003 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision methods for anechoic and hemi-anechoic rooms (Nationally Adopted International Standard)

ANSI S12.56-1999 (R 2004)/ISO 3746:1995 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Survey method using an enveloping measurement surface over a reflecting plane (Nationally Adopted International Standard)

ANSI/ASA S12.57-2002 (R 2007)/ISO 3747:2000 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Comparison method *in situ* (Nationally Adopted International Standard)

ANSI S12.60-2002 American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools

ANSI S12.65-2006 *(Revision of ANSI S3.14–1977)* American National Standard for Rating Noise with Respect to Speech Interference

ANSI/ASA S12.68-2007 American National Standard Methods of Estimating Effective A-Weighted Sound Pressure Levels When Hearing Protectors are Worn

ASA COMMITTEE ON STANDARDS (ASACOS)

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

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

ISO/TC 43 Acoustics, **ISO/TC 43 /SC 1** Noise (P.D. Schomer, U.S. TAG Chair)

ISO/TC 108 Mechanical vibration, shock, and condition monitoring (D.J. Evans, U.S. TAG Chair)

ISO/TC 108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures (A.F. Kilcullen and R.F. Taddeo, U.S. TAG Co-Chairs)

ISO/TC 108/SC 3 Use and calibration of vibration and shock measuring instruments (D.J. Evans, U.S. TAG Chair)

ISO/TC 108/SC 4 Human exposure to mechanical vibration and shock (D.D. Reynolds, U.S. TAG Chair)

ISO/TC 108/SC 5 Condition monitoring and diagnostics of machines (D.J. Vendittis, U.S. TAG Chair; R. Taddeo, U.S. TAG Vice Chair)

ISO/TC 108/SC 6 Vibration and shock generating systems (C. Peterson, U.S. TAG Chair)

IEC/TC 29 Electroacoustics (V. Nedzelnitsky, U.S. Technical Advisor)

STANDARDS NEWS FROM THE UNITED STATES

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

American National Standards Call for Comment on Proposals Listed

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

AMCA (Air Movement and Control Association)

Reaffirmations

BSR/AMCA 520–2004 (R200x), Laboratory Methods of Testing Actuators (reaffirmation of ANSI/AMCA 520–2004)

Establishes an industry standard for minimum rating and testing of actuators used on fire/smoke dampers. The testing requirements will cover torque or

force rating, long-term holding, operational life, elevated temperature performance, periodic maintenance, production, and sound testing for both pneumatic and electric operators. (December 8, 2008)

ASA (ASC S2) (Acoustical Society of America)

Reaffirmations

BSR/ASA S2.29-2003 (R200x), Guide for the Measurement and Evaluation of Vibration of Machine Shafts on Shipboard Machinery (reaffirmation and redesignation of ANSI S2.29–2003)

Gives guidelines for applying shaft vibration evaluation criteria, under normal operating conditions, measured at or close to the bearings of large shipboard machines with oil film bearings, such as main propulsion turbines and gears, and turbo-generators. These guidelines are presented in terms of steady vibration amplitudes and changes in amplitudes which may occur in these steady values. They apply to acceptance tests for new machinery, and in-situ testing for monitoring purposes. (November 3, 2008)

Withdrawals

ANSI S2.7-1982 (R2004), Balancing Terminology (withdrawal of ANSI S2.7–1982 (R2004))

Contains a collection of terms and definitions relating to balancing procedures and equipment. The terms are grouped into seven major categories, covering the subjects of: (1) Machines; (2) Rotors; (3) Unbalance; (4) Balancing; (5) Balancing Machines and equipment; (6) Flexible Rotors; and (7) Rotating Rigid Free-Bodies (i.e., aerospace vehicles). (November 17, 2008)

ANSI S2.17-1980 (R2004), Techniques of Machinery Vibration Measurement (withdrawal of ANSI S2.17-1980 (R2004))

Concerns vibration measurement quantities, equipment, and procedures involved in operating machinery. Calibration of vibration measurement and calibration equipment are discussed in the document. The standard has an application to preventive maintenance programs, equipment selection, and equipment quality. (November 24, 2008)

ASA (ASC S3) (Acoustical Society of America)

Reaffirmations

BSR S3.41-1990 (R200x), Audible Emergency Evacuation Signal (reaffirmation of ANSI S3.41-1990 (R2001))

Applies to an audible emergency signal used for and limited to situations requiring immediate evacuation from a building because of emergency. This standard specifies two parameters of the audible emergency evacuation signal, i.e., the temporal pattern and the required sound pressure level at all places with the intended signal reception area. It applies to the audible signal, not to the individual signaling system components. (October 27, 2008)

Revisions

BSR/ASA S3.45-200x, Procedures for Testing Basic Vestibular Function (revision and redesignation of ANSI S3.45-1999)

Defines test procedures, measurements, data analysis, and data reporting requirements for performing and reporting a battery of six different tests for the evaluation of human vestibular function ("Basic Vestibular Function Test Battery"). Stimuli are presented to evoke eye movement by a subject whose response is determined either by measurement of electrical signals generated by the eye movements or by image-processing methods applied to video eye movements. Test interpretation is not included. (November 24, 2008)

ASA (ASC S12) (Acoustical Society of America)

New Standards

BSR/ASA S12.67-200x, Pre-Installation Airborne Sound Measurements and Acceptance Criteria of Shipboard Equipment (new standard)

Describes the instrumentation and procedures for the pre-installation measurement and analysis of airborne noise generated by shipboard equipment. Maximum noise level criteria are presented for several types of equipment. This standard is based on MIL STD 740-1 "Airborne Sound Measurements and Acceptance Criteria of Shipboard Equipment" and MIL-STD-1474D, Requirement 5, "Shipboard Equipment Noise." (November 10, 2008)

ASTM

Reaffirmations

BSR/ASTM F2174-2002 (R200x), Practice for Verifying Acoustic Emission Sensor Response (reaffirmation of ANSI/ASTM F2174–2002) (October 27, 2008)

CSAA (Central Station Alarm Association)

New Standards

BSR/CSAA-CS-AUD-01–200x, Audio Verification Procedures for Burglar Alarms (new standard)

Defines monitoring procedures of burglar alarms by using the addition of audio and its transmission from the protected premises for the verification of alarm activity. The goal of this standard is to reduce the instances of false dispatches. (November 3, 2008)

TIA (Telecommunications Industry Association)

Reaffirmations

BSR/TIA 631-A-2002 (R200x), Telecommunications—Telephone Terminal Equipment—Radio Frequency Immunity Requirements (reaffirmation of ANSI/TIA 631-A-2002)

Specifies Radio Frequency (RF) immunity performance criteria for two-wire Telephone Terminal Equipment (TTE) having an acoustic output and two-wire TTE adjunct devices with connection port for Telephone Terminal Equipment (TTE) having an acoustic output. (October 27, 2008)

Revisions

BSR/TIA/EIA 136-440-C-200x, TDMA Third Generation Wireless Adaptive Multi Rate (AMR) Codec (revision and redesignation of ANSI/TIA/EIA 136.440-B-2006)

Provides a description of the AMR speech service, including speech coding, channel coding and link adaptation. (October 27, 2008)

Notice of Withdrawal: ANS at least 10 years past approval date

The following American National Standards have not been revised or reaffirmed within ten years from the date of their approval as American National Standards and accordingly are withdrawn:
ANSI/AWS F6.1-1978 (R1998), Method for Sound Level Measurement of Manual Arc Welding and Cutting Processes

Call for Members (ANS Consensus Bodies)

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

ASA (ASC S2) (Acoustical Society of America)

BSR/ASA S2.28-200x, Guide for the Measurement and Evaluation of Vibration of Shipboard Machinery (revision and redesignation of ANSI S2.28-2003)

ASA (ASC S12) (Acoustical Society of America)
BSR/ASA S12.42-200x, Microphone-In-Real-Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Hearing Protection Devices (revision and redesignation of ANSI S12.42-1995 (R2004))

ASA (ASC S3) (Acoustical Society of America)
BSR/ASA S3.25-200x, Occluded Ear Simulator (revision and redesignation of ANSI/ASA S3.25-200x)

American National Standards Final Action
The following American National Standards have received final approval from the ANSI Board of Standards Review.

ASA (ASC S3) (Acoustical Society of America)

Reaffirmations

ANSI/ASA S3.1-1999 (R2008), Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms (reaffirmation and redesignation of ANSI S3.1-1999 (R2003))

ANSI/ASA S3.7-1995 (R2008), Method for Coupler Calibration of Earphones (reaffirmation and redesignation of ANSI S3.7-1995 (R2003))

ANSI/ASA S3.20-1995 (R2008), Bioacoustical Terminology (reaffirmation and redesignation of ANSI S3.20-1995 (R2003))

ASA (ASC S12) (Acoustical Society of America)

Revisions

ANSI/ASA S12.6-2008, Methods for Measuring the Real-Ear Attenuation of Hearing Protectors (revision and redesignation of ANSI S12.6-1997 (R2002))

UL (Underwriters Laboratories, Inc.)

Revisions

ANSI/UL 1971-2008, Signaling Devices for the Hearing Impaired (revision of ANSI/UL 1971-2004)

Project Initiation Notification System (PINS)
ANSI Procedures require notification of ANSI by ANSI-accredited standards developers (ASD) of the initiation and scope of activities expected to result in new or revised American National Standards (ANS). Early notification of activity intended to reaffirm or withdraw an ANS and in some instances a PINS related to a national adoption is optional. The mechanism by which such notification is given is referred to as the PINS process. For additional information, see clause 2.4 of the ANSI Essential Requirements: Due Process Requirements for American National Standards.

ASA (ASC S2) (Acoustical Society of America)
BSR/ASA S2.28-200x, Guide for the Measurement and Evaluation of Vibration of Shipboard Machinery (revision and redesignation of ANSI S2.28-2003)
Contains procedures for the measurement and evaluation of the mechanical vibration of nonreciprocating machines, as measured on nonrotating parts. It contains criteria for evaluating new machines and for vibration monitoring. This American National Standard is related to the ISO 10816 series that provides guidelines for the evaluation of different types of machines. The type of machinery covered in this standard is shipboard machinery. Project Need: To revisit limits set on the broadband values, based on statistical analysis using new vibration data on 75 machines tested up to 20 times each. Stakeholders: Maritime industry, military, naval engineering, ship building.

ASA (ASC S3) (Acoustical Society of America)
BSR/ASA S3.25-200x, Occluded Ear Simulator (revision and redesignation of BSR/ASA S3.25-200x)
Specifies the acoustical performance of an occluded ear simulator. This device is designed to simulate the acoustical behavior of the ear canal between the tip of an earmold and the eardrum, including the acoustic impedance at the eardrum of a median adult human ear. The occluded ear simulator is also suitable as the basis for extensions intended to simulate the complete ear canal and the outer ear (e.g., head and torso simulators). Project Need: To acknowledge the compliance of different existing occluded ear simulators already deployed (and in use for 30 years), and new compliant devices. Harmonization with IEC 60318-4 is also desirable. Extension of the current frequency range of applicability may also be considered. Stakeholders: Hearing aid manufacturers, telecom manufacturers, consumer headphone and earphone manufacturers.

ASA (ASC S12) (Acoustical Society of America)
BSR/ASA S12.42-200x, Microphone-In-Real-Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Hearing Protection Devices (revision and redesignation of ANSI S12.42-1995 (R2004))
Provides two methods for measuring the insertion loss of any hearing protection device (HPD) that encloses the ears, caps the ears, or occludes the ear canals. It contains information on instrumentation, calibration, electroacoustic requirements, subject selection and training, procedures for locating ear-mounted microphones and HPDs to measure sound pressure levels at the ear, specifications describing suitable ATFs, and methods for reporting the calculated insertion-loss values. Project Need: To revise ANSI S12.42-1995 (R2004) so that it deals with earmuffs and earplugs. It also must be expanded to provide specifics on a variety of test signals to include impulse noise. The document must also be updated to correspond to the current version of ANSI S12.6-2008. Stakeholders: Industrial, military, and nonoccupational users of hearing protection devices and regulatory agencies.

ATIS (Alliance for Telecommunications Industry Solutions)
BSR ATIS 0600010.03-200x, Operational Vibration and Shock Requirements for Network Telecommunications Equipment Utilized in Central Office and Outside Plant Environments (new standard)
Provides evaluation criteria for industry to ensure that the effects of operational vibration and shock on network telecommunications are minimized. Project Need: To investigate the standardization of operational vibration and shock requirements for network telecommunications equipment. Stakeholders: Communications industry.

Standards News from Abroad

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

International Organization for Standardization (ISO)
Newly Published ISO and IEC Standards
Listed here are new and revised standards recently approved and promulgated by ISO—the International Organization for Standardization.

ISO Standards
MECHANICAL VIBRATION AND SHOCK (TC 108)
ISO 16063-12/Cor1:2008, Methods for the calibration of vibration and shock transducers—Part 12: Primary vibration calibration by the reciprocity method—Corrigendum
ISO 18436-4:2008, Condition monitoring and diagnostics of machines—Requirements for qualification and assessment of personnel—Part 4: Field lubricant analysis
ISO 18436-6:2008, Condition monitoring and diagnostics of machines—Requirements for qualification and assessment of personnel—Part 6: Acoustic emission
ISO 18436-7:2008, Condition monitoring and diagnostics of machines—Requirements for qualification and assessment of personnel—Part 7: Thermography

PERFORMANCE OF HOUSEHOLD ELECTRICAL APPLIANCES (TC 59)
IEC 60704-2-13 Amd.1 Ed. 1.0 b:2005, Amendment 1—Household and similar electrical appliances—Test code for the determination of airborne acoustical noise—Part 2-13: Particular requirements for range hoods

SMALL CRAFT (TC 188)

ISO 14509-1:2008, Small craft—Airborne sound emitted by powered recreational craft—Part 1: Pass-by measurement procedures

IEC Standards

ELECTRICAL ACCESSORIES (TC 23)

IEC 62080 Amd.1 Ed. 1.0 b:2008, Amendment 1—Sound signalling devices for household and similar purposes

ISO Draft Standards

ACOUSTICS (TC 43)

ISO/DIS 8253-1, Acoustics—Audiometric test methods—Part 1: Basic pure-tone air and bone conduction threshold audiometry (December 13, 2008)

ISO/DIS 10140-5, Acoustics—Laboratory measurement of sound insulation of building elements—Part 5: Requirements for test facilities and equipment (January 8, 2009)

ISO/DIS 10140-2, Acoustics—Laboratory measurement of sound insulation of building elements—Part 2: Measurement of airborne sound insulation (January 8, 2009)

ISO/DIS 10140-3, Acoustics—Laboratory measurement of sound insulation of building elements—Part 3: Measurement of impact sound insulation (January 8, 2009)

ISO/DIS 10140-4, Acoustics—Laboratory measurement of sound insulation of building elements—Part 4: Measurement procedures and requirements (January 8, 2009)

ISO/DIS 10140-1, Acoustics—Laboratory measurement of sound insulation of building elements—Part 1: Test codes (January 8, 2009)

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO/DIS 4866, Mechanical vibration and shock—Vibration of fixed structures—Guidelines for the measurement of vibrations and evaluation of their effects on structures (December 21, 2008)

ISO/DIS 18436-5, Condition monitoring and diagnostics of machines—Requirements for qualification and assessment of personnel—Part 5: Lubricant laboratory technician/analyst (January 6, 2009)

BOOK REVIEW

P. L. Marston
Physics Department, Washington State University, Pullman, Washington 99164

These reviews of books and other forms of information express the opinions of the individual reviewers and are not necessarily endorsed by the Editorial Board of this Journal.

Nonlinear Vibrations and Stability of Shells and Plates

Marco Amabili

Cambridge University Press, Cambridge, 2008. 374 pp. Price: $99.00 (hardcover) ISBN: 978-0-521-88329-0

Nonlinearity in the vibration of simple mechanical systems is a very important aspect of many branches of science. In acoustics it underlies the distortion in powerful loudspeakers and the shimmering sounds of cymbals and Chinese gongs, while in engineering it can lead to the buckling and failure of large structures associated with power sources. The cover of this book indicates its concern with the latter topics, since it displays photos of the lift-off of a space shuttle and the structure of an ultramodern building dome in Valencia, together with a plot of a vibrational mode of a square plate. The introduction adds photographs of the *Queen Mary 2*, the US Navy submarine *USS Asheville*, and the huge new Airbus 380 passenger aircraft.

A brief flick through the pages, however, shows that this is not a book for practical engineers but rather a detailed exposition of the underlying mathematics. Many of the pages in the first few chapters are mostly occupied by equations—there are 210 numbered equations in the 46 pages of Chapter 1. To assess the contents and the readership for whom it will be appropriate requires a much more detailed examination. The preface says that the book is aimed at "researchers, professionals, students and instructors" and that it contains "a complete text on the nonlinear mechanics of shells and plates" together with a "state of the art" exposition of their nonlinear vibrations and stability. This accords with my impressions. Despite their mathematical density, the early chapters are well set out and explained and do indeed tell you nearly all you need to know about the basis of the subject. After the basic equations of Chapter 1, the second chapter enters the more practical domain of doubly curved plates and plates with a laminated structure, which are common in aircraft and other applications. Chapter 3 then present an excellent introduction to nonlinear dynamics, bifurcations, and associated computational methods. A few terms for bifurcations and attractors, such as "dangerous," "explosive," and "blue sky catastrophe," are particularly attention grabbing!

The following chapters of the book concentrate on particular cases of practical importance and explore them in detail. Several chapters are devoted to circular cylindrical shells, both empty and filled with liquid, a topic of relevance to applications such as the fuel tanks of the space shuttle, which are considered explicitly. Some of the frequency response curves for this situation are really unusual! Along with computed results, the author also describes the results of his laboratory experiments on a system of this type and presents computed and experimental bifurcation diagrams and Poincaré maps. There is then a theoretical examination of the effects of axial load, boundary conditions, and other parameters, and experimental results are shown for some of these cases. Chapter 10 then leads us on to the behavior of shells with biaxial curvature, with the forward fuel tank of the space shuttle as an example, a matter that is returned to in a later chapter. The final subject treated is the stability of rotating circular plates, with relevance to computer hard-disk drives and similar systems.

Does the book cover most of the subjects I would expect? Perhaps not, since it is mostly limited to discussion of almost flat or circularly cylindrical shells with simple boundary conditions. Interesting cases not treated are, for example, a conical shell, a plate with a sharp change in slope, and a flat sheet rolled into a spiral like a clock spring. Certainly one can think of a large number of such deviant shapes, but these are generic enough that they might have been chosen as examples. The reason for limiting the geometries treated is presumably that their possible number is immense, and the book keeps to those with explicit solutions and clear engineering applications.

If it is read from an acoustical point of view, the book omits consideration of many of the subjects one might regard as important, such as harmonic generation and distortion as a function of amplitude and excitation frequency. Indeed the word "spectrum" does not appear anywhere in the index and figures 9.11 and 9.23, which do show spectra of a sort, are not what one would be looking for. This is not meant as a criticism, but simply to point out that the book is intended for mechanical engineers rather than acousticians.

The book is very well written and the mathematical analysis is clearly set out and explained. Each chapter has a list of typically 10 to 30 references and the text is illustrated with a good number of well-drawn diagrams, together with photographs of experimental equipment where this has been used. The index, at less than two double-column pages, is surprisingly short, but seems to be adequate.

The author is clearly an expert in this field, with many related publications, and he explains the subject very clearly, though the level of formality means that it is not easy to simply "dip in" to find the way to treat a problem—you need to read the whole chapter carefully, and perhaps the introductory chapters, too. It would be an excellent text for a graduate-level course in the subject for engineering students with a good background in mathematics. It should also appeal to established researchers in related areas who want to catch up with the latest theoretical approaches in this field without having to read all the published papers.

NEVILLE H. FLETCHER
Research School of Physical Sciences and Engineering
Australian National University
Canberra, ACT 0200, Australia

 0001-4966/2009/125(1)/576/1/$25.00

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, PB92
California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

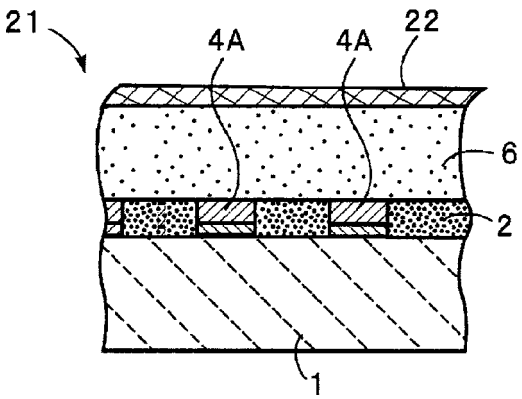
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026–2453*
JOHN ERDREICH, *Ostergaard Acoustical Associates, 200 Executive Drive, West Orange, New Jersey 07052*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue, M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,327,071

43.35.Pt SURFACE ACOUSTIC WAVE DEVICE

Kenji Nishiyama *et al.*, assignors to Murata Manufacturing Company, Limited
5 February 2008 (Class 310/313 A); filed in Japan 2 March 2004

A surface acoustic wave (SAW) filter and duplex element 21 having a zero resonance frequency temperature coefficient (TC) is claimed where the space between interdigital transducer fingers 4A is filled with silicon dioxide material 2 whose top surface is smoothed. A second SiO_2 layer 6 is applied,



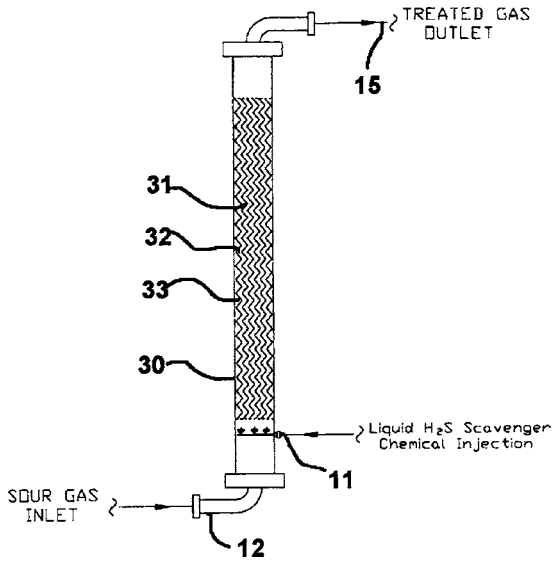
also smoothed, followed by a silicon nitride (SiN) topping 22. The latter is claimed to provide a temperature coefficient contrasting with that of SiO_2 and piezoelectric, resulting in a net zero TC for the SAW device.—AJC

7,326,392

43.35.Zc H_2S SCAVENGING METHOD

Kevin S. Fisher and Kenneth E. McIntush, assignors to Gas Technology Institute
5 February 2008 (Class 423/242.1); filed 2 August 2004

A gas scrubbing method enhanced by a mist produced at the face of an ultrasound transducer surface somewhere in the column 11-30-31. Removal



of hydrogen sulfide (H_2S) from gas flow 12 is enhanced by that mist of water droplets that are but a few microns in diameter, providing a droplet scrubbing area of 40 ft^2/ft^3 of gas and allowing the column length 30-31 to be much shorter.—AJC

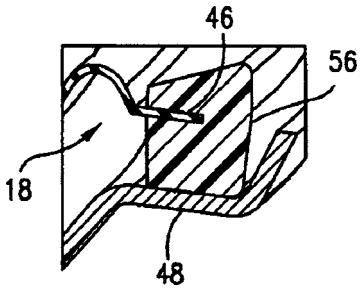
7,379,558

43.38.Dv LOUDSPEAKER WITH INTEGRATED SPIDER STANDOFF RING

Lucio Proni, assignor to JL Audio, Incorporated
27 May 2008 (Class 381/404); filed 16 December 2004

Aftermarket automotive loudspeakers suffer from various and sundry failure modes. One of these can be due to the variability in the attachment of the outer perimeter of the centering suspension to the speaker frame. Lower

 0001-4966/2009/125(1)/577/12/$25.00



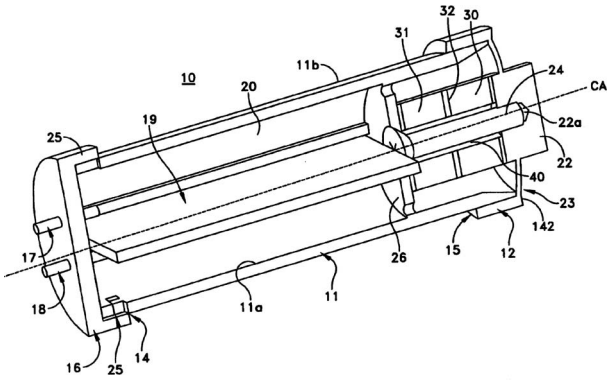
suspension element 18, commonly called a "spider," is attached to standoff 56 using an overmolding process. This assures that the outer rim of the spider is flat and can be repeatedly and assuredly attached to foot 48.—NAS

7,398,690

43.38.Fx ACOUSTIC PRESSURE SENSOR

David J. Erickson and Walter Chyrywaty III, assignors to Lockheed Martin Corporation
15 July 2008 (Class 73/702); filed 7 April 2006

This patent discloses the construction details of what appears to be a hydrophone receiver, although the application is not stated. The device is of a cylindrical design and is apparently intended for use as a receiver as it includes a cavity housing a preamplifier circuit board 19 next to the transducer elements 30, 31. The transducer elements are short, hollow cylinders



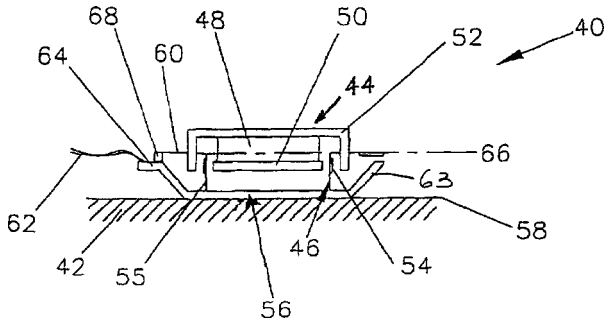
of piezoelectric material, most likely lead zirconate titanate for the stated frequency range of 20 Hz–20 kHz. It is not an especially unusual design, and no performance data are given to support the rather large claimed bandwidth. Most of the discussion concerns sealing methods for the housing.—JAH

7,372,968

43.38.Ja LOUDSPEAKER DRIVER

Andreas Buous and Martin Colloms, assignors to New Transducers Limited
13 May 2008 (Class 381/152); filed in United Kingdom 8 November 2000

Coupler 56 is used to attached driver 44 to planar acoustic radiator 42



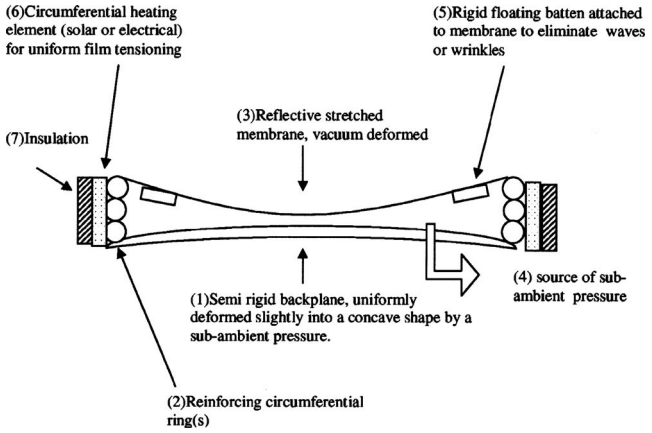
to allow easy replacement when said driver fails.—NAS

7,374,301

43.38.Ja STRETCHED MEMBRANE DEVICE

Douglas Evan Simmers, Massillon, Ohio
20 May 2008 (Class 359/847); filed 20 February 2005

Rigid batten 5 is attached to membrane 3 to prevent "wrinkles" in the



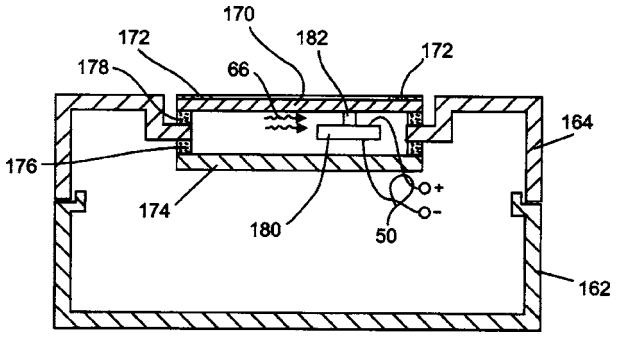
membrane when it is deformed to a concave shape.—NAS

7,403,628

43.38.Ja TRANSDUCER ASSEMBLY AND LOUDSPEAKER INCLUDING RHEOLOGICAL MATERIAL

Matthew J. Murray, assignor to Sony Ericsson Mobile Communications AB
22 July 2008 (Class 381/152); filed 7 April 2004

A number of prior patents describe ways in which some part of a cellular phone case can also serve as a sound emitting diaphragm. This patent argues that the driving mechanism for such an inertial transducer is fragile and can be damaged if the phone is bumped or dropped. So, instead



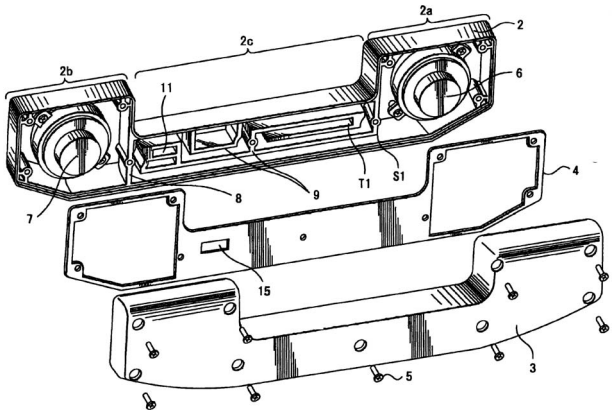
of a rigid connection between the two, a special resilient material is used that becomes rigid in the presence of a magnetic field, somewhat like an electromagnetic fluid clutch. The idea may or may not be practical, but it is certainly unusual—a welcome change from patents that require an electron microscope to detect any novel feature.—GLA

7,410,029

43.38.Ja BASS REFLEX TYPE LOUDSPEAKER APPARATUS, LOUDSPEAKER BOX AND IMAGE DISPLAY APPARATUS

Hideo Tanaami, assignor to Sony Corporation
12 August 2008 (Class 181/156); filed in Japan 9 March 2005

This is a good example of a "junk patent;" it displays negligible originality and provides no real protection. Its only apparent purpose is to act as



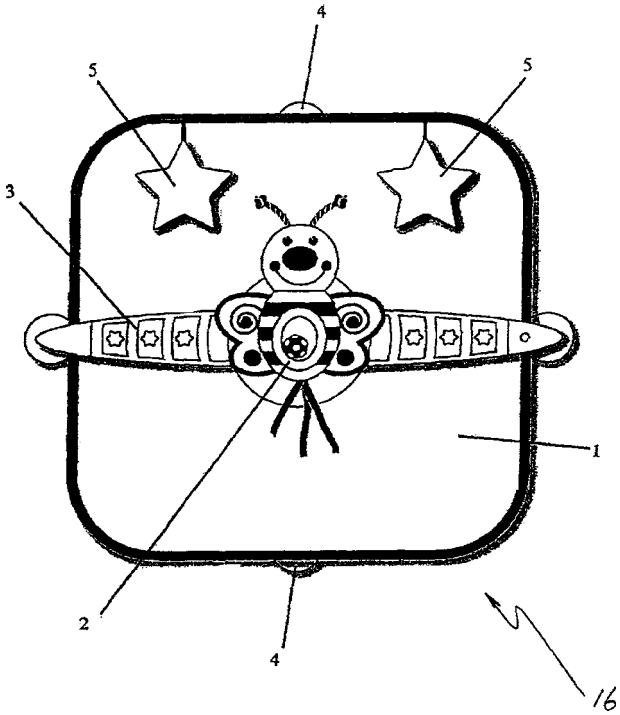
a corporate trading stamp for Sony. The area of interest is a box containing two loudspeakers, each with its own vented rear chamber (familiar prior art thus far). What the patent claims set forth is a specific geometry for folding and nesting the two vent tunnels. In the space of 20 min or so, any engineer could come up with a half-dozen equally acceptable alternatives.—GLA

7,398,816

43.38.Lc WINDOW SHADE

Andre Sala, assignor to Star H.K. Electronic Limited
15 July 2008 (Class 160/368.1); filed in United Kingdom 5 February 2004

Let us imagine a small child sitting next to a window in a residence or automobile. Bright sunlight streams through the window, suggesting the need for some kind of window shade. The child's parent or guardian, having acquired the special shade described in this patent, affixes it to the window



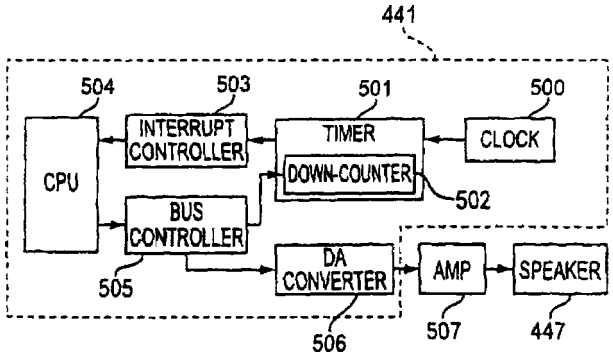
using its built-in suction cups. The aforesaid parent or guardian then operates a handheld remote control to activate the shade. Lights flash, stuffed animals dance, and music emanates from a concealed loudspeaker. The child is amused and, for a few minutes at least, the parent or guardian can attend to other matters.—GLA

7,406,355

43.38.Md METHOD FOR GENERATING PLAYBACK SOUND, ELECTRONIC DEVICE, AND ENTERTAINMENT SYSTEM FOR GENERATING PLAYBACK SOUND

Toru Morita, assignor to Sony Computer Entertainment Incorporated
29 July 2008 (Class 700/94); filed in Japan 21 January 1999

Although this patent is aimed at a specialized application relating to video games, the basic principle has wider potential uses. We are told that a game can be partially loaded into one or more control modules that can then be detached and used independently from the main game console. Sound functions must be included in each control module. Audio effects are triggered by an interrupt signal generated by a timer. Unless the timer operates



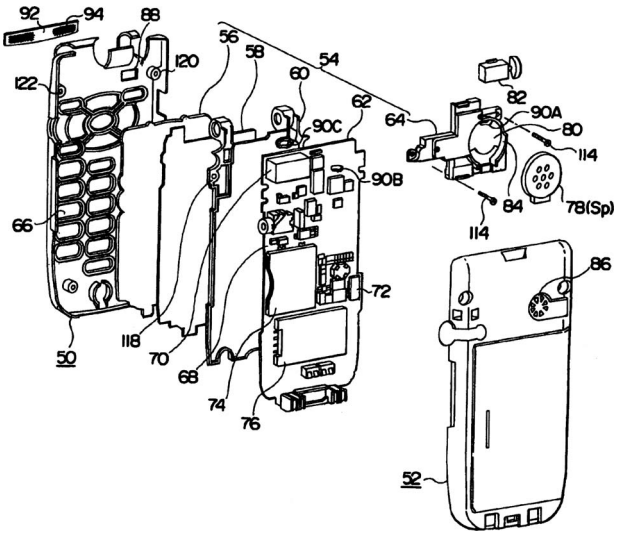
at a very high frequency, the digital audio data (say, a 1 kHz tone) may be initiated at a point well above or below a zero crossing, introducing undesired transient distortion. However, if the interrupt timing is deliberately perturbed to jibe with a period of the audio signal, the correct waveform is preserved.—GLA

7,400,910

43.38.Si SPEAKER SOUND ENHANCEMENT FOR A MOBILE TERMINAL

Yoshiya Matsumoto *et al.*, assignors to Fujitsu Limited
15 July 2008 (Class 455/570); filed in Japan 13 June 2003

How many engineers does it take to make a cellular phone audible whether the case is open or closed? In this instance, four. The novelty lies in



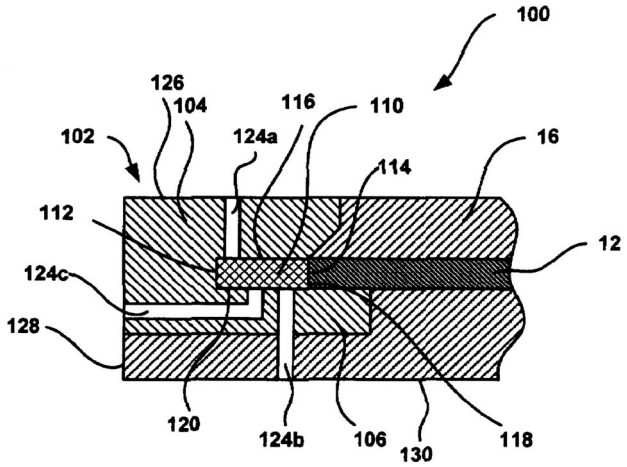
conducting sound from the back of the speaker through a short duct to a rear vent without allowing it to be hindered by "wiring boards and the like."—GLA

7,382,048

43.38.Si ACOUSTIC TRANSDUCER MODULE

Anthony D. Minervini, assignor to Knowles Electronics, LLC
3 June 2008 (Class 257/723); filed 28 April 2005

Microelectronic mechanical system (MEMS) 110 is connected to printed circuit board 12 via connecting surface 114. The MEMS has connections on the side toward 114 so that the MEMS does not need to be



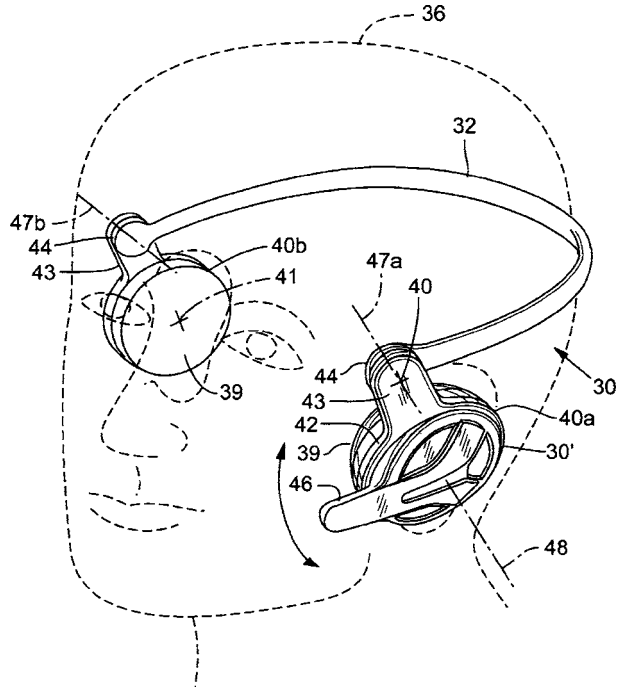
soldered as it is held in place by the various parts of housing 102. Acoustic ports 124 complete the microphone assembly.—NAS

7,406,172

43.38.Si REVERSIBLE BEHIND-THE-HEAD MOUNTED PERSONAL AUDIO SET WITH PIVOTING EARPHONE

Dominic Amae, assignor to Logitech Europe S.A.
29 July 2008 (Class 379/430); filed 16 February 2006

This patent teaches that combination headsets come in two basic varieties: stereo and "monoaural." In either case, only one microphone boom is



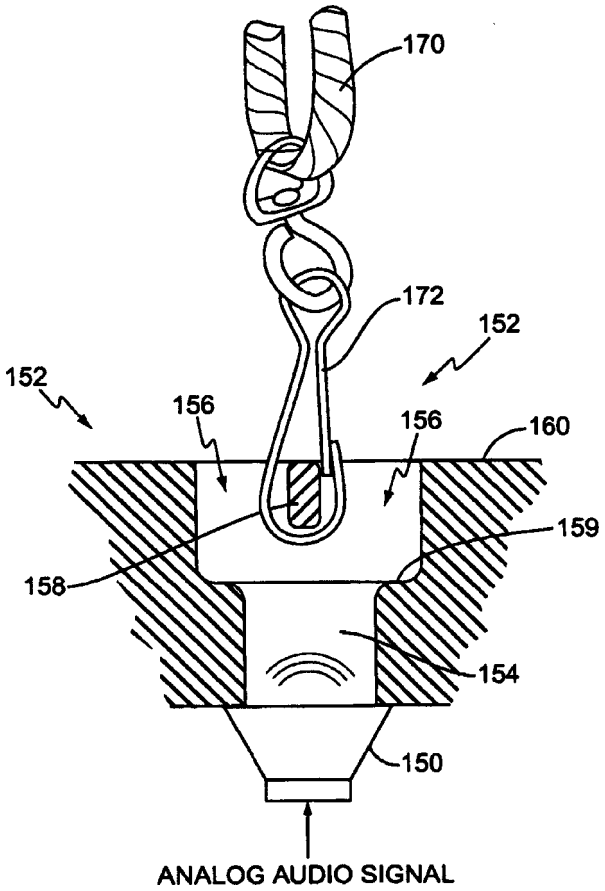
provided, usually attached to the right earcup. But suppose the user wants the microphone on the left? In this case simply flip and rotate and the job is done—no reassembly is required.—GLA

7,412,267

43.38.Si MOBILE DEVICE WITH A COMBINATION ATTACHMENT AND ACOUSTIC PORT

Chris Eaton and Matt Murray, assignors to Sony Ericsson Mobile Communications AB
12 August 2008 (Class 455/575.1); filed 30 October 2003

Small, handheld devices such as pagers and cellular phones are miniaturized to the greatest possible extent. There is no extra space for a loudspeaker grill, so sound may be squirted out through a small opening or a



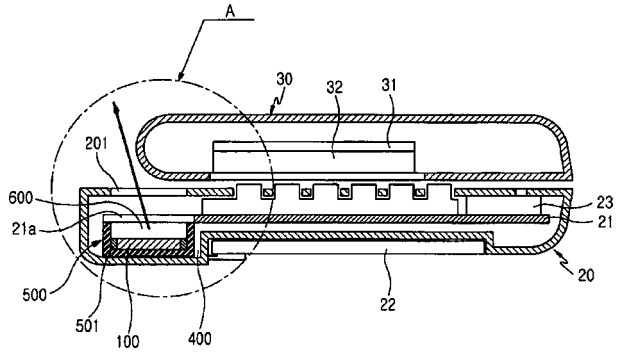
slot. Since many of these devices are fitted with a carrying strap, this patent suggests that the two functions can be combined as shown. From this modest concept the patent attorneys managed to derive 25 patent claims, an achievement more impressive than the invention itself.—GLA

7,408,514

43.38.Si SPEAKER OF A PORTABLE TERMINAL HAVING A RESONANCE SPACE

Sung-Soo Go, assignor to Samsung Electronics Company, Limited
5 August 2008 (Class 343/702); filed in Republic of Korea 18 October 2004

According to this patent, most cellular telephones suffer from a grievous design fault: "…the speaker device is mounted on the folder without a resonance space, thus producing a poor undertone so that the listener cannot hear a wide range of high quality sound." To remedy this sad state of affairs, speaker unit 100 is separated from sound opening 201, creating a resonance



space 600 in front of the diaphragm. If the diagram is accurate then there is no rear cavity at all, somewhat diminishing one's expectation of high quality sound. The diagram also indicates that the hinged cover only partly blocks opening 201; however, the patent claims tippy toe around this feature by stating only that "…sound generated by the speaker module advances in multiple directions, after passing through the resonance space, regardless of the folder being opened or closed."—GLA

7,401,397

43.40.Dx METHOD OF PRODUCING AN INERTIAL SENSOR

Kirill V. Shcheglov and A. Dorian Challoner, assignors to The Boeing Company
22 July 2008 (Class 29/595); filed 9 March 2006

This patent describes a microelectronic mechanical gyroscope made of silicon, said to have improved sensitivity and bias stability compared to other similar gyros due to its construction. The inventors compare the gyro to traditional quartz hemispheric resonator gyros and to silicon ring gyros of more recent invention. The device is essentially a silicon disk perforated by annular slots, with the annular cavities housing the drive and sense electrodes. The inventors claim that the sensitivity and stability improvements are gained by committing more surface area to electrodes and laser trimming the perimeter of the silicon disk after testing. Much of the text is given to discussion of the vibrational modes in and out of the plane of the disk, and their coupling. This appears to be an advance in the state of the art for compact commercial gyros, but it is hard to tell as no performance data are given.—JAH

7,401,000

43.40.Le ACOUSTIC VIBRATION ANALYZING APPARATUS AND ACOUSTIC VIBRATION ANALYZING METHOD, PROGRAM FOR ANALYZING ACOUSTIC VIBRATION, AND RECORDING MEDIUM, READABLE BY COMPUTER, ON WHICH PROGRAM FOR ANALYZING ACOUSTIC VIBRATION IS STORED

Takaaki Nakamura, assignor to Honda Motor Company, Limited
15 July 2008 (Class 702/145); filed in Japan 28 August 2003

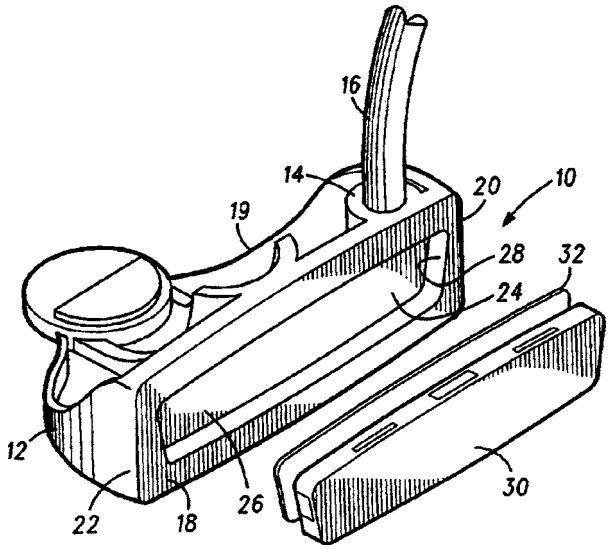
The apparatus described in this patent is intended to assist in identification of the sources of unusual sounds or vibrations in automotive transmissions. Sounds and vibrations from the transmission, from the engine, and other devices are sensed and recorded together with an elapsed time signal, with vehicle speed data, and with comments from the test driver. Frequency analysis of the sounds and vibrations is performed and the orders (multiples of rotational speed of a selected component) are determined. These are displayed and correlated with listener's observations. Later analysis permits reproduction of single sound components at a time so that a listener can compare these to his subjective observations.—EEU

7,374,499

43.40.Tm GOLF CLUB HEAD WITH DUAL DUROMETER FACE INSERT

David D. Jones and Leslie J. Bryant, assignors to Karsten Manufacturing Corporation
20 May 2008 (Class 473/340); filed 1 August 2005

The face of insert 30 of club head 10 is made from a high durometer material which provides the "desirable acoustic feedback" when striking the



golf ball. The surround of insert 30 is made of a lower durometer material to allow the insert to "float" within cavity 24 of body 12 which provides "a desirable feel."—NAS

7,401,690

43.40.Tm DAMPED CLUTCH PLATE SYSTEM AND METHOD

Jeff Lazowski and Eric Denys, assignors to Material Sciences Corporation
22 July 2008 (Class 192/30 V); filed 2 September 2005

A clutch plate according to this invention consists of a sandwich structure composed of a viscoelastic core between steel layers. The viscoelastic material can be selected to be effective at different temperatures and may also serve as the bond between the steel layers.—EEU

7,406,923

43.40.Tm SYSTEMS AND METHODS FOR REDUCING VIBRATIONS

Donald Wayne Allen *et al.*, assignors to Shell Oil Company
5 August 2008 (Class 114/243); filed 7 April 2006

Strakes attached to cylindrical structures that are exposed to fluid currents can reduce vibrations resulting from vortex shedding. This patent describes particular strake configurations that are claimed to be comparatively easy to manufacture, transport, store, and install, particularly for underwater applications.—EEU

7,406,939

43.40.Vn ENGINE VIBRATION SUPPRESSION DEVICE AND SUPPRESSION METHOD THEREOF

Yasuyuki Asahara and Kazuhiro Fujikane, assignors to Nissan Motor Company, Limited
5 August 2008 (Class 123/192.1); filed in Japan 7 November 2005

When the engine in a hybrid automobile is restarted there occur torque variations as the intake air is compressed and expanded, leading to vibration of the engine on its mountings. This patent describes an active control system that provides counteracting torques in order to reduce these vibrations. A controller estimates the internal cylinder pressure of the engine on the basis of the engine's operating conditions, calculates the desired opposing torque, and causes the hybrid's electrical motor to generate this torque.—EEU

7,409,741

43.40.Yq TOOTHBRUSH WITH TUNED VIBRATING HEAD

William S. Dworzan, Santa Ana, California
12 August 2008 (Class 15/22.1); filed 24 June 2003

In this toothbrush design a rotor with an eccentric mass is located near the brush's head. The rotor's rotational frequency matches the natural frequency of the toothbrush structure, enabling the generation of relatively intense resonant vibrations with minimal battery energy usage.—EEU

7,402,537

43.50.Gf AIR PERMEABLE, WATER REPELLENT, ACOUSTIC DOOR WATERSHIELD

Steven E. Lenda and Donald P. Marriott, assignors to Creative Foam Corporation
22 July 2008 (Class 442/50); filed 17 December 2003

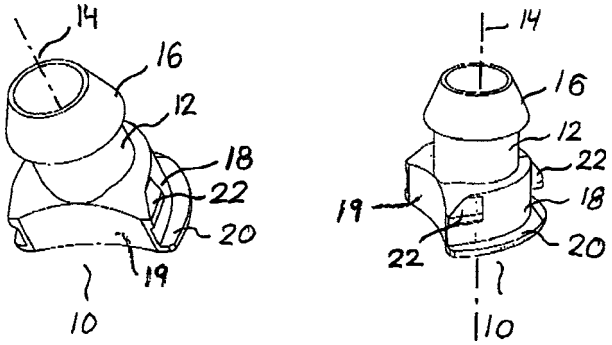
Watershields in automotive vehicles typically are applied to door panels in order to keep water, dirt, and sound from entering the passenger compartment. The watershield described in this patent is intended to be installed between the interior panel of a door and the attached interior trim. It is constructed of a layer of water repellant nonwoven scrim, which is air permeable, atop an acoustically absorptive layer formed from a lofted nonwoven fiber pad.—EEU

7,401,679

43.66.Ts ADAPTOR FOR MOUNTING A SOUND TUBE IN AN EARPIECE

Klaus Sommer Ipsen, assignor to GN Resound A/S
22 July 2008 (Class 181/129); filed in Denmark 14 February 2003

An adaptor connected to a sound tube inserts into the in-the-ear shell portion of an open behind-the-ear hearing aid fitting and also provides



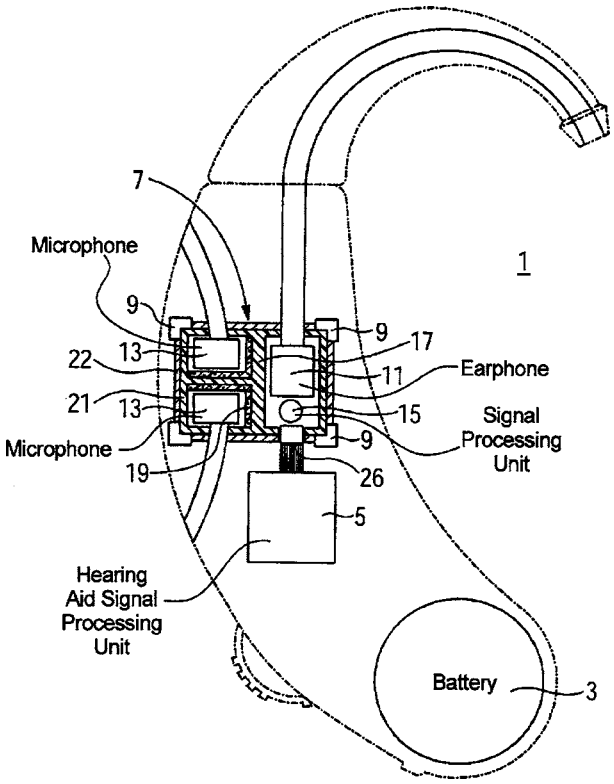
venting via at least one opening.—DAP

7,400,738

43.66.Ts ACOUSTIC MODULE FOR A HEARING AID DEVICE

Torsten Niederdränk *et al.*, assignors to Siemens Audiologische Technik GmbH
15 July 2008 (Class 381/318); filed in Germany 27 June 2002

A detachable acoustic module in a hearing aid houses the microphone(s) and receiver, seals them off acoustically, and mechanically isolates them from each other and from external vibration. The design is said to



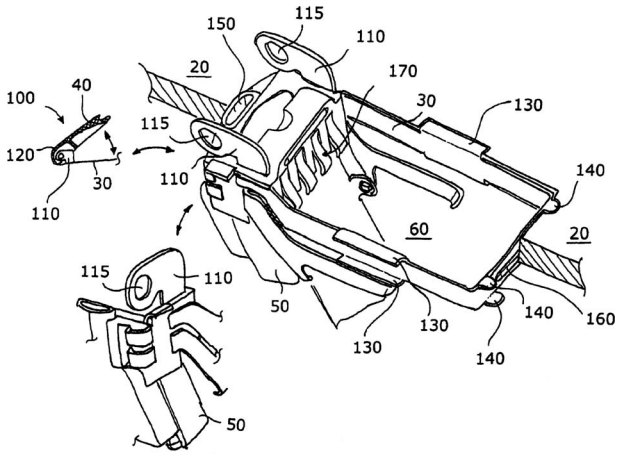
reduce size, shield against electromagnetic interference, and simplify manufacture since the module can be used in different hearing aid designs. The fixed vibration-coupling characteristics of the module are determined by the hearing aid signal processor and used to further improve feedback reduction.—DAP

7,403,630

43.66.Ts MINIATURE HEARING AID INSERT MODULE

Martin Bondo Jørgensen and Karsten Videbæk, assignors to Sonion Roskilde A/S
22 July 2008 (Class 381/324); filed 30 April 2004

An insertable module for a hearing aid contains a socket and battery lid that are attached to an electrically conductive metal frame for accommodating the hearing aid battery. The adherence of the removable module to the faceplate is made more robust by making the frame rigid, rectangular, and "closed loop" with only a small opening and with projections to engage the faceplate. Overall advantages are said to be easier manufacturability,



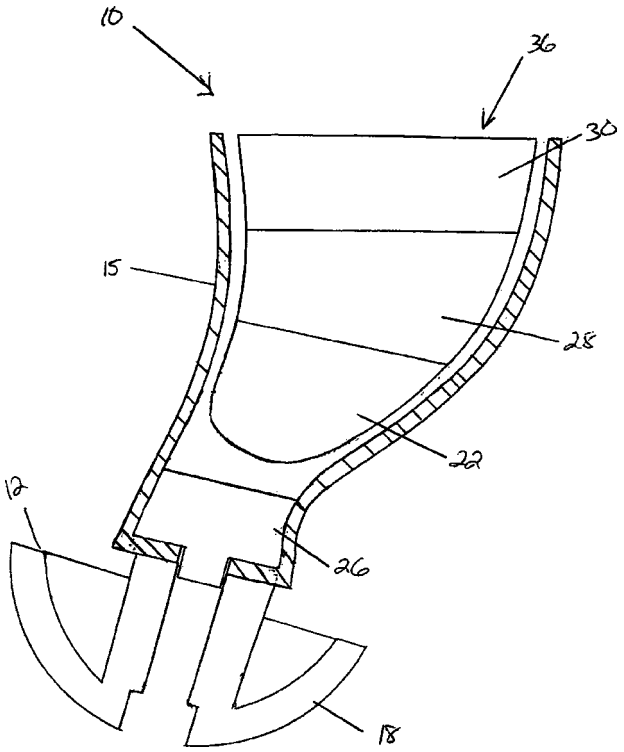
repair, and customization for different wearers' ears as well as greater battery concealment.—DAP

7,403,629

43.66.Ts DISPOSABLE MODULAR HEARING AID

John G. Aceti *et al.*, assignors to Sarnoff Corporation
22 July 2008 (Class 381/324); filed 13 March 2000

To allow replacement of a highly flexible earmold tip used with a deeply fitting disposable modular hearing aid, a replaceable base unit may contain in one embodiment a shell, electronics, battery, controls, receiver,



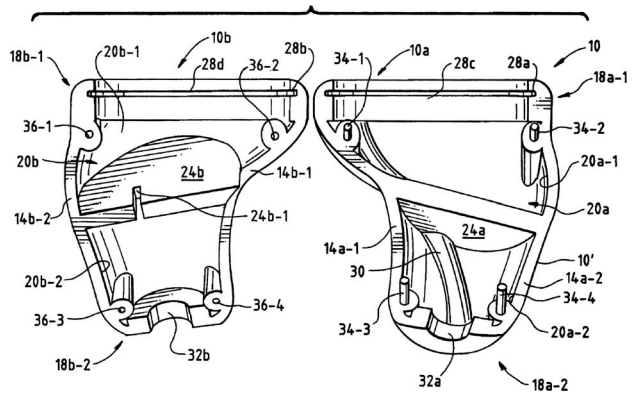
and microphone. The replaceable earmold, which may alternately house the receiver, electronics, and battery, has a mushroom-shaped ultrasoft tip and means for replaceably attaching to the base unit.—DAP

7,407,035

43.66.Ts SPLIT SHELL SYSTEM AND METHOD FOR HEARING AIDS

Robert Stinauer *et al.*, assignors to GN ReSound A/S
5 August 2008 (Class 181/135); filed 25 February 2003

To facilitate easier manufacturing of custom hearing aids, a two-piece shell is utilized to permit easier access to the inside of the hearing aid. The shell is split longitudinally, approximately in a plane along the axis of the



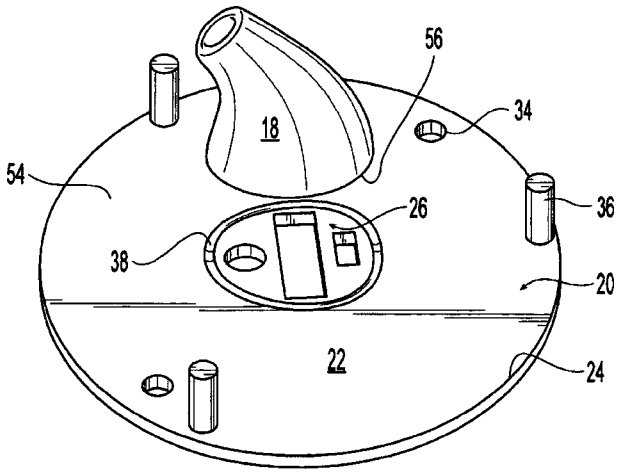
ear canal, and no faceplate is used. A modular component assembly may be used, and after joining the two parts which may have locking features, the shell may be inserted into an elastomeric sheath.—DAP

7,409,069

43.66.Ts FACEPLATE MOAT AND CUT OUT FOR HEARING INSTRUMENT

Thomas Geoffery Fletcher *et al.*, assignors to Sound Design Technologies, Limited
5 August 2008 (Class 381/322); filed 3 December 2004

To facilitate easier manufacturing of custom hearing aids, a moat or trench is cut into the surface of the faceplate that includes a wall structure



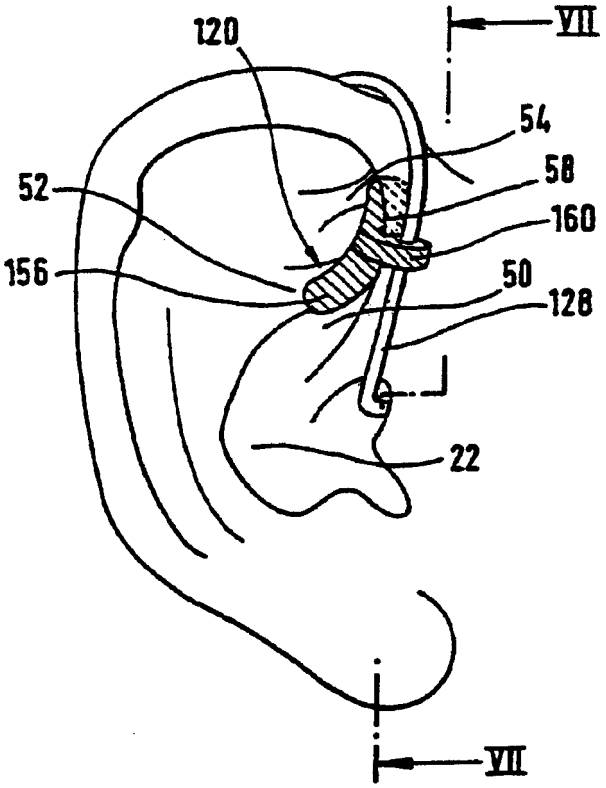
for coupling with the shell portion. The base of the shell is fabricated slightly undersized and presses into the moat in the faceplate after cement has been inserted into it.—DAP

7,412,068

43.66.Ts OTOPLASTY FOR BEHIND-THE-EAR (BTE) HEARING AIDS

Erich Bayer, Straubing, Germany
12 August 2008 (Class 381/328); filed in Germany 14 October 1999

An arc-shaped retention clip is used to better retain in place the flexible tubing and earpiece used with open behind-the-ear hearing aid fittings



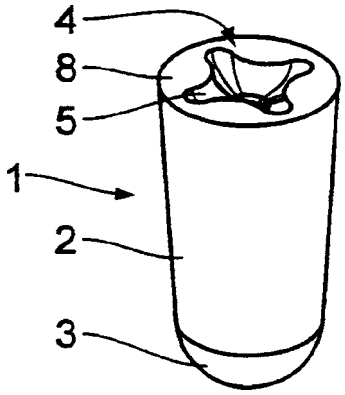
or cochlear implant systems with behind-the-ear processors. The clip adapts and attaches to the outer edge of the wearer's cavum concha near the cymba portion of the outer ear.—DAP

7,264,081

43.66.Vt HEARING PROTECTION EARPLUG

Stefan Bruck, assignor to Uvex Arbeitsschutz GmbH
4 September 2007 (Class 181/135); filed in Germany 10 June 2005

The addition of an "X" shaped cavity (5) in a foam earplug provides



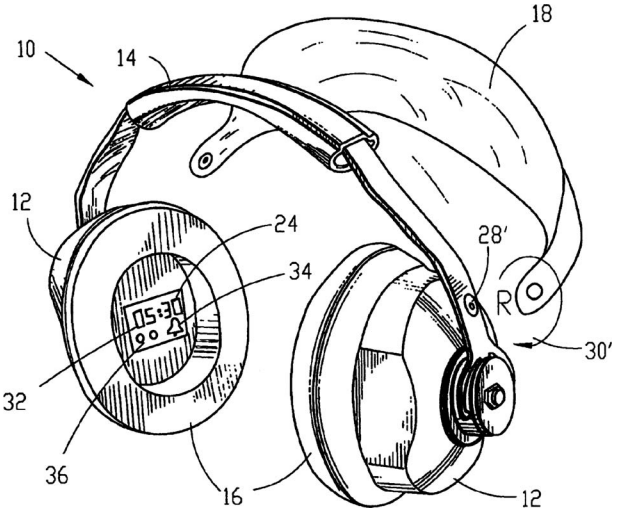
improved handling and comfort.—JE

7,268,669

43.66.Vt PERSONAL SENSORY REDUCTION SYSTEM, AND METHOD

Michael McEvoy, Pensacola, Florida
11 September 2007 (Class 340/309.16); filed 5 November 2004

The addition of a blindfold to an earmuff type hearing protector is



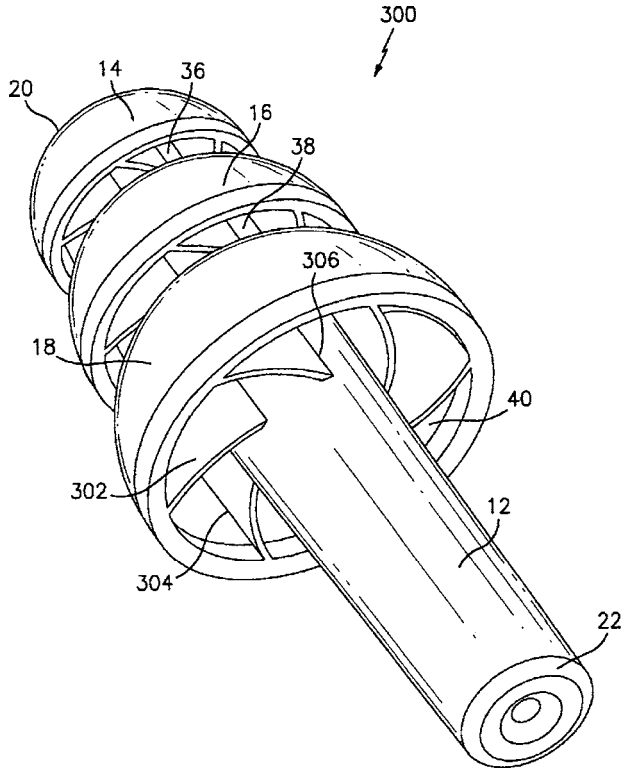
described. The earmuff can accommodate a programmable alarm clock (32).—JE

7,314,047

43.66.Vt HEARING PROTECTION DEVICE

Robert Falco, assignor to Cabot Safety Intermediate Corporation
1 January 2008 (Class 128/864); filed 27 January 2003

With the addition of ribs and/or flanges (302) to the inside surfaces (40) of the three primary flanges (18) of a molded hearing protector,



additional stiffness of the primary flanges is achieved to prevent collapse while preserving comfort.—JE

7,286,980

43.72.Dv SPEECH PROCESSING APPARATUS AND METHOD FOR ENHANCING SPEECH INFORMATION AND SUPPRESSING NOISE IN SPECTRAL DIVISIONS OF A SPEECH SIGNAL

Youhua Wang and Koji Yoshida, assignors to Matsushita Electric Industrial Company, Limited
23 October 2007 (Class 704/205); filed in Japan 31 August 2000

Sixteen embodiments of a method for speech enhancement (noise suppression) are presented, which in essence implement comb filtering. In theory, the passbands of the comb filter lie on the speech harmonics. The likelihood of this being true is said to be increased over the prior art in this method by highly accurate pitch detection, though this reviewer could not really understand how.—SAF

7,289,626

43.72.Dv ENHANCEMENT OF SOUND QUALITY FOR COMPUTER TELEPHONY SYSTEMS

George E. Carter and Bianka Skubnik, assignors to Siemens Communications, Incorporated
30 October 2007 (Class 379/387.02); filed 7 May 2001

A number of noise-reduction techniques are put forth in this patent, all of which apply frequency domain processing or filtering to the signal at the client side in a computer telephony application. It is suggested that the noise can be both detected and reduced by simple spectral processing and filtering. These rather obvious methods, which are outlined without any significant details, are offered as an alternative to client-side hardware noise gates, which is like a straw man to make the patent seem novel.—SAF

7,289,955

43.72.Dv METHOD OF DETERMINING UNCERTAINTY ASSOCIATED WITH ACOUSTIC DISTORTION-BASED NOISE REDUCTION

Li Deng *et al.*, assignors to Microsoft Corporation
30 October 2007 (Class 704/226); filed 20 December 2006

This patent has a tangled skein for its history, and constitutes a divisional of a continuation in part (now another patent) of a continuation in part (now a third patent) of yet a fourth, U.S. Patent No. 7,107,210, "Method of Noise Reduction Based on Dynamic Aspects of Speech." The background notes that, while there is uncertainty remaining in the cleaned signal used for speech recognition, prior art techniques either ignore it or overcompensate by eliminating speech frames with too much uncertainty. Methods are put forth in the patent which can, first, measure the uncertainty resulting from a noise reduction process, and then utilize what helpful information exists in speech frames having high uncertainty. Without noise, the observation vector y is the same as a clean vector x, and so evaluating the probability $p(y|m)$ for a Gaussian model component m would, under such unrealistic circumstances, be the same as $p(x|m)$. The method described instead determines $p(y|m)$ by marginalizing a joint conditional density function $p(y,x|m)$ over all possible (unseen) clean speech vectors.—SAF

7,295,968

43.72.Dv DEVICE AND METHOD FOR PROCESSING AN AUDIO SIGNAL

Franck Bietrix and Hubert Cadusseau, assignors to Wavecom
13 November 2007 (Class 704/200); filed in France 15 May 2001

"This invention relates to, in particular, the reduction or cancellation of noise in an audio signal via a digital communication device…." It is noted by the authors that, while international standards specify a 160-point frame for speech vocoding, noise cancellation processing generally uses 256-point frames to increase Fourier transform efficiency. The mismatch produces asynchrony among the various frame cuts of the signal, which in turn necessitates a number of processing accommodations in spectral domain noise cancellation that increase computational complexity. The patent describes, in unclear peculiar English, a method for synchronizing the differently sized frames by beginning the noise cancellation window 96 points before the beginning of a speech window so that they have the same end point. The method promises to eliminate all the processing tricks necessitated by the more usual methods.—SAF

7,292,977

43.72.Fx SYSTEMS AND METHODS FOR PROVIDING ONLINE FAST SPEAKER ADAPTATION IN SPEECH RECOGNITION

Daben Liu, assignor to BBNT Solutions LLC
6 November 2007 (Class 704/236); filed 16 October 2003

According to this patent, prior art techniques for speaker adaptation in a speech recognition system usually involves time-intensive and non-real-time techniques which require decoding everything twice, and also batch processing of a complete utterance's worth of segments to perform segment clustering by speaker. A number of techniques are suggested here to address these issues, including real-time speaker-change detection logic (another patent is referenced for this) and a decoding scheme involving a forward stage with reduced data rate, a backward stage with increased accuracy, and a rescoring stage with maximum accuracy (using the most complex sort of hidden Markov model). The level of detail in the patent is similar to that in this review.—SAF

7,295,970

43.72.Fx UNSUPERVISED SPEAKER SEGMENTATION OF MULTI-SPEAKER SPEECH DATA

Allen Louis Gorin *et al.*, assignors to AT&T Corporation
13 November 2007 (Class 704/221); filed 24 January 2003

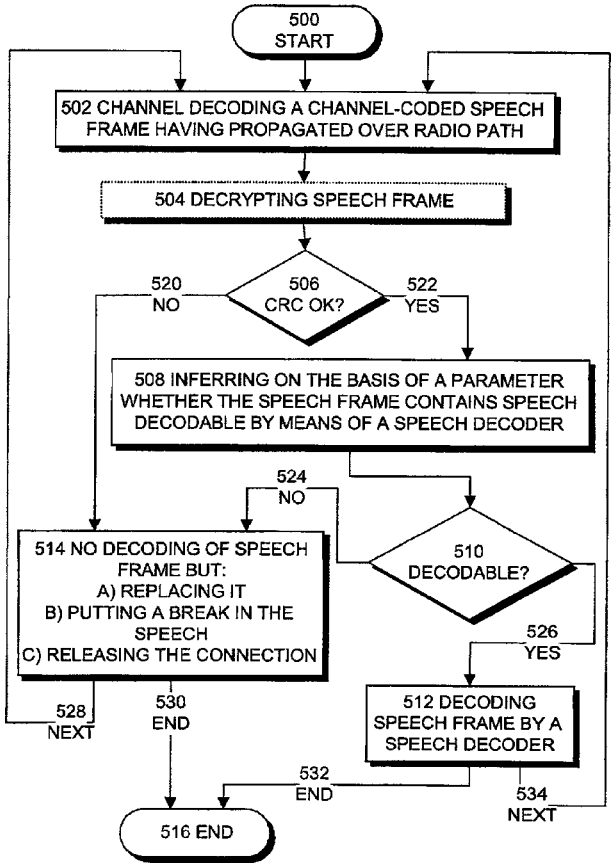
Multiple speakers are identified in a dialog by applying automatic segmentation routines in an iterative fashion. Between iterations and initially, speaker models are compared on each segment using a "generalized likelihood ratio" to determine the most likely speaker, and segments are clustered accordingly by speaker. A Gaussian mixture model is constructed for the pooled data from each segment cluster. The input speech is then compared with the model to obtain a detection score as a function of time; this error measure is used to obtain a new segmentation for the next iteration. The process can be repeated until a stable segmentation is achieved.—SAF

7,401,022

43.72.Gy PROCESSING A SPEECH FRAME IN A RADIO SYSTEM

Petri Ahonen, assignor to Nokia Corporation
15 July 2008 (Class 704/270); filed in Finland 19 September 2000

If a speech frame that has propagated over a radio path is found to be free of defects via channel code decoding, the value of at least one speech parameter in the channel-decoded speech frame is used to determine if the



speech frame contains speech that is decodable by a speech decoder. The speech frame is decoded only if speech is inferred to be present.—DAP

7,406,412

43.72.Gy REDUCED COMPUTATIONAL COMPLEXITY OF BIT ALLOCATION FOR PERCEPTUAL CODING

Stephen Decker Vernon *et al.*, assignors to Dolby Laboratories Licensing Corporation
29 July 2008 (Class 704/229); filed 20 April 2004

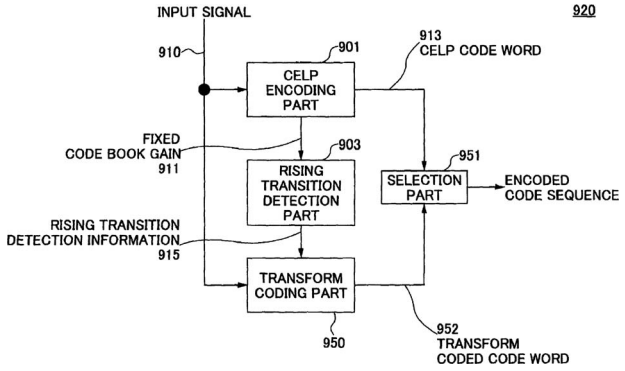
One or more coding parameters used in bit allocation for an audio coding system are determined by first making an estimate of an offset of a second masking curve from a calculated perceptual masking curve for an audio signal. The optimum value of the offset is determined by taking the difference between the number of bits that would be allocated if an initial value of offset is used and the number of bits that are actually available for allocation.—DAP

7,406,410

43.72.Gy ENCODING AND DECODING METHOD AND APPARATUS USING RISING-TRANSITION DETECTION AND NOTIFICATION

Kei Kikuiri *et al.*, assignors to NTT DoCoMo, Incorporated
29 July 2008 (Class 704/221); filed in Japan 8 February 2002

To prevent annoying noises caused by pre-echo from degrading subjective quality, an input signal that has been encoded by code-excited linear prediction (CELP) is decoded by detecting a rising transition in the gain of the excitation vectors that were determined by a first decoder. The presence



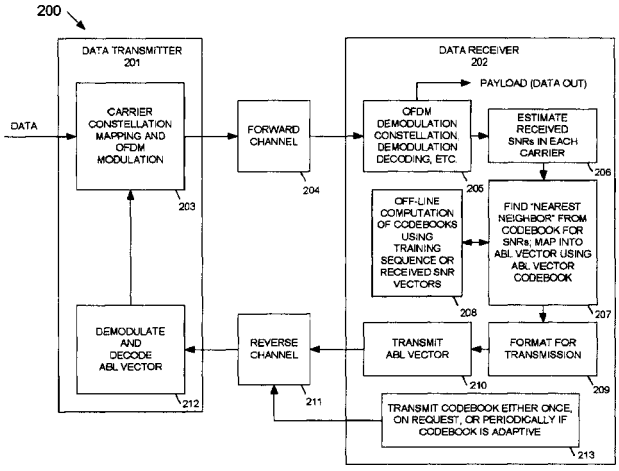
of the rising amplitude transition notifies a second decoder that decodes signals not encoded via CELP. As a result of the transition, the second decoder outputs a signal obtained by decoding an enhancement layer code word determined by encoding a difference between the input signal and a signal decoded by the first decoder.—DAP

7,408,998

43.72.Gy SYSTEM AND METHOD FOR ADAPTIVE BIT LOADING SOURCE CODING VIA VECTOR QUANTIZATION

John M. Kowalski, assignor to Sharp Laboratories of America, Incorporated
5 August 2008 (Class 375/260); filed 8 March 2005

To reduce computational complexity, a data receiver calculates a level of modulation for at least one carrier in a multicarrier communication



system via a transmitted index to a codeword in a codebook. The codeword index is determined as a function of the received signal-to-noise ratio for the carrier.—DAP

7,286,986

43.72.Ja METHOD AND APPARATUS FOR SMOOTHING FUNDAMENTAL FREQUENCY DISCONTINUITIES ACROSS SYNTHESIZED SPEECH SEGMENTS

David Talkin, assignor to Rhetorical Systems Limited
23 October 2007 (Class 704/268); filed in United Kingdom 2 August 2002

In state of the art concatenative speech synthesis, the general method involves the literal splicing together of brief segments of recorded speech (suitably encoded). A constant problem for this technique stems from mismatched fundamental frequencies of two segments that are being spliced together. This patent proposes techniques to ameliorate this difficulty, which involves adjusting the fundamental frequency contour of each concatenated speech segment according to a linear function of the beginning and end values for each segment. One ingenious and novel proposal uses a physical model of coupled springs to derive the linear function, thereby introducing natural costs of frequency shifting by modeling with spring constants.—SAF

7,292,976

43.72.Ne ACTIVE LEARNING PROCESS FOR SPOKEN DIALOG SYSTEMS

Dilek Z. Hakkani-Tur *et al.*, assignors to AT&T Corporation
6 November 2007 (Class 704/236); filed 29 May 2003

When a spoken dialog system is fully deployed, its speech recognition models require constant updating and retraining. This process generally involves human phoneticians who transcribe speech fragments that have been recently received and perhaps incorrectly recognized. This patent proposes some techniques to assist in reducing the labor of this activity. In one example, a finite state machine lattice is generated from each audio dialog sound file using the cepstral representation, together with costs derived from acoustic and language model probabilities. The information is used to estimate word confidence scores, so that utterances received are ranked according to mean word score. The utterances having the poorest recognition confidence can then be prioritized for the human experts to proceed with the updating and retraining. The level of detail in the patent is similar to that in this review.—SAF

7,401,020

43.72.Ja APPLICATION OF EMOTION-BASED INTONATION AND PROSODY TO SPEECH IN TEXT-TO-SPEECH SYSTEMS

Ellen M. Eide, assignor to International Business Machines Corporation
15 July 2008 (Class 704/258); filed 29 November 2002

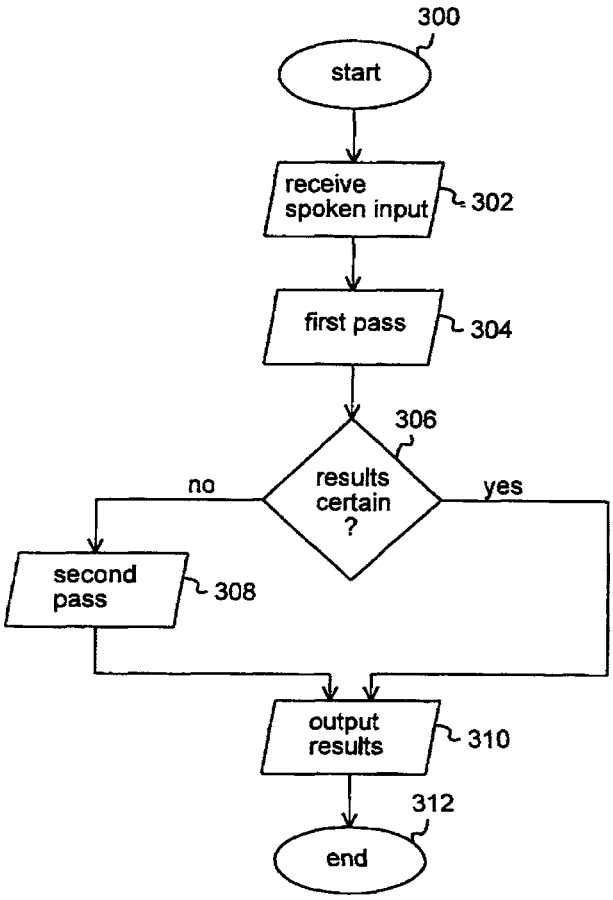
The patent would seem to be about methods for synthesizing speech with emotional content. However, almost nothing is said about how one might in, fact, perform such synthesis. The patent is primarily concerned with issues of how the input text could be marked up so as to indicate the desired emotion. Indeed, very little is said about that topic as well. All the patent really says is that the input text could be so marked up.—DLR

7,401,017

43.72.Ne ADAPTIVE MULTI-PASS SPEECH RECOGNITION SYSTEM

Hy Murveit *et al.*, assignors to Nuance Communications
15 July 2008 (Class 704/247); filed 4 April 2006

The goal is to speed up speech recognition without sacrificing accuracy. A first pass speech recognition produces several alternative speech expressions, each having an assigned score related to the certainty of matching the spoken input. Based on the first pass results, a second pass speech



7,289,958

43.72.Ne AUTOMATIC LANGUAGE INDEPENDENT TRIPHONE TRAINING USING A PHONETIC TABLE

Alexis P. Bernard and Lorin P. Netsch, assignors to Texas Instruments Incorporated
30 October 2007 (Class 704/255); filed 7 October 2003

A method is described for using a table of phonetic symbols and feature values in order to facilitate what might be called the "porting" of a speech recognition system from one language to another. Supposing one has a good speech recognizer for, say, English, the method indicates how best to transfer the monophone and triphone hidden Markov models (this is the architecture that is applicable here) to a new target language such as German. The phonetic table allows the selection of the most phonetically similar monophone in English to be used for each target German monophone model. Triphones follow from known techniques. The trouble is that the most crucial element of the procedure is only vaguely indicated, namely, what particular transformation must be made to the best English monophone model in order to yield a useful German model. The target model is simply stated to be a "compressed or stretched version" of the reference model.—SAF

recognition is selectively performed to match the spoken input only to those expressions identified in the first pass recognition. If one of the expressions in the first pass is assigned a score higher than a predetermined threshold, a second pass is not performed.—DAP

7,295,978

43.72.Ne SYSTEMS AND METHODS FOR USING ONE-DIMENSIONAL GAUSSIAN DISTRIBUTIONS TO MODEL SPEECH

Richard Mark Schwartz *et al.*, assignors to Verizon Corporate Services Group Incorporated
13 November 2007 (Class 704/240); filed 5 September 2000

In a typical speech recognition system using hidden Markov models, feature vectors obtained from the front end are organized into a number of multinomial Gaussian distributions. Dealing with these Gaussians takes up a large share of the computational complexity in speech recognition. This patent presents a technique by which "each dimension of the multinomial Gaussian distribution may be represented by a code for one of a small set of one-dimensional Gaussian distributions." This particular codebook approximation yields a dramatic reduction in the complexity of handling Gaussian models. The possible effects on performance of the system are not discussed.—SAF

7,406,408

43.72.Ne METHOD OF RECOGNIZING PHONES IN SPEECH OF ANY LANGUAGE

Bradley C. Lackey *et al.*, assignors to The United States of America as represented by the Director, National Security Agency
29 July 2008 (Class 704/8); filed 24 August 2004

In designing a speech recognition system capable of analyzing input speech expressed in any language or phonetic system, the recognizer should ideally have access to several specific components, which would include listings of the phonetic elements (or "phones") which occur in each of the languages to be handled, a database of the phonetic characteristics of each of the listed phones, a pronunciation dictionary for each language of interest, speech transcripts and frequency-of-occurrence information for each language, transcripts and frequency information for each utterance to be processed, and finally a sample of speech which has not yet been analyzed. The patent presents a variety of strategies by which (it is hoped) satisfactory analyses can be performed in the event that one or more of these prerequisites is not at hand. There is no detail whatsoever of how any of the listed steps might actually be accomplished.—DLR

7,412,390

43.72.Ne METHOD AND APPARATUS FOR SPEECH SYNTHESIS, PROGRAM, RECORDING MEDIUM, METHOD AND APPARATUS FOR GENERATING CONSTRAINT INFORMATION AND ROBOT APPARATUS

Erika Kobayashi *et al.*, assignors to Sony France S.A.
12 August 2008 (Class 704/267); filed in the European Patent Office 15 March 2002

At last, here is a patent that actually says how to do something. The task is to provide a humanlike robot with the capability of speaking with some degree of humanlike speech. A variety of situation analyzers would characterize the current situation in ways that cause the internal state of the system to make adjustments on each of two independent axes, labeled as "valence" and "arousal." This results in the internal state moving to one of several conditions, labeled as "calm," "sadness," "comfort," "happiness," or "anger." Depending on the internal state, the text sent to the speech synthesis system is marked up in any of several ways so as to produce synthetic speech with the appropriate emotional content.—DLR

7,402,136

43.80.Vj EFFICIENT ULTRASOUND SYSTEM FOR TWO-DIMENSIONAL C-SCAN IMAGING AND RELATED METHOD THEREOF

John A. Hossack *et al.*, assignors to University of Virginia Patent Foundation
22 July 2008 (Class 600/447); filed 15 January 2004

This system consists of a two-dimensional array transducer, transmission circuitry, and reception circuitry that includes channel memory buffers. The contents of the memory buffers are read into a beamformer with different focusing values for each buffer reading cycle. The beamformer calculates image values at multiple points along a line.—RCW